Supplementary Materials for GenPIE: A Time-Resolved Plenoptic Imager

ZIHENG WANG, The Chinese University of Hong Kong, Shenzhen, China
SIYUAN SHEN, ShanghaiTech University, China
HUANYU XU, ShanghaiTech University, China
KAICHUN QIAO, ShanghaiTech University and Deemos Technology, China
LONGWEN ZHANG, ShanghaiTech University and Deemos Technology, China
QIXUAN ZHANG, ShanghaiTech University and Deemos Technology, China
QILIN SUN, The Chinese University of Hong Kong, Shenzhen and Point Spread Technology, China
SHIYING LI*, ShanghaiTech University, China
JINGYI YU*, ShanghaiTech University, China

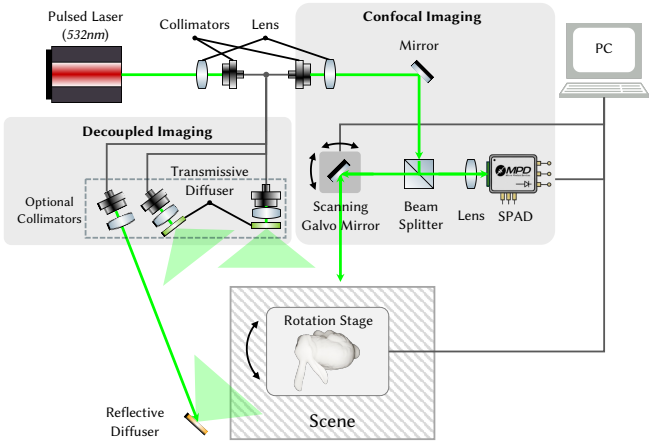

Fig. 1. 2D optical sketch of our imaging system. In the confocal mode, the pulsed laser is routed through the beam splitter and galvo mirrors, and the back-scattered signal is collected by the SPAD along the same optical path. In the decoupled mode, the illumination path is separated from the detection path using optional collimators and transmissive or reflective diffusers, enabling flexible illumination while the scene is rotated on the motorized stage.

In this supplementary material, we provide the 2D optical sketch of our imaging system in Fig. 1, discuss more applications as special cases of inverse transient rendering in Sec. A, provide more ablation studies in Sec. B, and the detail of mesh quality metrics in Sec. C.

A More Applications

We discuss two applications as constrained scenarios of inverse transient rendering. In NLOS imaging, part of the scene geometry, i.e., the relay wall, is known, while the hidden geometry remains unknown and is recovered by optimizing indirect transient measurements. In imaging through scattering media, if the volume extent and position of the scattering medium are known, the medium scattering properties and the hidden scene are what to recover. These two cases demonstrate how our framework can be specialized to different physical constraints by fixing known scene components and optimizing the remaining unknown parameters from time-resolved measurements.

