Non-Line-of-Sight Imaging using Phasor Field Virtual Wave Optics

Xiaochun Liu[1], Ibón Guillén[2], Marco La Manna[3], Ji Hyun Nam[1], Syed Azer Reza[3], Toan Huu Le[1], Adrian Jarabo[2], Diego Gutierrez[2], Andreas Velten[1,3,4]

[1]*Department of Electrical and Computer Engineering, University of Wisconsin Madison*
[2]*Graphics and Imaging Lab, Universidad de Zaragoza - I3A*
[3]*Department of Biostatistics and Medical Informatics, University of Wisconsin Madison*
[4]*Corresponding author: velten@wisc.edu*

Non-line-of-sight (NLOS) imaging allows to observe objects partially or fully occluded from direct view, by analyzing indirect diffuse reflections off a secondary relay surface. Despite its many potential applications[1–9], existing methods lack practical usability due to several shared limitations, including the assumption of single scattering only, lack of occlusions, and Lambertian reflectance. Line-of-sight (LOS) imaging systems, on the other hand, can address these and other imaging challenges despite relying on the mathematically simple processes of linear diffractive wave propagation. In this work we show that the NLOS imaging problem can also be formulated as a diffractive wave propagation problem. This allows to image NLOS scenes from raw time-of-flight data by applying the mathematical operators that model wave propagation inside a conventional line-of-sight imaging system. By doing this, we have developed a method that yields a new class of imaging algorithms mimicking the various capabilities of LOS cameras. To demonstrate our method, we derive three imaging algorithms, each with its own unique novel capabilities, modeled after three different LOS imaging systems. These algorithms rely on solving a wave diffraction integral, namely the Rayleigh-Sommerfeld Diffraction (RSD) integral. Fast solutions to RSD and its approximations are readily available, directly benefiting our method. We demonstrate, for the first time, NLOS imaging of complex scenes with strong multiple scattering and ambient light, arbitrary materials, large depth range, and occlusions. Our method handles these challenging cases without explicitly developing a light transport model. We believe that our approach will help unlock the potential of NLOS imaging, and the development of novel applications not restricted to laboratory conditions, as shown in our results.

We have recently witnessed large advances in transient imaging techniques[10], employing streak cameras[11], gated sensors[6], amplitude-modulated continuous waves[12], single-photon detectors (SPAD)[13], or interferometry[14]. Access to time-resolved image information has in turn led to advances in imaging of objects partially or fully hidden from direct view[1–3,5–7,15–18] (NLOS imaging). Other methods are able to use information encoded in the phase of continuous light and do not use time of flight[4]. In the basic configuration of an NLOS system, light bounces off a relay wall, travels to the hidden scene, then propagates back to the relay wall, and finally reaches the sensor.

Recent NLOS reconstruction methods are based on heuristic filtered backprojection[2,3,6,7,21], or attempt to compute inverse operators of simplified forward light transport models[5,9,19]. These simplified models do not take into account multiple scattering, surfaces with anisotropic reflectance or, with a few exceptions[19], occlusions and clutter. Moreover, the depth range that can be recovered is also limited, partially due to the difference in intensity between first- and higher-order reflections. Existing methods are thus limited to carefully controlled cases, imaging isolated objects of simple geometry with moderate or no occlusion. Moreover, while the goal of previous works is limited to the reconstruction of hidden geometry, we develop a novel theoretical framework for general NLOS imaging, reconstructing the irradiance at a virtual sensor; this enables applications beyond geometry reconstruction, as we demonstrate in this paper. Our data and reconstruction code can be found in a figshare repository[20].

Time-of-flight LOS imaging has used a phasor formalism together with Fourier domain ranging[12] to describe the emitted modulated light signal. Kadambi et al.[22] extended this concept to reconstruct NLOS scenes using phasors to describe hardware intensity modulation. We show that a similar description can be used to model the physics of light transport through the scene. The key insight of our method is that propagation through a scene of intensity-modulated light can be modeled using a Rayleigh-Sommerfeld diffraction (RSD) operator acting on a quantity we term the *phasor field*. This allows us to formulate any NLOS imaging problem as a wave imaging problem (Figure 1), and to transfer well-established insights and techniques from classic optics into the NLOS domain. Given a captured time-resolved dataset of light transport through a NLOS scene, and a choice of a *template* LOS imaging system, our method provides a recipe that results in a NLOS imaging algorithm mimicking the capabilities of the corresponding LOS system. This template system can be any real or hypothetical wave imaging system that includes a set of light sources and detectors. The resulting algorithms can then be efficiently solved using diffraction integrals like the RSD, for which a variety of fast exact and approximate solvers exist[24]. Section A in the supplemental illustrates this.

We start by mathematically defining our phasor field $\mathcal{P}(\mathbf{x}, t)$. Let $\mathcal{E}(\mathbf{x}, t)$ $[\sqrt{\mathrm{Wm}^{-2}}]$ be a quasi-monochromatic scalar field at position $\mathbf{x} \in \mathcal{S}$ and time t, incident on (or reflected from) a Lambertian surface $\mathcal{S}$, with center frequency Ω_0 and bandwidth $\Delta\Omega \ll \Omega_0$. We can then define

$$\mathcal{P}(\mathbf{x}, t) \equiv \left\langle \frac{1}{\tau} \int_{t-\tau/2}^{t+\tau/2} \left|\mathcal{E}(\mathbf{x}, t')\right|^2 \, dt' \right\rangle - \left\langle \frac{1}{T} \int_{t-T/2}^{t+T/2} \left|\mathcal{E}(\mathbf{x}, t')\right|^2 \, dt' \right\rangle \tag{1}$$

as the mean subtracted irradiance [Wm^{-2}] at point $\mathbf{x}$ and time t. The $\langle\cdot\rangle$ operator denotes spatial speckle averaging (for the reflected case) accounting for laser illumination, and τ represents the averaging of the intensity at a fast detector, with $\tau \ll 1/\Delta\Omega \ll T$. The second integral in the equation above is a long-term average intensity over an interval $T \gg \tau$ of the signal as seen by a conventional non-transient photodetector. Now, let us define the Fourier component of $\mathcal{P}(\mathbf{x}, t)$ for

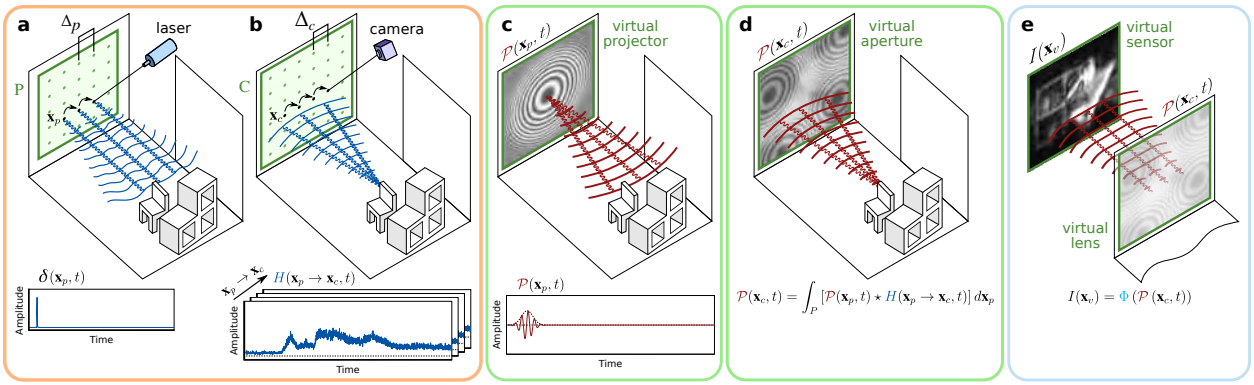

Figure 1: **NLOS as a virtual LOS imaging system.** Capturing scene data: **a**, A pulsed laser sequentially scans a relay wall; **b**, the light reflected back from the scene onto the wall is recorded at the sensor yielding an impulse response H of the scene. **c**, Virtual light source: The phasor field wave of a virtual light source $\mathcal{P}(\mathbf{x}_p, t)$ is modeled after the wavefront of the light source of the template LOS system. **d**, The scene response to this virtual illumination $\mathcal{P}(\mathbf{x}_c, t)$ is computed using H. **e**, The scene is reconstructed from the wavefront $\mathcal{P}(\mathbf{x}_c, t)$ using wave diffraction theory. The function $\Phi(\cdot)$ is also taken from the template LOS system.

frequency ω as

$$\mathcal{P}_{0,\omega}(\mathbf{x}) \equiv \int_{-\infty}^{+\infty} \mathcal{P}(\mathbf{x}, t)\, e^{-i\omega t} dt, \tag{2}$$

from which we can define a monochromatic component of the phasor field $\mathcal{P}_\omega(\mathbf{x}, t)$ as

$$\mathcal{P}_\omega(\mathbf{x}, t) \equiv \mathcal{P}_{0,\omega}(\mathbf{x})\, e^{i\omega t}. \tag{3}$$

Using the above, our phasor field $\mathcal{P}(\mathbf{x}, t)$ can be expressed as a superposition of monochromatic plane waves as $\mathcal{P}(\mathbf{x}, t) = \int_{-\infty}^{+\infty} \mathcal{P}_\omega(\mathbf{x}, t)\, d\omega/2\pi$. Since $\mathcal{P}(\mathbf{x}, t)$ is a real quantity, the Fourier components $\mathcal{P}_{0,\omega}(\mathbf{x})$ are complex and symmetric about $\omega = 0$. Note that in many places of this manuscript we assign $\mathcal{P}(\mathbf{x}, t)$ an explicitly complex value; in these cases it is implied that the correct real representation is $\frac{1}{2}(\mathcal{P}(\mathbf{x}, t) + \mathcal{P}^*(\mathbf{x}, t))$. In practice the complex conjugate can be safely ignored in our calculations. As can be seen in Section B in the supplemental, given an isotropic source plane S and a destination plane D, and assuming that the electric field at S is incoherent, the propagation of its monochromatic component $\mathcal{P}_\omega(\mathbf{x}, t)$ is defined by an RSD-like propagation integral:

$$\mathcal{P}_\omega(\mathbf{x}_d, t) = \gamma \int_S \mathcal{P}_\omega(\mathbf{x}_s, t) \frac{e^{ik|\mathbf{x}_d - \mathbf{x}_s|}}{|\mathbf{x}_d - \mathbf{x}_s|} d\mathbf{x}_s, \tag{4}$$

where γ is an attenuation factor, and $k = 2\pi/\lambda$ is the wave number for wavelength $\lambda = 2\pi/\omega$, $\mathbf{x}_s \in S$ and $\mathbf{x}_d \in D$. Note that, as described in Section B in the supplemental, we approximate γ as a constant over the plane S as $\gamma \approx 1/|\langle S \rangle - \mathbf{x}_d|$; this approximation has a minor effect on the signal

amplitude at the sensor, but does not change the phase of our phasor field. While Equation 4 is defined for monochromatic signals, it can be used to propagate broadband signals by propagating each monochromatic component independently; this can be efficiently done by time-shifting the phasor field (more details are provided in Section B.1 of the supplemental).

The key insight of Equation 4 is that, given the assumption of constant γ, the propagation of our phasor field is defined by the same RSD operator as any other physical wave. Therefore, in order to image a scene from a *virtual camera* with aperture at plane C, we can apply the image formation model of any wave-based LOS imaging system directly over the phasor field $\mathcal{P}(\mathbf{x}_c, t)$ at the aperture, with $\mathbf{x}_c \in C$. The challenge is how to compute $\mathcal{P}(\mathbf{x}_c, t)$ from an illuminating input phasor field $\mathcal{P}(\mathbf{x}_p, t)$, where $\mathbf{x}_p$ is a point in the *virtual projector* aperture P, given a particular NLOS scene (see Figure 1).

Since light transport is linear in space and time-invariant[23,25], we can characterize light transport through the scene as an impulse response function $H(\mathbf{x}_p \rightarrow \mathbf{x}_c, t)$, where $\mathbf{x}_p$ and $\mathbf{x}_c$ are the positions of the emitter and detector, respectively. The phasor field at the virtual aperture $\mathcal{P}(\mathbf{x}_c, t)$ can thus be expressed as a function of the input phasor field $\mathcal{P}(\mathbf{x}_p, t)$ and $H(x_p \rightarrow \mathbf{x}_c, t)$ as

$$\mathcal{P}(\mathbf{x}_c, t) = \int_P \left[\mathcal{P}(\mathbf{x}_p, t) \star H(\mathbf{x}_p \rightarrow \mathbf{x}_c, t)\right] d\mathbf{x}_p, \tag{5}$$

where $\star$ denotes the convolution operator. Any imaging system can be characterized by its image formation function $\Phi(\cdot)$, which transduces the incoming field into an image

$$I(\mathbf{x}_v) = \Phi\left(\mathcal{P}\left(\mathbf{x}_c, t\right)\right), \tag{6}$$

where $\mathbf{x}_v$ is the point being imaged (i.e., the point at the *virtual sensor*). This in turn can be formulated as an RSD propagator, requiring to solve a diffraction integral in order to generate the final image.

In an NLOS scenario, $H(\mathbf{x}_p \rightarrow \mathbf{x}_c, t)$ usually corresponds to 5D transients acquired via an ultrafast sensor focused on $\mathbf{x}_c$, and sequentially illuminating the relay wall with short pulses at different points $\mathbf{x}_p$ (see Figure 1, and the Methods Section). Points $\mathbf{x}_p$ and $\mathbf{x}_c$ correspond to a virtual LOS imaging system projected onto the relay wall. Once $H(\mathbf{x}_p \rightarrow \mathbf{x}_c, t)$ has been captured, both the wavefront $\mathcal{P}(\mathbf{x}_p, t)$ and the imaging operator $\Phi(\cdot)$ can be implemented computationally, so they are not bounded by hardware limitations. We can leverage this to employ different $\mathcal{P}(\mathbf{x}_p, t)$ functions from any existing LOS imaging system[26] to emulate its characteristics in an NLOS setting.