A.1 NLOS Imaging as Inverse Transient Rendering

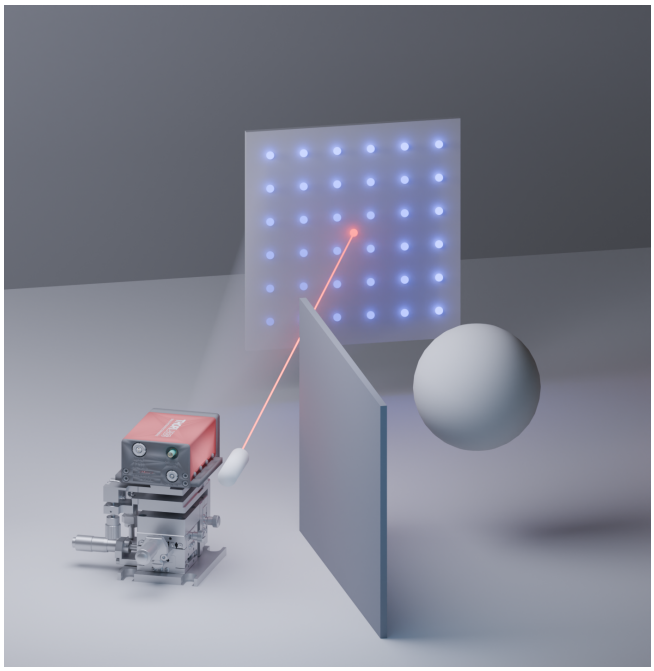

Fig. 2. Synthetic setup configuration. We conduct synthetic non-line-of-sight imaging under a non-confocal configuration. The transient camera's field of view uniformly covers the entire diffuse relay wall, with the illumination point positioned at the center of the wall.

Specially, we demonstrate our framework as an end-to-end solution for NLOS imaging. We conduct experiments where the relay

*Corresponding authors.

Authors' Contact Information: Ziheng Wang, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China, zihengwang3@link.cuhk.edu.cn; Siyuan Shen, ShanghaiTech University, Shanghai, China; Huanyu Xu, ShanghaiTech University, Shanghai, China; Kaichun Qiao, ShanghaiTech University and Deemos Technology, Shanghai, China; Longwen Zhang, ShanghaiTech University and Deemos Technology, Shanghai, China; Qixuan Zhang, ShanghaiTech University and Deemos Technology, Shanghai, China; Qilin Sun, The Chinese University of Hong Kong, Shenzhen and Point Spread Technology, Shenzhen, China; Shiying Li, ShanghaiTech University, Shanghai, China, lishy1@shanghaitech.edu.cn; Jingyi Yu, ShanghaiTech University, Shanghai, China, yujingyi@shanghaitech.edu.cn.

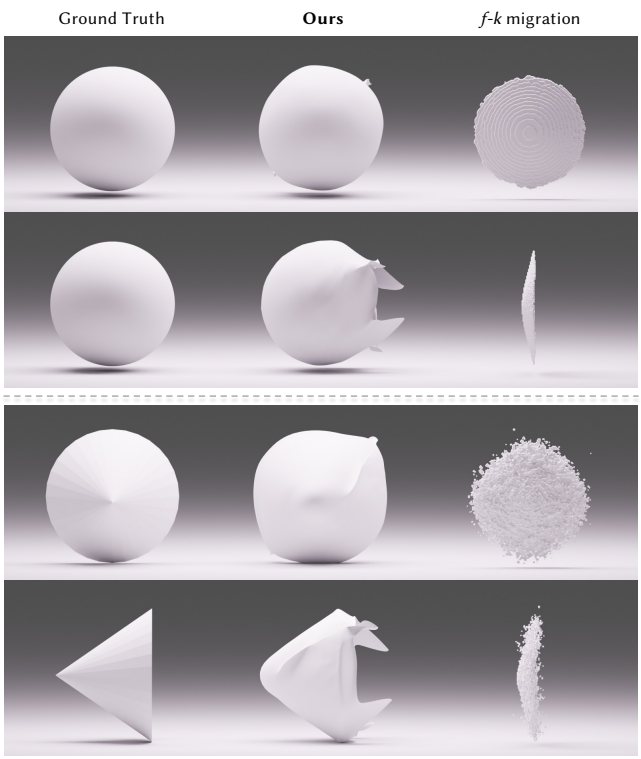

Fig. 3. Multi-view reconstruction results. **Top:** front and side views of a hidden sphere; **Bottom:** front and side views of a hidden cone.

wall geometry is roughly known while the hidden object remains unknown. Our synthetic setup (Fig. 2) mimics real-world NLOS imaging conditions: a pulsed laser is directed at the center of the diffuse relay wall, while a single-photon camera with a field of view covering the entire wall surface captures indirect light reflections. We record time-resolved photon arrivals at an optical path length resolution of 0.0096 m. No scanning of predefined patterns is required due to the full-wall coverage. We treat NLOS reconstruction as a constrained version of general inverse transient rendering. We begin with a lidar scan of the visible relay wall, which is used to prompt the foundation model to generate a visible planar surface. We initialize the hidden object geometry as a randomly positioned square surface within a predefined bounding box, then iteratively refine it through differentiable rendering. The known wall geometry remains fixed during optimization.