We illustrate the robustness and versatility of our method by implementing three virtual NLOS imaging systems based on common LOS techniques: a conventional photography camera capable of imaging NLOS scenes without knowledge of the timing or location of the illumination source; a transient photography system capable of capturing transient videos of the hidden scene revealing higher-order interreflections beyond 3rd bounce; and a confocal time-gated imaging system robust

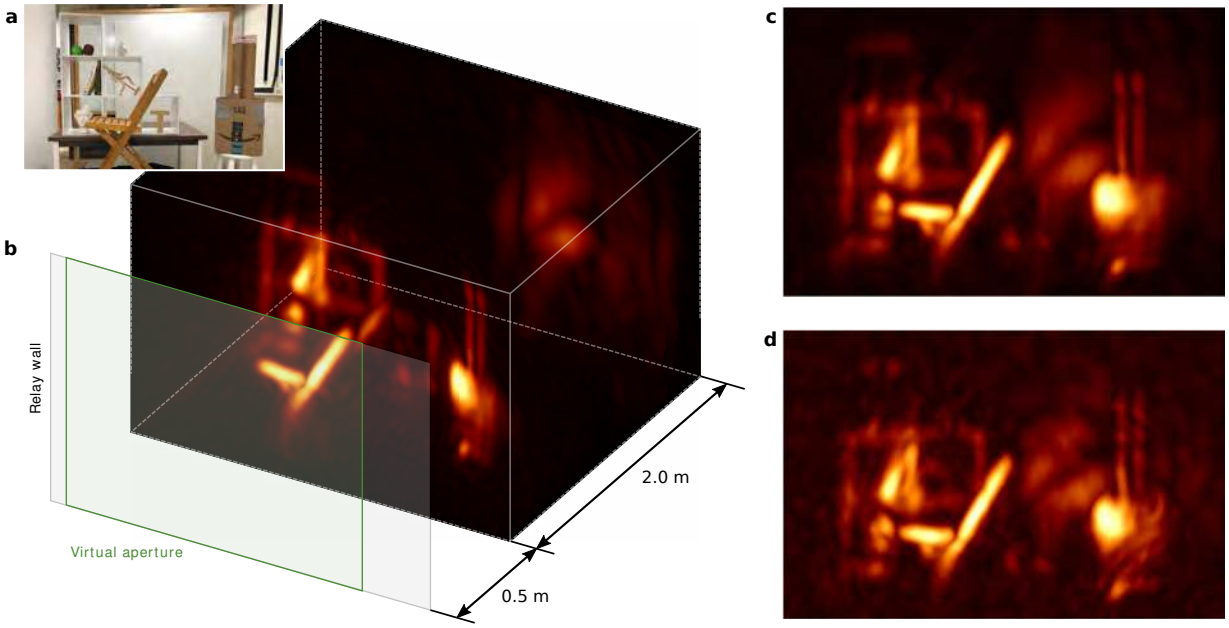

Figure 2: **Reconstructions of a complex NLOS scene. a**, Photograph of the scene as seen from the relay wall. The scene contains occluding geometries, multiple anisotropic surface reflectances, large depth, and strong ambient and multiply scattered light. **b**, 3D visualization of the reconstruction using phasor fields ($\lambda = 6$ cm). We include the relay wall location and the coverage of the virtual aperture for illustration purposes. **c**, Frontal view of the scene, captured with an exposure time of 10 ms per laser position. **d**, Frontal view captured with just a 1 ms exposure time (24 seconds for the complete scan).

to interreflections. An in-depth description of these example imaging systems is provided in Section C of the supplemental, including their corresponding $\mathcal{P}(\mathbf{x}_p, t)$ functions and imaging operators $\Phi(\cdot)$, while Section D describes some practical solver examples.

The spatial resolution of our virtual camera is $\Delta_x = 0.61\lambda L/d$, where d is the virtual aperture diameter, and L is the imaging distance. The distance Δ_p between sample points $\mathbf{x}_p$ in P (see Figure 1) has to be small enough to sample H at the phasor field wavelength. We fix $\Delta_p = 1$ cm, and unless stated otherwise $\lambda = 4$ cm. The minimum sampling rate is $\Delta_p < \lambda/2$; in practice we found $\Delta_p = \lambda/4$ to provide the best trade-off between reconstruction noise and resolution.

The computational cost of our algorithm is bounded by the RSD solver computing the image formation model $\Phi(\cdot)$. Fast diffraction integral solvers exist[24], with complexity $O(N^3 \log(N))$. For the particular case of our confocal system, we formulate the algorithm as a backprojection (see Section D.2 of the supplemental for details), so we are bounded by the computational cost of the backprojection algorithm used.

One common application of NLOS imaging is the reconstruction of hidden geometry. Figure 2

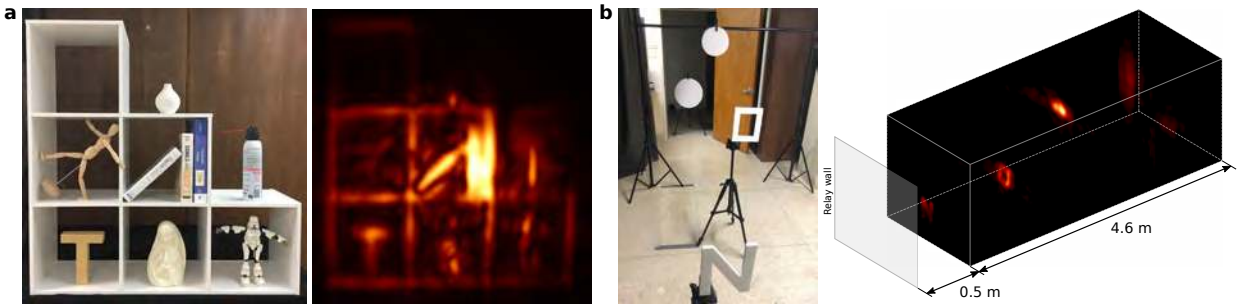

Figure 3: **Robustness of our technique.** **a**, Reconstruction in the presence of strong ambient illumination (all the lights on during capture). **b**, Hidden scene with a large depth range, leading to very weak signals from objects farther away.

shows the result on a complex scene with our virtual confocal camera. The scene contains multiple objects with occlusions distributed over a large depth, a wide range of surface reflectances and albedos, and strong interreflections. Despite this challenging scenario, our method is able to image many details of the scene, at the correct depths, even with an ultra-short, 1 ms exposure. More analysis on the robustness of our method to capture noise can be found in the Methods Section. For simpler scenes (no occlusions, limited depth, controlled reflectance, and no interreflections), our method yields results on par with current techniques, which already approach theoretical limits regarding reconstruction quality (see the Methods Section).

In Figure 3 we demonstrate the robustness of our method when dealing with other challenging scenarios, including strong multiple scattering and ambient illumination (3.a), or a high dynamic range from objects spanning very large depths range (3.b). Last, our method also allows to implement novel NLOS imaging systems and applications, leveraging the wealth of tools and processing methods available in LOS imaging. Figure 4 (top) demonstrates NLOS refocusing using our virtual photography camera, computed with both the exact RSD operator and using a faster Fresnel approximation, while the bottom row shows frames of NLOS femto-photography reconstructed using our virtual transient photography system, revealing 4th and 5th bounce components in the scene. The first, second and fourth frames, in green, show how light first illuminates the chair, then propagates to the shelf, and finally hits the back wall three meters away. The frames in orange show higher order bounces. The third frame shows the chair being illuminated *again* by light bouncing back from the relay wall, while the last two frames show how the pulse of light travels from the wall back to the scene (see the supplementary video). A description of the Fresnel approximation to the RSD operator, as well as the LOS projector-camera functions used in these examples, appear in the supplemental (Sections D.1 and C.2).

In the Methods Section, we include comparisons against ground truth for two synthetic scenes, inside a corridor of 2 m x 2 m x 3 m to create interreflections, simulated using a open source transient renderer[27]; these scenes are included in a publicly available database[28]. We analyze the robustness

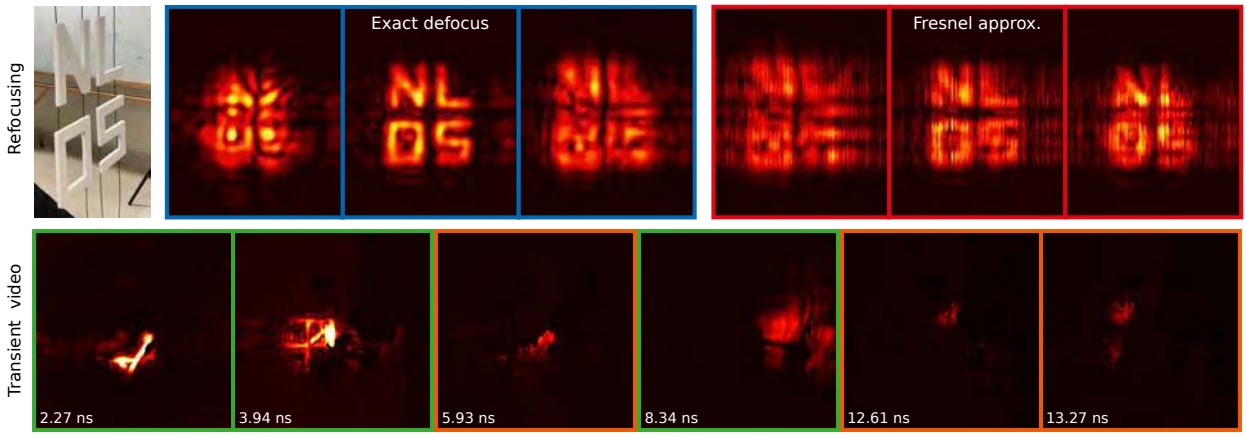

Figure 4: **Additional novel NLOS imaging applications of our method. (Top)** NLOS refocusing: The hidden letters (left) are progressively brought in and out of focus as seen from a virtual photography camera at the relay wall, using the exact lens integral (blue), and the faster Fresnel approximation (red). **(Bottom)** NLOS transient video: Example frames of light traveling through a hidden office scene when illuminated by a pulsed laser. Timestamps indicate the propagation time from the relay wall. Frames with a green border show third bounce objects, frames on orange show 4th and 5th bounce effects.

of our method with and without such interreflections; the reconstruction mean square error does not increase, remaining below 5 mm. Last, we progressively vary the specularity of the hidden geometry, from purely Lambertian to highly specular; again, the quality of the reconstructions does not vary significantly (mean square error of about 2 mm).

The examples shown highlight the primary benefit of our approach: By turning NLOS into a virtual LOS system, the intrinsic limitations of previous approaches no longer apply, enabling a new class of NLOS imaging methods that leverage existing wave-based imaging methods. Formulating NLOS light propagation as a wave does not impose limitations on the types of problems that can be addressed, nor the datasets that can be used. Any signal can be represented as a superposition of phasor field waves; our formulation thus can be viewed as a choice of basis to represent any kind of NLOS data. Expressing the NLOS problem this way allows to create a direct analogy to LOS imaging, which can be exploited to derive suitable imaging algorithms, and to implement them efficiently.

We have shown three novel imaging algorithms derived from our method. Our results include significantly more complex scenes than any NLOS reconstruction shown so far in the literature, as well as novel applications. In addition, our approach is flexible, fast, memory-efficient, and lacks computational complexity since it does not require inverting a light transport model. In the future, we anticipate its application to other LOS imaging systems to, for instance, separate light transport into direct and global components, or utilize the phase of $\mathcal{P}_\omega$ for enhanced depth resolution. Our

virtual imaging system could also be used to create a *second* virtual imaging system to see around two corners, assuming the presence of a secondary relay Lambertian surface in the hidden scene, or to select and manipulate individual light paths to isolate specific aspects of the light transport in different NLOS scenes. In that context, combining our theory with light transport inversions, via, e.g., an iterative approach, could potentially lead to better results, and is an interesting avenue of future work.

Acknowledgements

This work was funded by DARPA through the DARPA REVEAL project (HR0011-16-C-0025), the NASA Innovative Advanced Concepts (NIAC) program (NNX15AQ29G), the Air Force Office of Scientific Research (AFOSR) Young Investigator Program (FA9550-15-1-0208), the Office of Naval Research (ONR, N00014-15-1-2652), the European Research Council (ERC) under the EU's Horizon 2020 research and innovation programme (project CHAMELEON, grant No 682080), the Spanish Ministerio de Economía y Competitividad (project TIN2016-78753-P), and the BBVA Foundation (Leonardo Grant for Researchers and Cultural Creators). We would like to acknowledge Jeremy Teichman for helpful insights and discussions in developing the phasor field model. We would also like to acknowledge Mauro Buttafava, Alberto Tosi, and Atul Ingle for help with the gated SPAD detector, and Belen Masia, Sandra Malpica and Miguel Galindo for careful reading of the manuscript.

Author Contributions

X. L., S. A. R., M. L. M., and A. V. conceived the method; X. L., I. G., M. L. M., and J. H. N. implemented the reconstruction; M. L. M., X. L., J. H. N., and, T. H. L. built and calibrated the system; I. G., D. G., and A. J., developed the simulation system; A. J., D. G., and A. V. coordinated the project; All authors contributed to writing the paper.

Author Information

Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the paper. Correspondence and requests for materials should be addressed to the corresponding author Andreas Velten (velten@wisc.edu).

Data Availability

The measured data and the Phasor Field NLOS code supporting the findings of this study are available in the figshare repository https://doi.org/10.6084/m9.figshare.8084987. Additional data and code are available from the corresponding authors upon request.

References

1. Kirmani, A., Hutchison, T., Davis, J., & Raskar, R. Looking around the corner using ultrafast transient imaging. *International Journal of Computer Vision* **95**, 13-28 (2011).

2. Gupta, O., Willwacher, T., Velten, A., Veeraraghavan, A., & Raskar, R. Reconstruction of hidden 3D shapes using diffuse reflections. *Optics Express* **20**, 19096-19108 (2012).

3. Velten, A. et al. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature Communications* **3**, 745 (2012).

4. Katz, O., Small, E. & Silberberg, Y. Looking around corners and through thin turbid layers in real time with scattered incoherent light. *Nature Photonics* **6**, 549-553 (2012).

5. Heide, F., Xiao, L., Heidrich, W., & Hullin, M. B. Diffuse mirrors: 3D reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors. *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), 3222-3229 (2014).

6. Laurenzis, M., Velten, A. Nonline-of-sight laser gated viewing of scattered photons. *Optical Engineering* **53**, 023102 (2014).

7. Buttafava, M., Zeman, J., Tosi, A., Eliceiri, K. & Velten, A. Non-line-of-sight imaging using a time-gated single photon avalanche diode. *Optics Express* **23**, 2099721011 (2015).

8. Arellano, V., Gutierrez, D., & Jarabo, A.. Fast back-projection for non-line of sight reconstruction. *Optics Express* **25**, 11574-11583 (2017).

9. O'Toole, M., Lindell, D. B., & Wetzstein, G. Confocal non-line-of-sight imaging based on the light-cone transform. *Nature* **555**, 338 (2018).

10. Jarabo, A., Masia, B., Marco, J., & Gutierrez, D. Recent advances in transient imaging: A computer graphics and vision perspective. *Visual Informatics* **1**, 65-79 (2017).

11. Velten, A. et al. Femto-photography: capturing and visualizing the propagation of light. *ACM Transactions on Graphics* **32**, 44 (2013).

12. Gupta, M., Nayar, S. K., Hullin, M. B., & Martin, J. Phasor imaging: A generalization of correlation-based time-of-flight imaging. *ACM Transactions on Graphics* **34**, 5 (2015).

13. O'Toole, M., Heide, F., Lindell, D., Zang, K., Diamond, S., & Wetzstein, G. Reconstructing Transient Images from Single-Photon Sensors. *2017 IEEE International Conference on Computational Photography* (IEEE, 2017).

14. Gkioulekas, I., Levin, A., Durand, F. & Zickler, T. Micron-scale light transport decomposition using interferometry. *ACM Transactions on Graphics* **34**, 37 (2015).

15. Xin, S. et al. A Theory of Fermat Paths for Non-Line-of-Sight Shape Reconstruction, *In IEEE Intl. Conf. Computer Vision and Pattern Recognition* (CVPR), 2019.

16. Tsai, C., Sankaranarayanan, A. & Gkioulekas, I. Beyond Volumetric Albedo A Surface Optimization Framework for Non-Line-of-Sight Imaging, *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), 2019.

17. Liu, X., Bauer, S., & Velten, A. Analysis of feature visibility in non-line-of-sight measurements. *In IEEE Intl. Conf. Computer Vision and Pattern Recognition* (CVPR), 2019.

18. Wu, R. et al. Adaptive Polarization-Difference Transient Imaging for Depth Estimation in Scattering Media. *Optics Letters* **43**, 6 (2018).

19. Heide, F., O'Toole, M., Zang, K., Lindell, D., Diamond, S., & Wetzstein, G. Non-line-of-sight Imaging with Partial Occluders and Surface Normals. *ACM Transactions on Graphics* (2019).

20. Liu, Xiaochun; Guillen, Ibon; Manna, Marco La; Nam, Ji Hyun; Reza, Syed Azer; Le, Toan Huu; et al.; Datasets and reconstruction code for a virtual wave non-line-of-site imaging approach. figshare. https://doi.org/10.6084/m9.figshare.8084987 (2019)

21. Laurenzis, M., & Velten, A. Feature selection and back-projection algorithms for nonline-of-sight laser-gated viewing. *Journal of Electronic Imaging* **23**, 6 (2014).

22. Kadambi, A., Zhao, H., Shi, B. & Raskar, R. Occluded imaging with time-of-flight sensors. *ACM Transactions on Graphics* **35**, 2 (2016).

23. Sen, P., Chen, B., Garg, G., Marschner, S. R., Horowitz, M., Levoy, M., & Lensch, H. Dual photography. *ACM Transactions on Graphics* **24**, 3. (2005)

24. Shen, F. & Wang, A. Fast-fourier-transform based numerical integration method for the Rayleigh-Sommerfeld diffraction formula. *Applied optics*, **45**, 6. (2006).

25. O'Toole, M. et al. Temporal frequency probing for 5D transient analysis of global light transport. *ACM Transactions on Graphics* **33**, 4. (2014)

26. Goodman, J. *Introduction to Fourier optics* 3rd edn (Roberts and Company Publishers, 2005).

27. Jarabo, A. et al. A framework for transient rendering. *ACM Transactions on Graphics* **33**, 6. (2014)

28. Galindo, M. et al. A Dataset for Benchmarking Time-Resolved Non-Line-of-Sight Imaging. *2019 IEEE International Conference on Computational Photography (Posters)* (IEEE, 2019).

Methods

Details on Data Acquisition

Hardware configuration: Our capture system, shown in the extended data Fig. 1, consists of a Onefive Katana HP amplified diode laser (1 W at 532 nm, and a pulse width of about 35 ps used at a repetition rate of 10 MHz), and a gated SPAD detector processed by a Time-Correlated Single Photon Counter (TCSPC, PicoQuandt HydraHarp), with a time resolution of about 30 ps and a dead time of 100 ns. Two additional CCD cameras are used to calibrate the laser's position. The measured time resolution of our system is approximately 65 ps, a combination of the pulse width of the laser and the time jitter of the system.

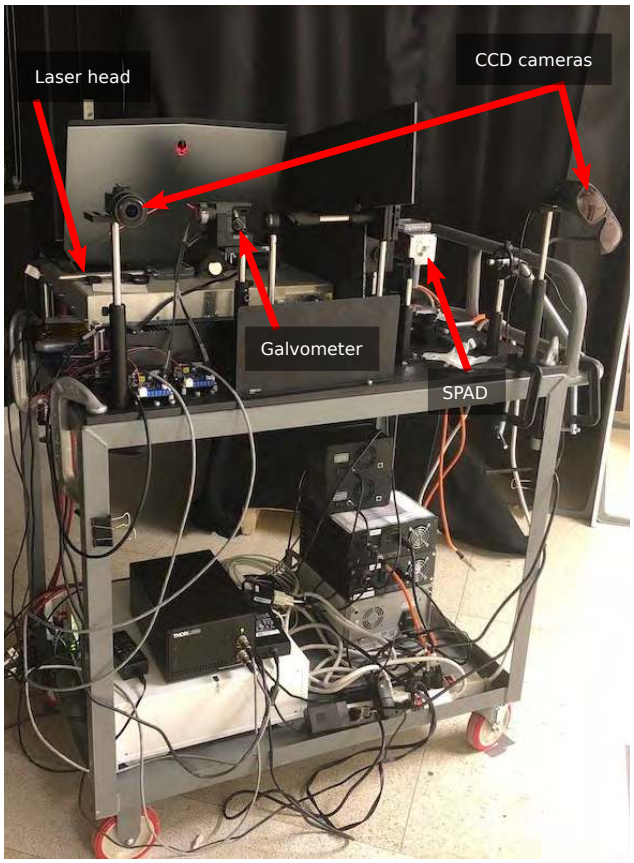

Figure M.1: *Capture hardware* used for the results shown in the paper.

NLOS measurement geometry: We obtain an impulse response function $H(\mathbf{x}_p \rightarrow \mathbf{x}_c, t)$ of the scene by sequentially illuminating points $\mathbf{x}_p$ on the relay wall with a short pulse, and detecting the

signal returning at points $\mathbf{x}_c$.

Our hardware device is located 2.5 m from the relay wall, with the NLOS scenes hidden from direct view. The FOV is 25 degrees. The walls are made of standard white styrofoam. The scanning area in the relay wall (virtual camera) is 1.8 m x 1.3 m, with laser points $\mathbf{x}_p$ spaced by $\Delta_p = 1$ cm in each direction. The SPAD detector is focused at a position near the center of the grid. We avoid scanning a small square region around the SPAD focused position (the confocal position) since the signal becomes very noisy at this location. Figures 2 and 3 in the main paper provide additional details for the specific scenes shown.

Exposure time: Our capture setup includes CCD cameras (extended data Fig. 1) to confirm the 3D position of every laser during the measurement; these are a limiting factor in the speed of our experiments. Since the capture process runs in parallel, we use a very long 1 s exposure time per laser position for our main datasets. They are used for all results unless otherwise specified. In addition we capture scenes without without the additional CCD photographs that can be collected much faster and with much shorter exposure times. In Figure 2 we show datasets of an office scene captured with exposure times of 1 to 10 ms per laser position, which results into a total capture time as low as 24 s. Further reconstructions of a shelf dataset are shown later as additional results showing that we can reduce exposure times at least down to 50 ms per data point without a significant loss in quality even with ambient light. This results in less than 20 minutes of total capture time. Note also that in our current prototype, we capture data sequentially with a single SPAD. Prototype SPAD arrays are currently under development, and it seems likely that a 16x16 array will be available by the end of the year. We thus expect to be able to capture 256 data points in less than 0.1 s in the near future.

Collected data: In total we use 10 experimental and two simulated datasets in the paper. All experimental datasets use a single SPAD location and 130 by 180 laser positions. The datasets and exposure times are:

- An Office Scene collected with 1 second exposure per laser position. This dataset is used to create the video shown in the supplementary video, frames of which are shown in Figure 4(bottom). A photograph and reconstruction of this scene is also shown in the supplementary video. The data is analyzed in extended data Fig.3 and extended data Table 1.
- An Office Scene collected with 10 ms, 5 ms, and 1 ms, used in Figures 2, extended data Fig.6, extended data Fig.7, and extended data Fig.8, and extended data Table 1.
- A scene of a bookshelf used in Figure 3(a) and in extended data Table 1.

	Total Photons	Photons/bin	Max. bin	Avg./ laser pos.
Large depth scene	3215722952	9.7	552	13742
NLOS letters	6502986696	19.6	2889	27791
Shelf	6158590767	18.6	2074	26319
Office Scene	6201680972	18.7	1406	26503
Office Scene 10 ms	48017499	0.14	18	2716
Office Scene 5 ms	24012257	0.072	15	1026
Office Scene 1 ms	4801568	0.014	6	205

Table 1: **Photon statistics for the captured data:** used in the paper and supplement. The first four scenes were captured with 1 s exposure time. The first column shows the total photons counted, the second the average photon count per time bin, the third is the maximum count over all time bins and the last contains the average number of photons collected in each laser position in the dataset.

- A scene of a bookshelf captured with various exposure times and ambient light conditions shown in extended data Fig.2 and extended data Fig.5.
- A scene with letters distributed over a large depth used in Figure 3(b) and extended data Table 1.
- A scene of the letters NLOS in a plane used in Figure 4(Top) and extended data Table 1.

To provide further insight into the noise and artifacts present in our data we go through an analysis of the raw data from our 1 second exposure office scene. We compare the maximum and average number of photons per second and laser position $\mathbf{x}_p$ for our captured scenes in the extended data Table 1. The dark count rate of our detector is 10 photons per second. We do not explicitly subtract dark counts nor ambient light or backgrounds.

Note that the high total photon numbers in the transient responses (extended data Table 1) are due to the long responses associated with the large depth and volume of the scenes, and not due to a particularly bright signal. Example data for a scene of a shelf is shown in extended data Fig.2 (whose reconstruction can be found on extended data Fig.5). In this scene, our longest 1 s exposure time peaks at about 150 photons/s (such peaks are probably due to the presence of specular surfaces), and the captured signal is extremely noisy. In comparison, the recent method by O'Toole et al.[9] acquires a brighter, cleaner signal in 0.1 s, peaking at about 600 photons/s, due to the use of retroreflective paint applied on the hidden objects (data from their `data_resolution_chart_40cm` dataset).