As shown in Fig. 3, our method is compared against f-k migration [Lindell et al. 2019]. The results demonstrate that our explicit representation of hidden objects successfully recovers complete geometry by directly optimizing multi-bounce signals, including self-reflection effects. In contrast, f-k migration suffers from limited spatial resolution and elevated noise sensitivity due to its reliance on linear inversion.

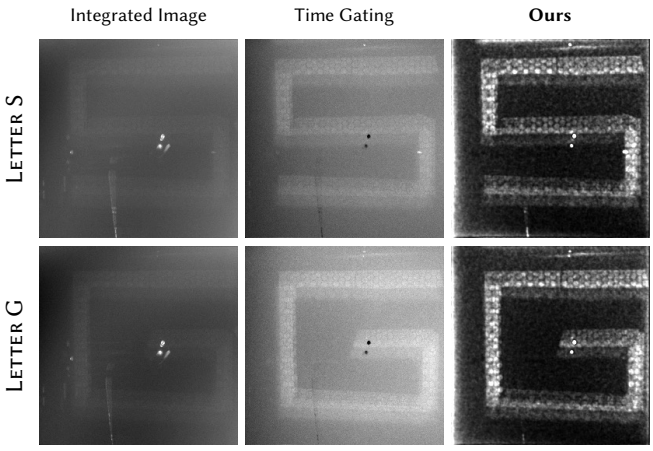

Fig. 4. **Left:** Integrated total intensity from measured transient response. **Middle:** Integrated images after gating the backscattered photons. **Right:** Results after applying our deconvolution model to gated transient measurements.

A.2 Imaging through Scattering Media

The ability of our framework to jointly optimize scene parameters, including illumination intensity and volumetric scattering coefficients, naturally extends to a constrained scenario of imaging through scattering media.

Existing time-resolved approaches for scattering imaging [Du et al. 2022; Lindell and Wetzstein 2020] typically rely on pre-calibrated scattering coefficients as fixed priors [Patterson et al. 1989] and are unable to explicitly model the photons backscattered by the medium itself. This limits their applicability in practical scenarios where backscatter dominated the signal [Maccarone et al. 2019, 2015; Wang et al. 1991], often forcing these methods to treat backscattering as unstructured noise. Furthermore, obtaining accurate scattering coefficients usually requires a difficult separate calibration procedure.

In contrast, our framework utilizes the backscattered signal to fit the medium's properties. In practice, directly optimizing the geometry of the hidden object through the medium using a volumetric path tracer is computationally prohibitive due to high-variance gradients and strong multi-scattering coupling. To address this, we show that this problem can be solved using confocal imaging by combining a convolution approximation with forward transient path tracing under known scattering coefficients.

Modeling light transport through scattering media. To derive a practical forward model for our confocal setup, we directly formulate the transient response under point-by-point confocal scanning, rather than starting from a general directional radiance field. Let $\mathbf{x}_f$ denote the scanned illumination/detection position at the front surface of the tank, and let $M(\mathbf{x})$ denote the hidden planar target response on the back surface. We assume that the scattering medium is approximately homogeneous within the field of view, so that propagation through a medium layer of depth D can be approximated by a shift-invariant spatio-temporal kernel $h_D(\mathbf{x}, t)$.