Let us further analyze the captured data. In extended data Fig.3(a) we show a visualization of our data matrix for the 1s exposure office scene using the function `imagesc` in Matlab, where each row is the data collected for a different location of the laser illumination spot, and each column

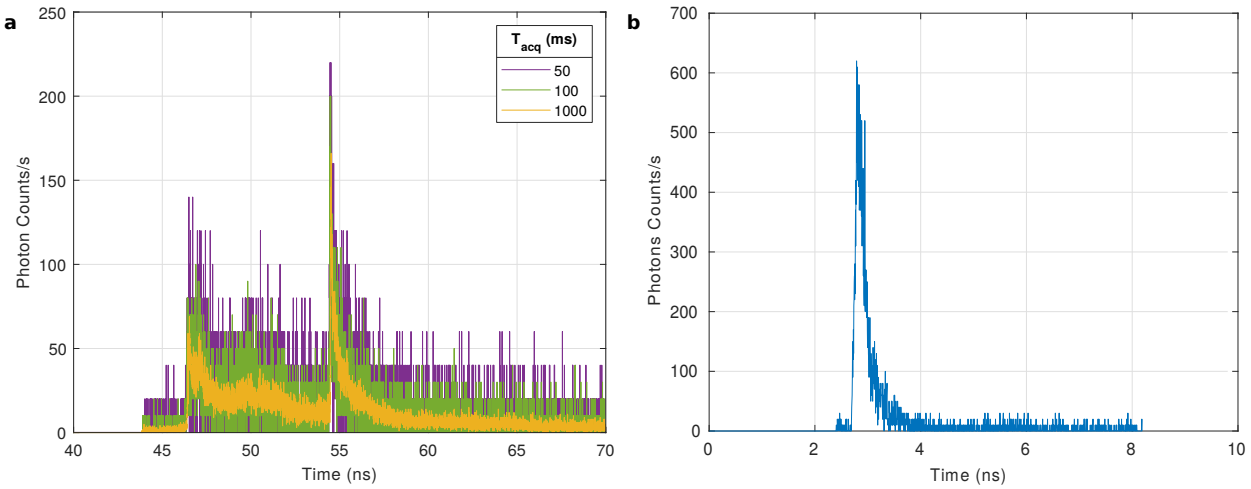

Figure M.2: **Data Comparison:** **a**, Raw data for one of the laser positions $\mathbf{x}_p$. Shown is the number of photons per second accumulated in each time bin (i.e. the collected histogram divided by the integration time in seconds). Time bins are 4 ps wide. As expected, all three curves appear to follow the same mean, but have a larger variance for lower exposure times. The raw data thus gets significantly noisier as exposure time decreases. The effects on the reconstruction, however, are rather minor as Extended Data Fig.4 shows. **b**, Example dataset from O'Toole et al.[9] for comparison.

contains a different time bin. The first time bin corresponds to the time when the illumination laser pulse leaves the relay wall. In the images we do not show time bins 10001 to 15000 as they are mostly empty due to the closing of the gate. As can be seen there are some sparse, very large peaks in the dataset that saturate the counting registers of our TCSPC ($2^{16} - 1$ counts). As we will see, these artifacts in the data are likely due to imperfections in the gating or optical setup.

Let us focus on the first instants of the captured data shown in extended data Fig.3(a), which reveal features that look like straight diagonal lines in the first few time bins. The fact that there are straight lines in this plot, indicates that they are likely related to a first-bounce signal, rather than the actual scene response. NLOS signals should show up as hyperbolas or sections of hyperbolas in this type of visualization and the curvature of the hyperbolas should be highest at the earliest time bins. The image contains many more features that look like straight lines that do not appear to have the correct hyperbolic curvature to be actual NLOS signals. Many of them also appear identically again in the other datasets which is another hint that they are probably not real NLOS data, but artifacts related to the measurement system. Our algorithm is completely agnostic to the presence of these artifacts. The brightest peaks also appear too early in the data to be associated with a NLOS object. To see this, consider that the closest object in any of our scenes is the chair in the office scene and it is more than 50 cm away from the wall. Consequently the first time response from an actual object cannot arrive at the SPAD earlier than 3.3 ns after the laser illuminates the

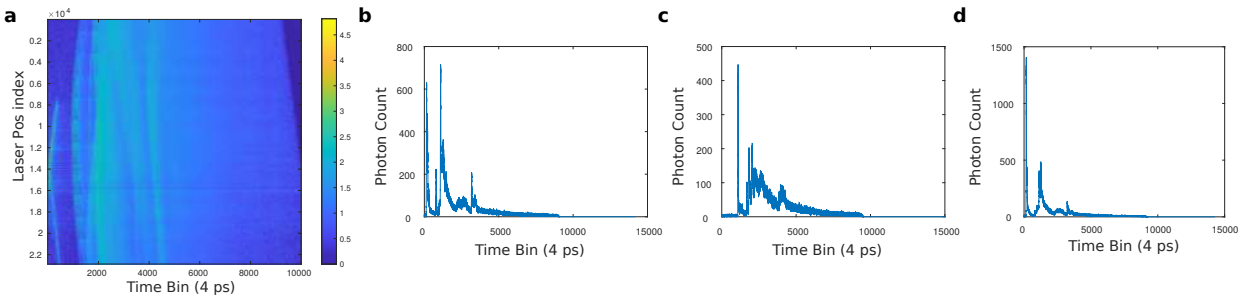

Figure M.3: **Visualization of the raw data** for our long exposure office scene. From left to right: **a**, Base 10 logarithm of the photon counts in all time bins. **b**, After removing the first 833 time bins in each dataset the plots show the photon counts for the laser position that received the largest total number of photons in the datset, **c**, the counts for the laser position that received the median number of photon counts, and **d**, the laser position that contains the time bin with the global maximum count in the entire set.

relay wall. Time bins are 4 ps wide. Any data before time bin 833 therefore can only be an artifact. We will speculate more about the origin of these artifacts later.

If we ignore those first 833 time bins that contain no useful data we obtain a dataset that can yield some meaningful statistics about the data. In this dataset the largest photon count in all our over 200 million time bins is smaller than 1400 photons. Note that as we show below, this 1400 maximum is likely still due to a gate artifact that just happened to occur slightly later than 3 ns into the dataset. Statistics for all datasets are shown in extended data Table 1.

Maximum photon counts usually come from the objects in the scene closest to the wall. Considering the large depth and specularities of our scenes, most of the reconstructed scene volume is actually using signals much weaker than the maximum signal as voxels are further away from the wall. Signals from a given surface are expected to drop in magnitude with one over distance to the power of four as distance is increased. An object generating 100 photon peaks at 50 cm distance in the front of our scene would therefore only create 100/8 photons if placed at at 1 m and 100/625 = 0.2 photons at 2.5 m towards the end of our office scene. This ability to handle scenes with large dynamic range in the data is another stated advantage of our algorithm.

In extended data Fig.3(b) we show a plot of the photon counts over time bins for the laser position that received the most total photons. We again see the extreme peak of $2^{16} - 1$ counts in the beginning of the dataset. Again, this peak can not possibly be a real third bounce signal as it would require for the pulse to travel between laser position and SPAD position significantly faster than the speed of light. The actual NLOS data starting around time bin 1000 and peaking at just above 50 photons.

Finally we show a plot of the laser position that received the total photon count closest to the

median of all laser positions (extended data Fig.3(c)). We can see that this data generally stays below 150 photons with what are likely specular peaks reaching up to 200 photons and a large 450 photon peak at the beginning of the dataset that is either a specular peak or another gate artifact. At this point we want to also point out that since we illuminate only a grid of points at the wall, we do not capture all the specular peaks in our data. In order to see a specular reflection peak from a scene surface we have to get lucky and illuminate the exact spot on the wall that results in the specular reflection to overlap with the the SPAD position (see in the Supplementary Material Figure S.2 for an illustration). Therefore specular peaks in our measurements can vary greatly based on how close to the peak the laser actually sampled the wall. Again we want to point out that this type of uncontrolled artifact does not affect our algorithm.

As we stated above, the time bin with the highest photon count when ignoring obvious early artifacts contains about 1400 photons. Next we plot the laser position that contains this time bin (extended data Fig.3(d)). Note that zero on the x-axis here corresponds to time bin 834. As we see the 1400 photon peak appears very close to the beginning of the transient and may be a gating artifact that occurs in the data just after the opening of the gate. This type of data distortion is described further below. If not a gating artifact, the peak is likely a specular reflection as it is very short and could only be caused by a small isolated diffuse patch or a specular surface in the scene. Peaks from extended diffuse surfaces are necessarily longer in duration.

We thus conclude that while our data contains significant spurious artifacts, the actual photon counts useful for reconstructions are no higher nor cleaner than in previous methods. Note that the removal of early artifacts is only done here to generate Figures M.3(b-d) to allow visualization. All reconstructions shown in the manuscript contain the full recorded data without the removal of any potential artifacts or time bins.

Even though an understanding of the origin of the artifacts in the data is not needed for our method, we can offer some additional speculation for sources of some of them:

- Many of the early peaks in our data are likely related to imperfections in our gating method. When the SPAD gate opens just after the laser pulse has passed, photoelectrons in the SPAD may can cause a detection event that is not due to an actual photon, but rather might be due to the electrons excited by the first bounce light and trapped in long lived states in the SPAD. Even so these electrons are not amplified, they need to be transported off the SPAD junction or they can cause counts as soon as the gate opens.

- In certain cases it is also possible that the gate does not actually block the pulse for some laser positions. The gate has to be positioned such that it blocks the laser in all laser positions, while not blocking any actual signal. This is not always possible, and we do not re-adjust the gate for each position while scanning.

- In our past setups we have observed many effects inside the imaging system that can keep

light trapped long enough to cause a peak at the time when the the NLOS data arrives. This can be due to multiple reflections between lenses, multiphoton fluorescence in the glass or coating of the lenses or stray light reflecting off a random surface at the right distance. We have confirmed some of these effects, but suspect there are many more.

- In particular we can see light that travels from the laser spot to the SPAD, reflects off the surface of the SPAD pixel, is imaged back to the relay wall and comes back to the SPAD. In confocal or near confocal configurations this can create a peak that is many times brighter than the actual data.
- Retro-reflective targets can be used to reduce many of these artifacts, most of which are created either by the laser or a first bounce reflection of the laser. If the hidden target is retroreflecting, the ratio between the brightness of the laser and its first bounce and the brightness of the 3rd bounce NLOS data is reduced by multiple orders of magnitude.

Helmholtz reciprocity: Ideally, we would capture $H(\mathbf{x}_p \rightarrow \mathbf{x}_c, t)$ sampling points on both the projector aperture $\mathbf{x}_p \in P$ and the camera aperture $\mathbf{x}_c \in C$. In our current setup with a single SPAD detector, we only sample a single point for $\mathbf{x}_c$. Using Helmholtz Reciprocity we can interpret these datasets as having a single $\mathbf{x}_p$ and and array of $\mathbf{x}_c$. The choice of capture arrangement is made for convenience since it is easier to calibrate the position of the laser spot on the wall. Improved results are anticipated once array sensors become available (currently under development).

Additional Validation and Discussion

Resolution Limits

The resolution limit for NLOS imaging systems with an aperture diameter d at imaging distance L is closely related to the Rayleigh diffraction limit[7]: $\Delta_x = 1.22c\sigma L/d$, with c the speed of light, for a pulse of full width at half maximum σ. O'Toole et al.[9] derive a criterion for a resolvable object based on the separability of the signal in the raw data, not in the reconstruction, resulting in a similar formula of $\Delta_x = 0.5c\sigma L/d \approx 0.5\lambda L/d$.

In our virtual LOS imaging system, we can formulate a resolution limit that ensures a minimum contrast in the reconstruction, based on the well-known resolution limits of wave based imaging systems. The resolution limit therefore depends on the particular choice of virtual imaging system. For an imaging system that uses focusing only on the detection or illumination side, this limit is approximated by the Rayleigh criterion. For an imaging system that provides focusing both on the light source and the detector side, the resolution doubles (as it does for example in a confocal or structured illumination microscope) and becomes $\Delta_x = 0.61\lambda L/d$.

Effect of Strong Interreflections

In order to confirm the presence and effect of strong interreflections in our captured data, we compare it qualitatively with primary data from a synthetic bookshelf scene, with and without interreflections. The bookshelf is placed in a corridor of 2 m x 2 m x 3 m, with only a single lateral aperture of 1 m x 2 m to allow imaging the hidden scene. The shelf has a size of 1.4 m x 0.5 m, placed at 1.7 m from the relay wall, and 0.3m from the lateral walls. The virtual aperture has a size of 1.792m x 1.7920m and a granularity of 256 x 256 laser points; we use a $\lambda = 4$ cm, $\Delta_p = 2.8$ cm.

As can be seen in extended data Fig.4, the synthetic data clearly shows how the presence of interreflections adds, as expected, low-frequency information resembling echoes of light. This same behavior can be seen in the real captured data, revealing the presence of strong interreflections.

Additionally, we evaluate the robustness of our method in the presence of such interreflections. Similar to recent work[9], we compare between a voxelization of the ground-truth geometry and a reconstructed voxel-grid obtained from our irradiance reconstructions, with and without including interreflections; the resulting mean square error is:

MSE without interreflections (extended data Fig.4.a): 4.93 mm.

MSE with interreflections (extended data Fig.4.b): 4.66 mm.

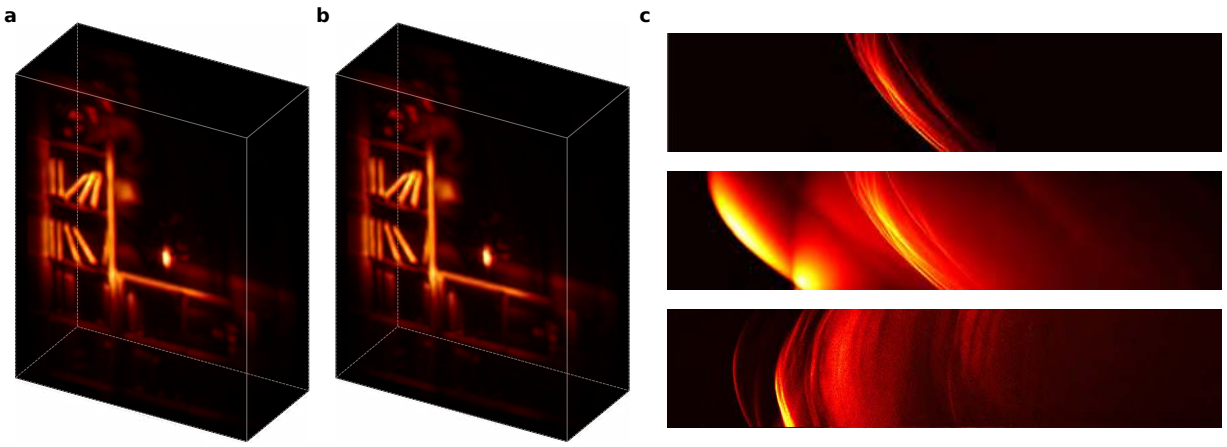

Figure M.4: **Robustness to multiple reflections:** Result on the synthetic bookshelf scene. **a**, Without interreflections. **b**, Including high-order interreflections. The quality of the results is very similar. **c**, Primary data (streak images) from the same scene without (top), and with interreflections (middle). This synthetic data clearly shows how the presence of interreflections adds, as expected, low-frequency information resembling echoes of light. The bottom image shows primary data captured from the real office scene in Figure 2. It follows the same behavior as the middle image, revealing the presence of strong interreflections.

Effect of Exposure Time

Ambient light: To analyze how well our technique works in ambient light and with much faster exposure times, we perform several additional measurements using progressively shorter exposure times, showing that we can reduce exposure times at least down to 50 ms per data point without a significant loss in quality (see extended data Fig.5). Extended data Fig.2 shows raw data for one of the laser positions. In particular, it shows the number of photons per second accumulated in each time bin (i.e. the collected histogram divided by the integration time in seconds). As expected, all three curves appear to follow the same mean, but have a larger variance for lower exposure times. The raw data thus gets significantly noisier as exposure time decreases. The effects on our reconstruction, however, are rather minor as extended data Fig.5 shows.

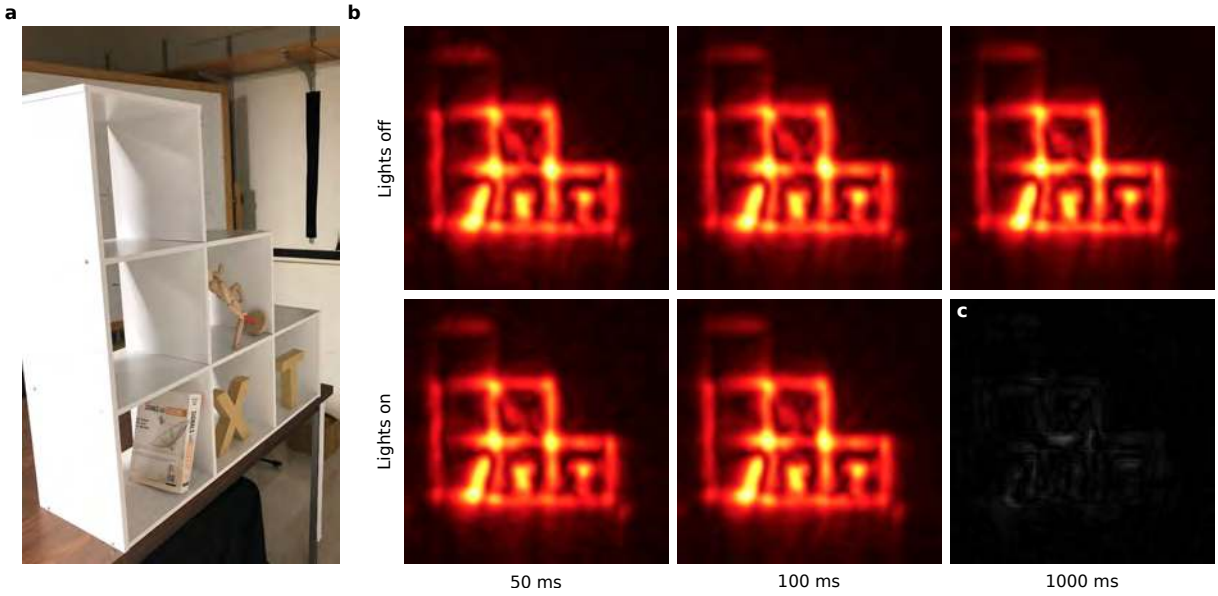

Figure M.5: **Robustness to ambient light and noise:** **a**, Hidden bookshelf. **b**, Imaging results with increasingly higher exposure times; even at 50 ms, there is no significant loss in quality. Top row: Using only the pulsed laser as illumination source. Bottom row: adding a large amount of ambient light (same conditions as the photograph in (a)), the quality also remains constant. **c**, Difference between the 50 ms and 1000 ms exposure captures for the lights off case.

Short exposure captured data: Extended data Fig.6 shows the reconstruction of the office scene (Figure 2) for short exposure times of 10 ms, 5 ms, and 1 ms for each of the roughly 24000 laser positions. This leads to total capture times of about 4 minutes, 2 minutes and 24 seconds respectively. Plots showing raw data from those datasets are shown in extended data Fig.7.

We compare the results of our reconstructions on the 1 ms-data, against filtered backprojection

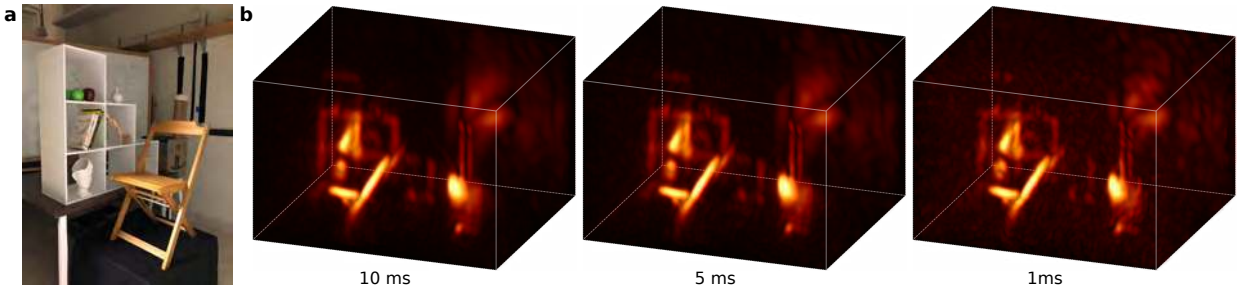

Figure M.6: **Short exposure reconstructions:** Reconstruction of the office scene using very short capture times. **a**, Photograph of the captured scene. **b**, From left to right, reconstructions for data captured with 10, 5, and 1 ms exposure time per laser. The total capture time was about 4 minutes, 2 minutes, and 24 seconds, respectively.

with a Laplacian filter[3], as well as the LOG-filtered backprojection[21], which generally achieves better results. In fact, we are not aware of any reconstruction method that has been shown to consistently outperform a LOG-filtered backprojection. Extended data Fig.8 shows the result of this comparison.

Non-Lambertian Surfaces

To validate the robustness of our method in the presence of non-Lambertian materials in the hidden scene, we have created a synthetic scene made up of two letters, R and D, one partially occluding the other, placed in a corridor of 2m x 2m x 3m, with only a single lateral aperture of 1 m x 2 m to allow imaging the hidden scene. The letters have a size of 0.75 m x 0.8 m, placed at 1.25 m and 1.7 m from the relay wall, respectively, and 0.5m from the lateral walls (see extended data Fig.9.a). The virtual aperture has a size of 1.792 m x 1.792 m and a granularity of 128 x 128 laser points; we use a $\lambda = 4\Delta_p$ with $\Delta_p = 5.6$ cm. We start with purely Lambertian targets, and progressively increase their specularity; we use the Ward BRDF model[29], decreasing the surface roughness, using available transient rendering software[27]. The simulation includes up to the fifth indirect bounce.

Extended data Fig.9.b shows the resulting irradiance reconstructions. Since our method does not make any assumption about the surface properties of the hidden scene, the changes in material appearance do not significantly affect our irradiance reconstructions. Similar to recent work[9] we compare between a voxelization of the ground-truth geometry, and the reconstructed voxel-grid; the resulting mean square error for each of the different reflectances is as follows:

MSE for a surface roughness of 1 (perfect Lambertian): 2.1 mm.

MSE for a surface roughness of 0.4: 2.2 mm.

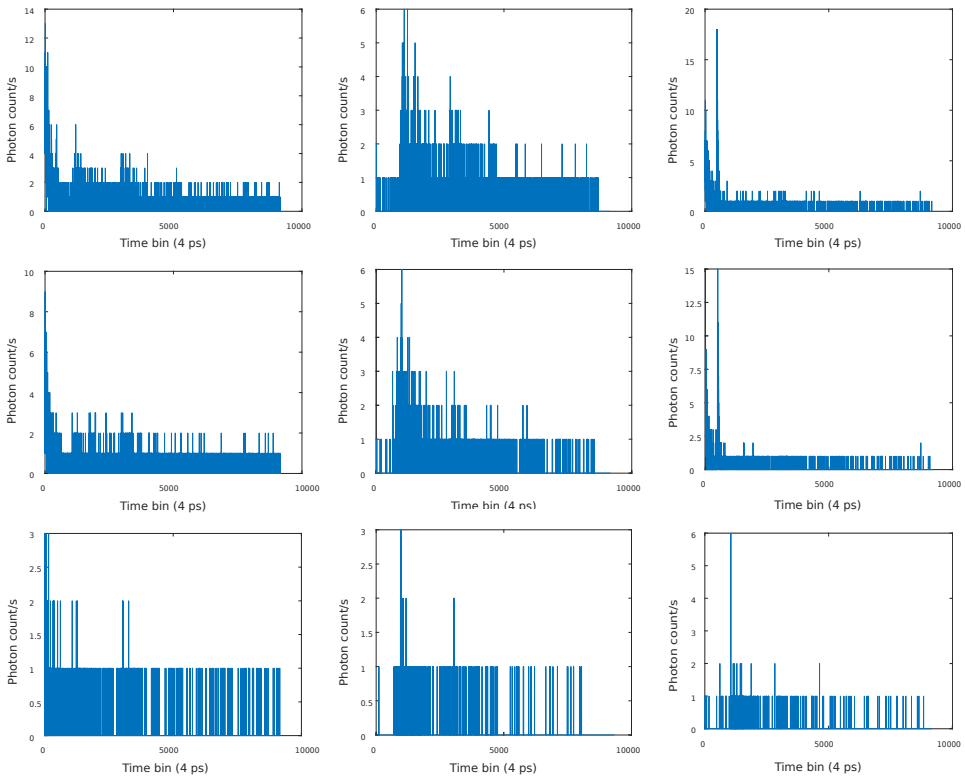

Figure M.7: **Short exposure data:** Photon counts in the raw data for our office scene for 10 ms (top row), 5 ms (center row), and 1 ms (bottom row) exposure times per laser. After removing the first 833 time bins in each dataset the columns show (left) the photon counts for the laser position that received the largest total number of photons in the dataset, (center) the counts for the laser position that received the median number of photon counts, and (right) the laser position that contains the time bin with the global maximum count in the entire set.

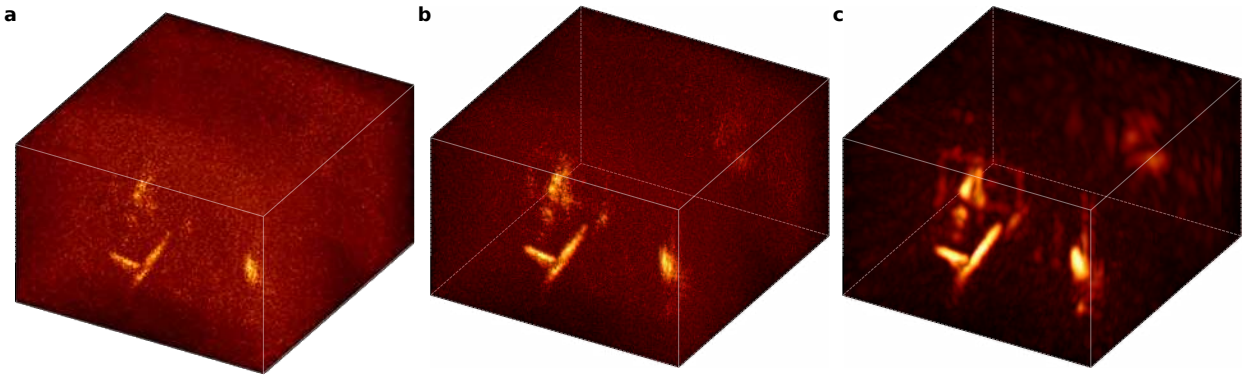

Figure M.8: **Comparison to prior methods:** Reconstruction of the office scene using very short capture times of 1 ms per laser (24 seconds in total). **a**, Filtered backprojection using the Laplacian filter. **b**, LOG-filtered backprojection. **c,** Our method.

MSE for a surface roughness of 0.2: 2.2 mm.

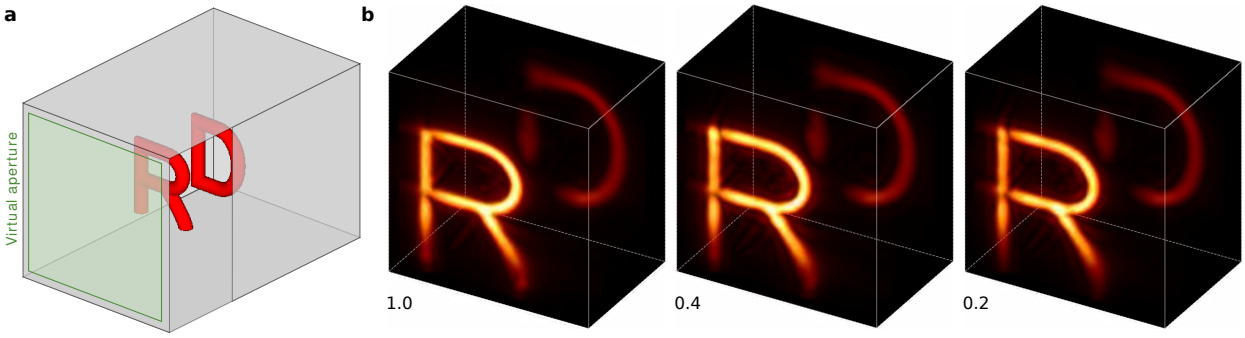

Figure M.9: **Robustness to scene reflectance: a**, Geometry of our experimental setup. **b**, From left to right, imaging results for the Lambertian targets (roughness 1), and increasingly specular surfaces (roughness 0.4 and roughness 0.2). The reconstructed irradiance is essentially the same for all cases.