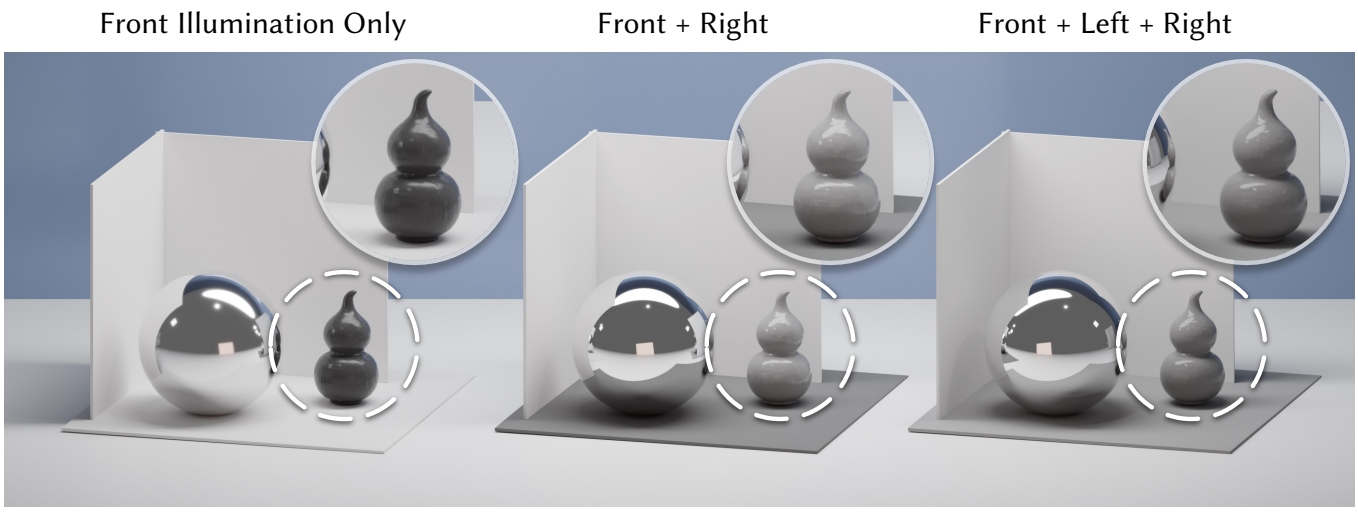

Fig. 5. Ablation on the number of illumination probes. Additional illumination probes provide stronger constraints for ambiguous metallic and specular regions, leading to more stable reconstructions than using front illumination alone. This figure is rendered using Blender by taking the optimized BSDFs of different objects.

Under confocal scanning, each measurement is generated by a single point source incident at $\mathbf{x}_f$, rather than by a spatially extended illumination field. Ignoring the return scattering path for the moment, the transient response induced on the hidden planar target can therefore be written as

$$P_{\text{virt}}(\mathbf{x}_f, t) = \int_S h_D(\mathbf{x}_f - \mathbf{x}_s, t)\, M(\mathbf{x}_s)\, d\mathbf{x}_s = (h_D \circledast M)(\mathbf{x}_f, t), \quad \text{(A.1)}$$

where $\circledast$ denotes spatio-temporal convolution and S is the hidden planar surface. Here, P_{virt} can be interpreted as a virtual self-emissive transient pattern induced on the hidden surface by the confocal point illumination through the scattering medium.

To account for the return path from the hidden surface back to the detector, we apply a second scattering operator with the same shift-invariant approximation. This yields the measured transient response

$$R_{\text{model}}(\mathbf{x}_f, t) = (h_D \circledast P_{\text{virt}})(\mathbf{x}_f, t) = (h_D \circledast h_D \circledast M)(\mathbf{x}_f, t). \quad \text{(A.2)}$$

More generally, if the hidden surface introduces an additional local scattering or blur effect, we model it by a surface kernel $\Phi(\mathbf{x}, t)$, leading to

$$R_{\text{model}}(\mathbf{x}_f, t) \approx (h_D \circledast \Phi \circledast h_D \circledast M)(\mathbf{x}_f, t). \quad \text{(A.3)}$$

Using the cascade property of convolution, we further define an effective two-way scattering kernel

$$h_{2D}(\mathbf{x}, t) \approx h_D(\mathbf{x}, t) \circledast h_D(\mathbf{x}, t), \quad \text{(A.4)}$$

so that the forward model becomes

$$R_{\text{model}}(\mathbf{x}_f, t) \approx (h_{2D} \circledast \Phi \circledast M)(\mathbf{x}_f, t). \quad \text{(A.5)}$$

In the simplest case where the surface blur is negligible, Φ reduces to a delta kernel and the model simplifies to

$$R_{\text{model}}(\mathbf{x}_f, t) \approx (h_{2D} \circledast M)(\mathbf{x}_f, t). \quad \text{(A.6)}$$

In this formulation, the overall spatio-temporal PSF of the confocal imaging-through-scattering system is given by the effective two-way kernel h_{2D}.