Reconstruction Comparison with Other Methods

Our imaging system allows to reconstruct hidden geometry. For this particular application, we show here a comparison using the publicly available confocal dataset[9]. This set can be reconstructed using different NLOS methods; we show results for the CNLOS deconvolution[9], filtered backprojection[7], and our proposed method. For these confocal measurements, backprojection can be expressed as a convolution with a pre-calculated kernel, and thus all three methods are using the same backprojection operator. Neither our method nor filtered backprojection are limited to confocal data, and can be acquired making use of simpler devices and capture configurations. They can thus be applied to a broader set of configurations and considerably more complex scenes. For the CNLOS deconvolution method[9], we leave the optimal parameters unchanged. For our proposed virtual wave method, we use the aperture size and its spatial sampling grid published in the supplementary materials to calculate the optimal phasor field wavelength. For the filtered backprojection it is important to choose a good discrete approximation of the Laplacian operator in the presence of noise. Previous works implicitly do the denoising step by adjusting the reconstruction grid size to approximately match the expected reconstruction quality[2,3,7], or by downsampling across the measurements[9]. All of them can be considered as proper regularizers. To provide a fair comparison without changing the reconstruction grid size, we convolve a Gaussian denoising kernel with the Laplacian kernel, resulting in a LOG filter, which we apply over the backprojected volume.

Note that a large improvement in reconstruction quality for the simple scenes included in the dataset (isolated objects with no interreflections) is not to be expected, since existing methods already deliver reconstructions approaching their resolution limits. We nevertheless achieve im-

proved contrast and cleaner contours in our wave camera method, due to our better handling of multiply scattered light, which pollutes the reconstructions in the other methods (see extended data Fig.10).

In the noisy datasets (extended data Fig.11), filtered backprojection fails. CNLOS includes a Wiener filter that performs well removing uniform background noise, although a noise level must be *explicitly* estimated. Our phasor field virtual wave method, on the other hand, performs well automatically, without the need to explicitly estimate a noise level. This is particularly important in complex scenes with interreflections, where the background is not uniform across the scene, and the noise level cannot be reliably estimated.

Nevertheless, our main contribution is not improving the reconstruction for simple, 3rd-bounce scenes. Instead, our method allows to derive a new class of NLOS algorithms, which can successfully handle scenes of much larger complexity.

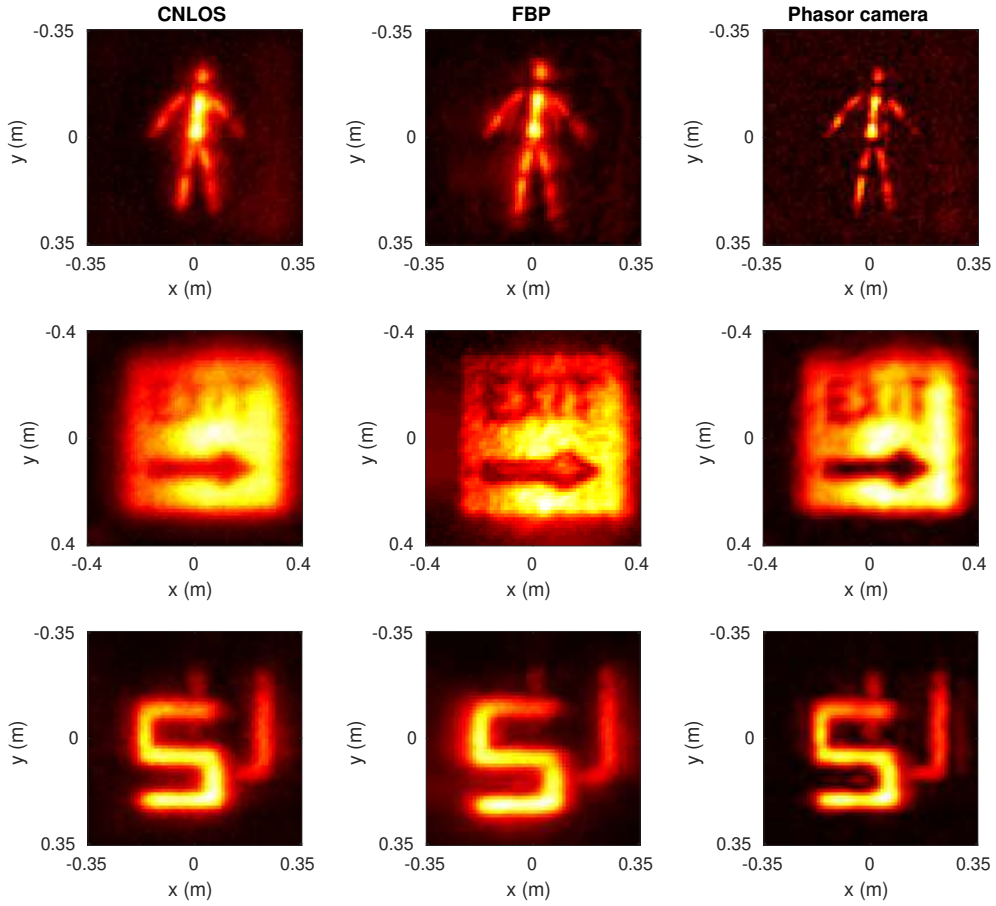

Figure M.10: **Reconstruction comparison on a public dataset**. From left to right: CNLOS deconvolution, filtered (LOG) backprojection, and our proposed method. A large improvement in reconstruction quality for the simple scenes included in the dataset (isolated objects with no interreflections) is not to be expected, since existing methods already deliver reconstructions approaching their resolution limits. Nevertheless, our method achieves improved contrast and cleaner contours, due to better handling of multiply scattered light.

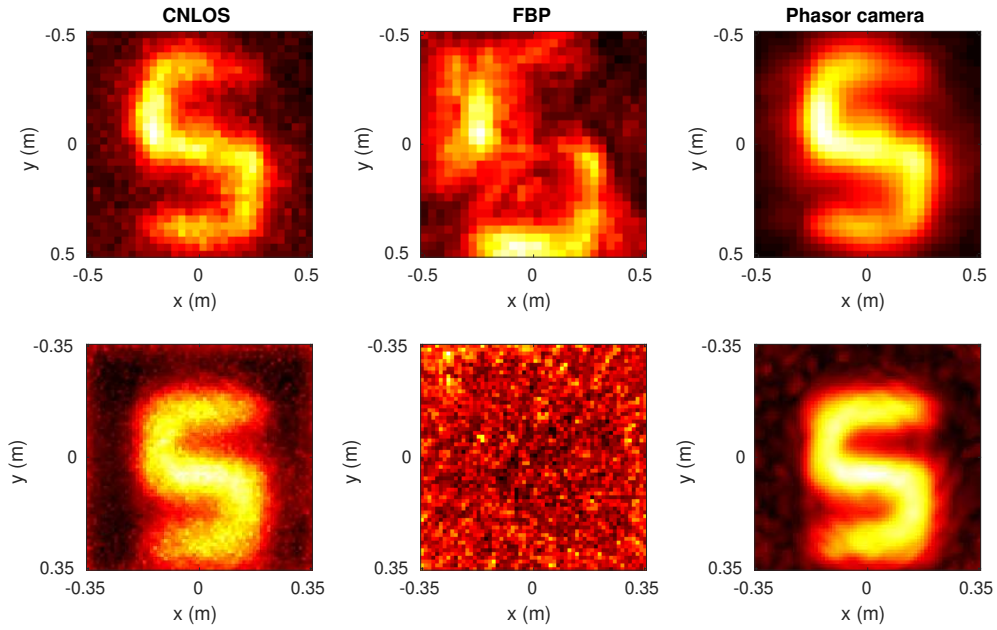

Figure M.11: **Reconstruction comparison on a public dataset**. From left to right: CNLOS deconvolution, filtered (LOG) backprojection, and our proposed method. A large improvement in reconstruction quality for the simple scenes included in the dataset (isolated objects with no interreflections) is not to be expected, since existing methods already deliver reconstructions approaching their resolution limits. Nevertheless, our method achieves improved contrast and cleaner contours, due to better handling of multiply scattered light.

References

29. Ward, G. J. Measuring and modeling anisotropic reflection. *Computer Graphics* **26.2**, 265-272 (1992).

Supplementary Materials

Non-Line-of-Sight Imaging using Phasor Field Virtual Wave Optics

Xiaochun Liu[1], Ibón Guillén[2], Marco La Manna[3], Ji Hyun Nam[1], Syed Azer Reza[3], Toan Huu Le[1], Adrian Jarabo[2], Diego Gutierrez[2], Andreas Velten[1,3,4]

[1]*Department of Electrical and Computer Engineering, University of Wisconsin Madison*
[2]*Graphics and Imaging Lab, Universidad de Zaragoza - I3A*
[3]*Department of Biostatistics and Medical Informatics, University of Wisconsin Madison*
[4]*Corresponding author: velten@wisc.edu*

This supplemental document contains the following information:

A Overview of the Method

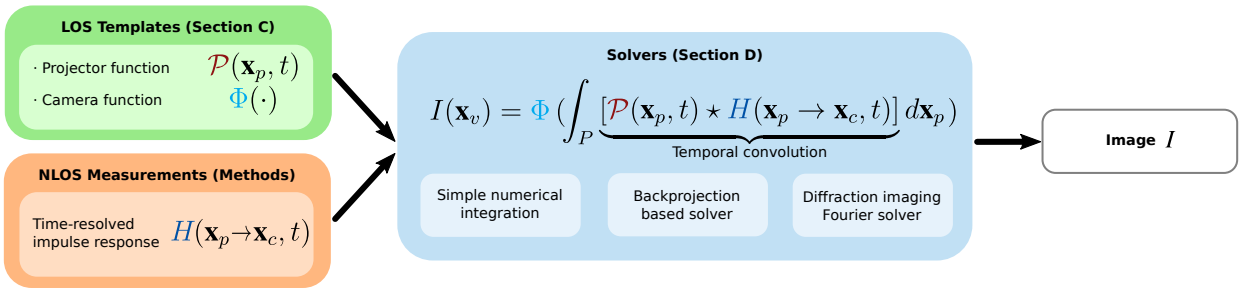

Figure S.1: Our imaging framework: The flowchart describes how to image an NLOS scene from a chosen template LOS system. Please refer to the text for details.

Figure S.1 illustrates the flowchart of our method, while Table S.1 summarizes the symbols used in the main paper and throughout this document. One of the key ideas is that we transform the relay wall into the *virtual aperture* or lens of any LOS system (see Figure 1 in the main paper). By formulating the NLOS problem as a diffractive wave propagation problem, we can image hidden scenes of unprecedented complexity from raw time-of-flight data, by applying the mathematical operators that model wave propagation inside our virtual LOS imaging system (Section B).

First, a *template* LOS imaging system is selected (green box); our corresponding NLOS algorithm will inherit the same capabilities and features as the chosen template LOS system. Note that only a computational model is needed, not a physical implementation. The LOS system is characterized by the source function $\mathcal{P}(\mathbf{x}_p, t)$ describing the illumination wavefront, and the camera function $\Phi(\cdot)$ describing the imaging process at virtual LOS aperture. Both functions can be derived from existing diffraction and LOS literature. In Section C we derive such functions for the three LOS examples we have implemented: A conventional photography camera; a transient camera; and a confocal time-gated camera.

The blue box illustrates our imaging equation, using the chosen $\mathcal{P}(\mathbf{x}_p, t)$ and $\Phi(\cdot)$ functions:

$$I(\mathbf{x}_v) = \Phi\left(\int_P [\mathcal{P}(\mathbf{x}_p, t) \star H(\mathbf{x}_p \rightarrow \mathbf{x}_c)] d\mathbf{x}_p\right), \tag{S.1}$$

where $\star$ is the convolution operator in the temporal domain. Background to derive this equation is included in Section B. Equation S.1 can be solved numerically, which is straightforward, but slow. We propose two alternative solvers: Since the camera function $\Phi(\cdot)$ is essentially an RSD propagator, we can rewrite Equation S.1 in terms of an RSD operator, for which fast algorithms exist. Moreover, we note that Equation S.1 can be expressed as a backprojection, enabling the use existing backprojection methods[1,2]. We illustrate the flexibility of our approach by applying these different solvers to our three template LOS systems. This is described in more detail in Section D.

Last, Methods Section provides details about our NLOS data acquisition step ($H(\mathbf{x}_p \rightarrow \mathbf{x}_c, t)$, orange box), which is similar to other NLOS methods. Additional results, validation and discussions are included in Methods Section.

Symbol	Definition	Units
S	Source plane of a propagating wave	
D	Destination plane of a propagating wave	
P	Virtual projector aperture at the relay wall (Figure 1, a)	
C	Virtual camera aperture at the relay wall (Figure 1, a)	
$\mathbf{x}_s, \mathbf{x}_d, \mathbf{x}_p, \mathbf{x}_c$	Point at planes S, D, P, and C respectively	
Δ_p, Δ_c	Distance between points at planes P and C respectively	$[m]$
$\mathbf{x}_v$	Reconstructed (voxel) point of a volumetric space	
Δ_x	Size of each voxel of the reconstructed volume	$[m]$
t	Time	
$\mathcal{P}(\mathbf{x}, t)$	Phasor field	$[Wm^{-2}] \in \mathbb{R}$
$\mathcal{P}_\omega(\mathbf{x}, t)$	Monochromatic phasor field	$[Wm^{-3}] \in \mathbb{C}$
λ	Wavelength of the monochromatic phasor field	$[m]$
ω	Frequency of the monochromatic phasor field	$[s^{-1}]$
$\mathcal{P}_{0,\omega}(\mathbf{x})$	Amplitude of the monochromatic phasor field	$[Wm^{-2}] \in \mathbb{C}$
$\mathbf{x}_{ls}$	Position of the virtual laser	
t_0	Time of emission of the virtual laser	$[s]$
σ	Pulse's $e^{-\frac{1}{2}}$ attenuation radius	$[s]$
$\mathcal{E}(\mathbf{x}, t)$	Electric field	$[Vm^{-1}]$
$\mathcal{E}_0(\mathbf{x})$	Amplitude of the electric field	$[Vm^{-1}]$
Ω_0	Center frequency of the electric field E	$[s^{-1}]$
$\Delta\Omega$	Frequency bandwidth of the electric field E	$[s^{-1}]$
τ	Average time interval of the ultrafast detector	$[s]$
T	Long term average time interval ($\tau << \Delta\Omega^{-1} << T$)	$[s]$
$H(\mathbf{x}_p \rightarrow \mathbf{x}_c, t)$	Time-resolved impulse response function	
$\Phi(\cdot)$	Image formation model	
I	Image reconstructed by operator $\Phi(\cdot)$	

Table S.1: Table of symbols used in our paper.

B Derivation of the Phasor Field

Here we derive the Rayleigh-Sommerfeld diffraction (RSD) integral for the phasor field. Consider a point light source at a location $\mathbf{x}_s$ that emits light with a sinusoidal time-varying irradiance $L(\mathbf{x}, t) = \text{Re}[L_0\,(e^{-i\omega t} + 1)]$ with amplitude L_0 and modulation frequency ω. More formally, $L(t)$ and L_0 relate with the electromagnetic field $\mathcal{E}(\mathbf{x}, t)$ as $L(\mathbf{x}, t) = \left\langle \frac{1}{\tau} \int_{t-\tau/2}^{t+\tau/2} |\mathcal{E}(\mathbf{x}, t')|^2 dt' \right\rangle$ and $L_0(\mathbf{x}) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{+T/2} |\mathcal{E}(\mathbf{x}, t)|^2 dt$, with τ a sufficiently small value. The operator $\langle\cdot\rangle$ is the spatial averaging operator that takes into account multiple possible measurements for e.g. removing the effect of laser's speckle.

We define the real-valued phasor field $\mathcal{P}(\mathbf{x}, t)$ at a point in space as

$$\mathcal{P}(\mathbf{x}, t) = L(\mathbf{x}, t) - L_0(\mathbf{x}). \tag{S.2}$$

Since $L(\mathbf{x}, t)$ is modulated with a single frequency ω, this allows us to consider $\mathcal{P}(\mathbf{x}_s, t)$ as a monochromatic phasor field wave $\mathcal{P}_\omega$ emitted from a point light source at location x_s, with amplitude $\mathcal{P}_{0,\omega}(\mathbf{x}_s)$ and oscillating at a frequency ω:

$$\mathcal{P}(\mathbf{x}_s, t) = \mathcal{P}_\omega(\mathbf{x}_s, t) = \mathcal{P}_{0,\omega}(\mathbf{x}_s) e^{i\omega t}. \tag{S.3}$$

In the following, wherever we write an explicitly complex expression for $\mathcal{P}(\mathbf{x}_s, t)$, it is implied that the actual real phasor field is $\frac{1}{2}(\mathcal{P}(\mathbf{x}_s, t) + \mathcal{P}^*(\mathbf{x}_s, t))$. In practice we can safely ignore the complex conjugate component in all our computations. Note that the constant $L_0(\mathbf{x})$ term is only necessary to link the phasor field wave to a measurable physical quantity, since real intensities cannot be negative. We can think of it as the monochromatic wave component at frequency $\omega = 0$. Since our propagator is linear and does not mix different frequency components it can safely be ignored as it can only create zero-frequency contributions.

To determine the light intensity and thereby the phasor field at any point in space and time $(\mathbf{x}_d, t)$ we have to account for the travel time from $\mathbf{x}_s$ to $\mathbf{x}_d$, defined as $t_p = |\mathbf{x}_d - \mathbf{x}_s|/c$, with c the propagation speed, and the radial drop-off in light intensity:

$$\mathcal{P}_\omega(\mathbf{x}_d, t) = \mathcal{P}_{0,\omega}(\mathbf{x}_s) \frac{e^{i\omega(t+t_p)}}{|\mathbf{x}_d - \mathbf{x}_s|^2} = \mathcal{P}_{0,\omega}(\mathbf{x}_s) \frac{e^{i\omega(t+|\mathbf{x}_d-\mathbf{x}_s|/c)}}{|\mathbf{x}_d - \mathbf{x}_s|^2} = \mathcal{P}_{0,\omega}(\mathbf{x}_s) \frac{e^{i\omega t + ik|\mathbf{x}_d-\mathbf{x}_s|}}{|\mathbf{x}_d - \mathbf{x}_s|^2}, \tag{S.4}$$

where $k = 2\pi/\lambda$ is the wave number at the modulation wavelength, λ. If instead of a single light source we have a collection of incoherent sources comprising a surface S, we have

$$\mathcal{P}_\omega(\mathbf{x}_d, t) = \int_S \mathcal{P}_{0,\omega}(\mathbf{x}_s) \frac{e^{i\omega t + ik|\mathbf{x}_s-\mathbf{x}_d|}}{|\mathbf{x}_s - \mathbf{x}_d|^2} d\mathbf{x}_s. \tag{S.5}$$

This equation looks like the Rayleigh-Sommerfeld propagator, except for the squared denominator, and the missing $1/i\lambda$. The $1/i\lambda$ term is a global constant that does not qualitatively affect our

propagator. We approximate $|\mathbf{x}_s - \mathbf{x}_d|^2 \approx |\mathbf{x}_s - \mathbf{x}_d||\mathbf{x}_{avgS} - \mathbf{x}_d|$ where $\mathbf{x}_{avgS}$ is the average position of all source points in S. Pulling this constant term out of the integral, we obtain

$$\begin{aligned}\mathcal{P}_\omega(\mathbf{x}_d, t) &\approx \frac{1}{|\mathbf{x}_{avgS} - \mathbf{x}_d|} \int_S \mathcal{P}_{0,\omega}(\mathbf{x}_s) \frac{e^{i\omega t + ik|\mathbf{x}_s - \mathbf{x}_d|}}{|\mathbf{x}_s - \mathbf{x}_d|} d\mathbf{x}_s \\ &= \frac{1}{|\mathbf{x}_{avgS} - \mathbf{x}_d|} \int_S \mathcal{P}_\omega(\mathbf{x}_s, t) \frac{e^{ik|\mathbf{x}_s - \mathbf{x}_d|}}{|\mathbf{x}_s - \mathbf{x}_d|} d\mathbf{x}_s, \end{aligned} \tag{S.6}$$

which is the RSD (Equation 4 in the main text) for scalar waves, with $\gamma = 1/|\mathbf{x}_{avgS} - \mathbf{x}_d|$. This approximation does not affect the phase term, causing only a slow-varying error in amplitude. Given a known source plane, this error can be precomputed. Since it does not alter the phase of the signal, it has no effect on the reconstructed geometry.

Furthermore, as we show in Section C.1, most real imaging systems do not invert the $1/r$ term in the RSD propagator. Further research may also lead to alternative formulations of the phasor field that deal with this error in a more elegant way.

In the following, we use $\mathcal{R}_{\mathbf{x}_d}\left(\mathcal{P}_{0,\omega}(\mathbf{x}_s, t)\right)$ as a shorthand for the RSD operator:

$$\mathcal{R}_{\mathbf{x}_d}\left(\mathcal{P}_{0,\omega}(\mathbf{x}_s, t)\right) = \frac{1}{|\mathbf{x}_{avgS} - \mathbf{x}_d|} \int_S \mathcal{P}_\omega(\mathbf{x}_s, t) \frac{e^{ik|\mathbf{x}_s - \mathbf{x}_d|}}{|\mathbf{x}_s - \mathbf{x}_d|} d\mathbf{x}_s. \tag{S.7}$$

B.1 Propagating Broadband Signals

The derived RSD operator propagates only monochromatic waves $\mathcal{P}_\omega(\mathbf{x}_s, t) = \mathcal{P}_{0,\omega}(\mathbf{x}_s)e^{i\omega t}$. Any broadband signal can be propagated by first writing it as a superposition of monochromatic waves, then propagating each one individually. For a general $\mathcal{P}(\mathbf{x}, t)$ we therefore define the RSD operator as:

$$\mathcal{R}_{\mathbf{x}_d}(\mathcal{P}(\mathbf{x}, t)) = \mathcal{R}_{\mathbf{x}_d}\left(\int_{-\infty}^{+\infty} \mathcal{P}_\omega(\mathbf{x}, t) \frac{d\omega}{2\pi}\right) = \int_{-\infty}^{+\infty} \mathcal{R}_{\mathbf{x}_d}\left(\mathcal{P}_\omega(\mathbf{x}, t)\right) \frac{d\omega}{2\pi}. \tag{S.8}$$

Alternatively, a broadband RSD operator can be implemented in the time domain by shifting the components of $\mathcal{P}$ in time as follows:

$$\begin{aligned}\mathcal{R}_{\mathbf{x}_d}\left(\mathcal{P}(\mathbf{x}, t)\right) &= \int_{-\infty}^{+\infty} \mathcal{R}_{\mathbf{x}_d}\left(\mathcal{P}_\omega(\mathbf{x}, t)\right) \frac{d\omega}{2\pi} \\ &= \int_{-\infty}^{+\infty} \int_S \mathcal{P}_{0,\omega}(\mathbf{x}_s) e^{i\omega t - ik|\mathbf{x}_s - \mathbf{x}_d|} d\mathbf{x}_s \frac{d\omega}{2\pi} \\ &= \int_S \int_{-\infty}^{+\infty} \mathcal{P}_{0,\omega}(\mathbf{x}_s) e^{i\omega(t - \frac{1}{c}|\mathbf{x}_s - \mathbf{x}_d|)} \frac{d\omega}{2\pi} d\mathbf{x}_s \\ &= \int_S \mathcal{P}(\mathbf{x}_s, t - \frac{1}{c}|\mathbf{x}_s - \mathbf{x}_d|) d\mathbf{x}_s \end{aligned} \tag{S.9}$$

We will make use of this property when deriving the camera functions for our example cameras (see Table S.3).

B.2 Non-Lambertian Surfaces

The RSD propagator we derive requires that the source plane S be Lambertian. Since all our cameras rely primarily on RSD propagators from the aperture of the relay wall, the Lambertian constraint only applies to the relay wall. Rather than reconstructing the geometry and BRDF of the scene, our virtual cameras reconstruct phasor field irradiance from the scene towards our virtual aperture as a function of position and time, analogous to their LOS counterparts. The reconstructed signal thus corresponds to the averaged irradiance for the entire aperture. This is illustrated in Figure S.2. Prior methods seek to reconstruct NLOS geometry, which requires correct modelling of albedo, BRDF, occlusions, and interreflections, resulting in a nonlinear inverse problem[3]. In the absence of such data from the hidden scene, these prior methods need to rely on simplifying assumptions, thus limiting the range of scenes that can be reconstructed. Since our method does not make any assumption about the surface properties of the hidden scene, the changes in material appearance do not significantly affect our irradiance reconstructions (see Results in Method section). Simulations with varying BRDFs can be found in the Method Section.

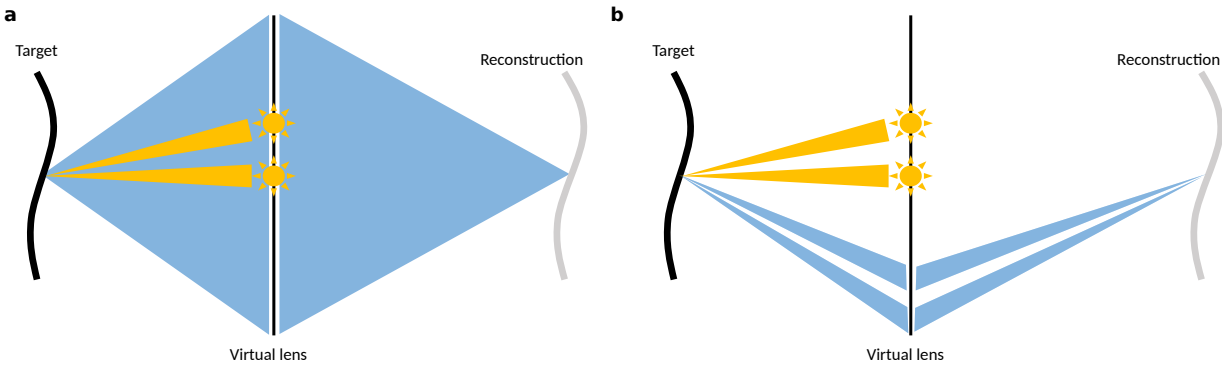

Figure S.2: Light hitting the relay wall (in yellow) illuminates the target scene, and is scattered and bounced back by unknown objects in the scene (in blue). **a)** When the target object is perfectly Lambertian, it bounces light back to the entire captured surface of the relay (our virtual lens); our virtual imaging system then focuses the incoming irradiance. **b)** Increasingly specular surfaces may cause the returning light to be reflected towards specific directions; however, like a conventional camera, as long as such reflected light hits some area of the virtual lens, it will be imaged correctly, with a potential spatial resolution loss if few light sources are used. Increasing the number of light sources allows to obtain a progressively more complete irradiance reconstruction and improve resolution.