Experimental results. In practice, we estimate the effective two-way PSF h_{2D} by constructing a synthetic scene with a scattering medium of depth $2D$ and a point light source placed immediately behind the volume, and then rendering the detector's time-resolved impulse response with a transient path tracer. Under the forward model in Eq. (A.5), recovering the effective hidden planar response from the measured transient light transport R_{meas} reduces to a spatio-temporal deconvolution problem. Specifically, letting $M_{\text{eff}} \triangleq \Phi * M$, we solve for M_{eff} using Wiener filtering [Wiener 1949]:

$$\widehat{M}_{\text{eff}} = \mathcal{F}^{-1}\left[\frac{\mathcal{F}\{R_{\text{meas}}\}\ \mathcal{F}\{h_{2D}\}^*}{|\mathcal{F}\{h_{2D}\}|^2 + 1/\lambda}\right], \quad \text{(A.7)}$$

where $\mathcal{F}$ denotes the Fourier transform. In the simplest case where the surface blur is negligible, Φ reduces to a delta kernel and $\widehat{M}_{\text{eff}}$ directly approximates the hidden planar pattern M.

We conduct experiments under a confocal setup similar to Du et al. [2022]. We prepare a turbid medium by mixing 8 L of water with approximately 1.5 mL of 20% lipid emulsion (Intralipid) into a 15×15 cm tank of height 20 cm, and attach a black cardboard sheet striping with retro-reflective tape onto the back wall of the tank. We only gate photons originating from internal system reflections, excluding those reflected by the scene.

Fig. 4 presents two experiments with planar targets displaying the letters S and G. For each case, we compare the integrated image, a time-gated reconstruction, and our method. While integration and time gating suffer from strong backscatter and reduced contrast, our approach recovers the underlying sharp letter patterns.

B More Ablation Studies

We provide additional ablation studies to further analyze the effect of measurement diversity. Specifically, we evaluate the influence of the number of illumination probes, the number of views, and the stability of the foundation model prompt under a specular geometry.

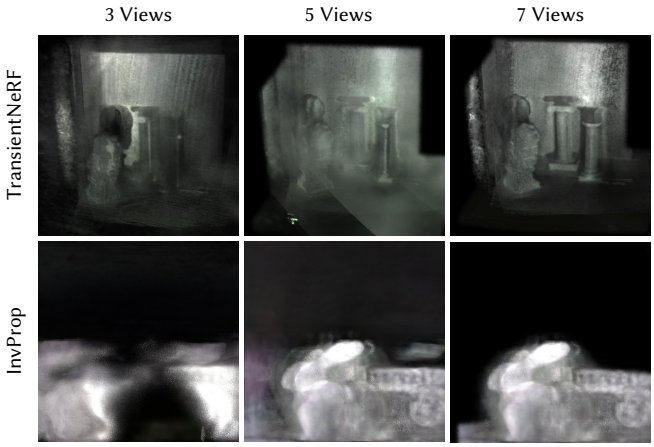

Fig. 6. **Ablation on the number of input views.** We compare reconstruction results under 3, 5, and 7 input views. Increasing the number of views provides stronger geometric constraints and improves reconstruction quality under sparse-view settings.

B.1 Effect of the Number of Illumination Probes

We first study how the number of decoupled illumination probes affects the reconstruction quality of our method. We vary only the number of illumination positions used for inverse transient rendering. As Fig. 5 shows, adding more illumination probes improves the conditioning of the inverse problem compared with using a single illumination. If we only apply front illumination, the metal sphere and the gourd looks rougher than their actual appearances. Adding more illuminations improves the material reconstruction.

B.2 Effect of the Number of Views

We further show how the number of input views affects transient reconstruction baselines under different view configurations. As Fig. 6 shows, increasing the number of views provides stronger geometric constraints and reduces ambiguities caused by self-occlusion.

C Mesh Smoothness and Geometric Fairness Metrics

C.1 Laplacian Smoothness Energy

Laplacian smoothing minimizes the squared magnitude of the discrete Laplacian operator [Meyer et al. 2003]. The Laplacian smoothness energy measures how much a surface deviates from a locally harmonic shape. Using the uniform graph Laplacian, the discretized Laplacian coordinate of vertex i is defined as

$$(LV)_i = v_i - \frac{1}{d_i} \sum_{j \in \mathcal{N}(i)} v_j$$

where d_i is the vertex degree. The Laplacian smoothness energy is then defined as

$$E_\delta = \frac{1}{N} \sum_{i=1}^{N} \|(LV)_i\|^2$$

where N is the number of vertices.

This formulation corresponds to the special case of Laplacian surface editing where the target Laplacian coordinates are set to zero, enforcing maximal local smoothness. The squared Laplacian energy is a smoothness prior in geometry processing [Stein et al. 2018]. This makes E_δ a convenient mesh-quality indicator since it directly penalizes large discrete Laplacians. So lower values correspond to smoother, more locally harmonic surfaces under consistent normalization.

References

Dongyu Du, Xin Jin, Rujia Deng, Jinshi Kang, Hongkun Cao, Yihui Fan, Zhiheng Li, Haoqian Wang, Xiangyang Ji, and Jingyan Song. 2022. A Boundary Migration Model for Imaging within Volumetric Scattering Media. *Nature Communications* 13, 1 (June 2022), 3234.

David B Lindell and Gordon Wetzstein. 2020. Three-dimensional imaging through scattering media based on confocal diffuse tomography. *Nature Communications* 11, 4517 (2020), 1–13.

David B Lindell, Gordon Wetzstein, and Matthew O'Toole. 2019. Wave-based non-line-of-sight imaging using fast fk migration. *ACM Transactions on Graphics* 38, 4 (2019), 1–13.

Aurora Maccarone, Francesco Mattioli Della Rocca, Aongus McCarthy, Robert Henderson, and Gerald S. Buller. 2019. Three-Dimensional Imaging of Stationary and Moving Targets in Turbid Underwater Environments Using a Single-Photon Detector Array. 27, 20 (2019), 28437.

Aurora Maccarone, Aongus McCarthy, Ximing Ren, Ryan E. Warburton, Andy M. Wallace, James Moffat, Yvan Petillot, and Gerald S. Buller. 2015. Underwater Depth Imaging Using Time-Correlated Single-Photon Counting. 23, 26 (2015), 33911.

Mark Meyer, Mathieu Desbrun, Peter Schröder, and Alan H. Barr. 2003. Discrete Differential-Geometry Operators for Triangulated 2-Manifolds. In *Visualization and Mathematics III*, Hans-Christian Hege and Konrad Polthier (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 35–57.

Michael S. Patterson, B. Chance, and B. C. Wilson. 1989. Time Resolved Reflectance and Transmittance for the Noninvasive Measurement of Tissue Optical Properties. 28, 12 (1989), 2331.

Oded Stein, Eitan Grinspun, Max Wardetzky, and Alec Jacobson. 2018. Natural Boundary Conditions for Smoothing in Geometry Processing. *ACM Trans. Graph.* 37, 2, Article 23 (May 2018), 13 pages.

L. Wang, P. P. Ho, C. Liu, G. Zhang, and R. R. Alfano. 1991. Ballistic 2-D Imaging Through Scattering Walls Using an Ultrafast Optical Kerr Gate. 253, 5021 (1991), 769–771.

Norbert Wiener. 1949. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications.* The MIT Press.