C LOS Template Functions

In this section we show how to model the outgoing illumination wavefront $\mathcal{P}(\mathbf{x}_p, t)$, as well as the image formation model $\Phi(\cdot)$ for our example *template* LOS imaging systems, using standard diffraction optics. We begin with some preliminary considerations regarding the phase transformation by an ideal lens, which is essential for deriving any arbitrary image formation model $\Phi(\cdot)$. We then derive $\mathcal{P}(\mathbf{x}_p, t)$ and $\Phi(\cdot)$ for three example systems: (1) Conventional Photography Camera, (2) Transient Camera, and (3) Confocal Time-Gated Camera.

C.1 Phase Operator of an Ideal Computational Thin Lens

We define an ideal lens as an element that focuses a planar wavefront into a point at the focal distance f from the lens, on the optic axis. This is equivalent to converting light coming from a point $\mathbf{x}_f$ and turning it into a planar wave, i.e., a wave with a phase that is independent of the position $\mathbf{x}_l$ on the plane of the lens. Light leaving from a point at $\mathbf{x}_f$ creates spherical wavefronts, i.e., the phase at a plane perpendicular to the z-direction at a distance f from $\mathbf{x}_f$ is

$$\phi_w(\mathbf{x}_l, \mathbf{x}_f) = e^{i\omega \frac{|\mathbf{x}_f - \mathbf{x}_l|}{c}}. \tag{S.10}$$

The lens phase shift, $\phi_l(\mathbf{x}_l, \mathbf{x}_f)$, has to cancel this phase term, and thus the lens acts on the wavefront of a monochromatic wave $\mathcal{P}_\omega(\mathbf{x})$ as

$$\mathcal{P}'_\omega(\mathbf{x}_l, t) = \mathcal{P}_\omega(\mathbf{x}_l, t) \cdot \phi_l(\mathbf{x}_f, \mathbf{x}_l). \tag{S.11}$$

where $\mathcal{P}'_\omega(\mathbf{x}_l, t)$ is the wavefront *after* the lens and $\phi_l(\mathbf{x}_l, \mathbf{x}_f) = e^{-i\omega |\mathbf{x}_f - \mathbf{x}_l|/c}$.

To understand how this lens affects a general broadband signal, consider a wave $\mathcal{P}(\mathbf{x}, t)$ expressed as a superposition of monochromatic waves:

$$\mathcal{P}(\mathbf{x}, t) = \int_{-\infty}^{+\infty} \mathcal{P}_\omega(\mathbf{x}, t) \frac{d\omega}{2\pi} = \int_{-\infty}^{+\infty} \mathcal{P}_{0,\omega}(\mathbf{x}) e^{i\omega t} \frac{d\omega}{2\pi}; \tag{S.12}$$

Considering $\mathbf{x}_f$ and $\mathbf{x}_l \in L$ and applying the phase shift of the lens to this wavefront, we find

$$\mathcal{P}(\mathbf{x}, t)\, \phi_l(\mathbf{x}_l, \mathbf{x}_f) = \int_{-\infty}^{+\infty} \mathcal{P}_{0,\omega}(\mathbf{x})\, e^{i(\omega t - \omega \frac{|\mathbf{x}_f - \mathbf{x}_l|}{c})} \frac{d\omega}{2\pi} = \mathcal{P}(\mathbf{x}, t - \frac{|\mathbf{x}_f - \mathbf{x}_l|}{c}). \tag{S.13}$$

Like the RSD propagation, the phase shift of an ideal lens can thus also be described as a shift in time.

System	$\mathcal{P}(\mathbf{x}_p, t)$
(1) Photo Camera (ambient light)	$e^{i\omega t}$
(2) Transient Camera (pulsed point light)	$e^{i\omega t}\delta(\mathbf{x}_p - \mathbf{x}_{ls})e^{-\frac{(t-t_0)^2}{2\sigma^2}}$
(3) Confocal Time-Gated Camera (pulsed focused light)	$e^{i\omega(t-\frac{1}{c}\lvert\mathbf{x}_v-\mathbf{x}_p\rvert)}e^{-\frac{(t-t_0-\frac{1}{c}\lvert\mathbf{x}_v-\mathbf{x}_p\rvert)^2}{2\sigma^2}}$

Table S.2: Illumination wave functions for different light sources, used in our three example imaging systems.

System	$\Phi(\mathcal{P}(\mathbf{x}_c, t))$
(1) Photo Camera	$\lvert\int_C \mathcal{P}(\mathbf{x}_c, t-\frac{1}{c}\lvert\mathbf{x}_v-\mathbf{x}_c\rvert)d\mathbf{x}_c\rvert^2 = \lvert\mathcal{R}_{\mathbf{x}_v}(\mathcal{P}(\mathbf{x}_c, t))\rvert^2$
(2) Transient Camera	$\lvert\int_C \mathcal{P}(\mathbf{x}_c, t-\frac{1}{c}\lvert\mathbf{x}_v-\mathbf{x}_c\rvert)d\mathbf{x}_c\rvert^2 = \lvert\mathcal{R}_{\mathbf{x}_v}(\mathcal{P}(\mathbf{x}_c, t))\rvert^2$
(3) Confocal Time-Gated Camera	$\lvert\int_C \mathcal{P}(\mathbf{x}_c, -\frac{1}{c}\lvert\mathbf{x}_v-\mathbf{x}_c\rvert)d\mathbf{x}_c\rvert^2 = \lvert\mathcal{R}_{\mathbf{x}_v}(\mathcal{P}(\mathbf{x}_c, -\frac{1}{c}\lvert\mathbf{x}_v-\mathbf{x}_c\rvert))\rvert^2$

Table S.3: Imaging operators to implement our three example imaging systems. The evaluation functions essentially describe the imaging transform of a lens with the resulting image being read out at different times with respect to the illumination.

Imaging with a lens: A lens that images a point $\mathbf{x}_v$ onto a sensor pixel $\mathbf{x}_r$ can be described as a combination of two co-located lenses. One to collimate the light from $\mathbf{x}_v$, and one to focus it onto $\mathbf{x}_r$. This results in a phase shift $\phi(\mathbf{x}_l) = -\omega\frac{|\mathbf{x}_v-\mathbf{x}_l|-|\mathbf{x}_l-\mathbf{x}_r|}{c}$. In the camera, this is followed by a propagation from the lens to the sensor. If we use Equations S.9 and S.13 for the propagation and lens, we obtain:

$$\begin{aligned}\mathcal{P}(\mathbf{x}_r, t) &= \int_L \mathcal{P}(\mathbf{x}_l, t - |\mathbf{x}_v - \mathbf{x}_l| + |\mathbf{x}_l - \mathbf{x}_r| - |\mathbf{x}_l - \mathbf{x}_r|)d\mathbf{x}_l \\ &= \int_L \mathcal{P}(\mathbf{x}_l, t - |\mathbf{x}_v - \mathbf{x}_l|)d\mathbf{x}_l \\ &= \mathcal{R}_{\mathbf{x}_v}(\mathcal{P}(\mathbf{x}_l, t)),\end{aligned} \tag{S.14}$$

where the $1/r$ factor in the RSD propagator can be omitted. The imaging lens thus effectively propagates the field from the aperture $\mathbf{x}_l$ back into the scene.

C.2 Example Projector and Camera Functions

Our theoretical model allows us to implement any arbitrary (virtual) camera system by defining the projector function $\mathcal{P}(\mathbf{x}_p, t)$ and imaging operator $\Phi(\cdot)$. Methods for modeling such function using Fourier optics are available in the literature[4]. In our work we implement three of them: (1) Conventional Photography Camera, (2) Transient Camera, and (3) Confocal Time-Gated Camera.

Each has capabilities never before demonstrated in NLOS imaging. The derived $\mathcal{P}(\mathbf{x}_p, t)$ and $\Phi(\cdot)$ functions are listed in Tables S.2 and S.3.

(1) Our first example is a **conventional photography camera** system with a $\mathcal{P}$-field monochromatic illumination source at frequency ω. It reconstructs the hidden scene with low computational effort. Like a LOS photography camera it does not require knowledge of the position or timing of the light source.

In other words, the reconstruction is independent of the position of $\mathbf{x}_p$, which reduces the need of careful calibration of the laser positions in the relay wall. Since our illumination for this system is ambient light, the projector function $\mathcal{P}(\mathbf{x}_p, t)$ can be anything. However, like in conventional imaging, the resolution of the image is determined by the temporal bandwidth of $\mathcal{P}(\mathbf{x}_p, t)$, corresponding to the wavelength in the conventional camera. We thus choose a function with a short phasor field wavelength: $\mathcal{P}(\mathbf{x}_p, t) = e^{i\omega t}$. The camera operator is represented by a lens that creates an image on a set of detector pixels that record the absolute value squared of the field. Implementing the lens using the time shift property or the RSD propagator (Equation S.14) yields $\Phi(\mathcal{P}(\mathbf{x}_c, t)) = |\int_C \mathcal{P}(\mathbf{x}_c, t - \frac{1}{c}|\mathbf{x}_v - \mathbf{x}_c|)d\mathbf{x}_c|^2 = |\mathcal{R}_{\mathbf{x}_v}(\mathcal{P}(\mathbf{x}_c, t))|^2$. Note that this expression is constant with time, just like the intensity in the sensor of a LOS photography camera, so it can be evaluated at any time t.

(2) The second example is a **NLOS transient camera** system. Like its LOS counterpart[5], this camera captures the propagation of light through the scene, revealing complex multibounce light transport phenomena. As a consequence this virtual camera may be used to identify direct and global components of such light transport. In this case we model a monochromatic point light source at a single point $\mathbf{x}_{ls}$ which illuminates the scene with a short gaussian shape flash of $\sigma = \frac{6\lambda}{2.36}$ at time t_0. The illumination function is thus $\mathcal{P}(\mathbf{x}_p, t) = e^{i\omega t}\delta(\mathbf{x}_p - \mathbf{x}_{ls})e^{\frac{-(t-t_0)^2}{2\sigma^2}}$. The camera is the same as the conventional photography camera, except that the reconstructed intensity on the sensor now depends on time t, capturing frames at each t_f. We assume that the camera focus follows the light pulse.

(3) Last, we implement a **confocal time-gated** imaging system, which images specific voxels $\mathbf{x}_v$ of a volumetric space, illuminated with a focused ultrashort pulse of width σ.

Note that our virtual imaging system is confocal, but the data for H is not necessarily captured with a confocal arrangement as in prior NLOS work[2]. Our illumination is a light pulse focused on a voxel $\mathbf{x}_v$: $\mathcal{P}(\mathbf{x}_p, t) = e^{i\omega(t - \frac{1}{c}|\mathbf{x}_v - \mathbf{x}_p|)}e^{-\frac{(t - t_0 - \frac{1}{c}|\mathbf{x}_v - \mathbf{x}_p|)^2}{2\sigma^2}}$. In the design of this system we can choose the phasor field pulse width σ. As this width increases, the depth resolution of the virtual imaging system worsens, although the signal-to-noise ratio improves. In practice, we found that a pulse full width at half maximum of about six wavelengths $\sigma = \frac{6\lambda}{2.36}$ yields the best results. Longer pulses are effective for canceling more noise in the reconstruction. The camera is again implemented as an imaging lens, like the cases above. However, in this case the camera needs

to focus on the same point $\mathbf{x}_v$ as the light source. Since we are only concerned with the 3rd bounce return directly reflected by a scene surface at $\mathbf{x}_v$, we evaluate the signal only at a time $t = -\frac{1}{c}|\mathbf{x}_v - \mathbf{x}_c|$, when 3rd-bounce light from this location is seen. This results in a function $\Phi(\mathcal{P}(\mathbf{x}, t)) = |\int_C \mathcal{P}(\mathbf{x}_c, -\frac{1}{c}|\mathbf{x}_v - \mathbf{x}_c|)d\mathbf{x}_c|^2 = |\mathcal{R}_{\mathbf{x}_v}(\mathcal{P}(\mathbf{x}_c, -\frac{1}{c}|\mathbf{x}_v - \mathbf{x}_c|))|^2$.

D Implementation Details of the RSD Solvers

Here we describe an RSD diffraction integral solver to implement our conventional photographic camera system, and a backprojection solver for the transient confocal systems (refer to Section C.2). Note that both solvers can be applied to any of the systems.

D.1 Conventional Photography Camera using RSD

Using Equation S.1 we write

$$\begin{aligned} I(\mathbf{x}_v) &= \Phi\left(\int_P \left[\mathcal{P}(\mathbf{x}_p, t) \star H(\mathbf{x}_p \to \mathbf{x}_c, t)\right] d\mathbf{x}_p\right) \\ &= \Phi\left(\int_P \int_{-\infty}^{+\infty} \mathcal{P}(\mathbf{x}_p, t-\tau)\, H(\mathbf{x}_p \to \mathbf{x}_c, \tau) d\tau\, d\mathbf{x}_p\right). \end{aligned} \tag{S.15}$$

Plugging in the terms from Tables S.2, and S.3 for the conventional photographic camera, we find:

$$I(\mathbf{x}_v) = \left|\mathcal{R}_{\mathbf{x}_v}\left(\int_P \left[e^{i\omega t} \star H(\mathbf{x}_p \to \mathbf{x}_c, t)\right] d\mathbf{x}_p\right)\right|^2. \tag{S.16}$$

After this convolution, each time response can be represented entirely by a single complex number. The result is the phasor field (complex amplitude) at the virtual aperture, which is propagated back into the scene using an RSD propagator.

Solving the RSD propagator numerically for each voxel in the scene would be computationally expensive. For a voxel space of side-length N, and $N*N$ points $\mathbf{x}_c$, the complexity is N^5. However, there are multiple algorithms that solve the RSD integral for a plane of voxels as a 2D convolution. For all planes making up the reconstruction space this results in a much lower complexity of $N^3 \log(N)$. While there are efficient solvers for the exact RSD[7], we rely on the well-known Fresnel approximation[4], to implement an efficient solver.

The Fresnel diffraction from a source plane S to a parallel destination plane D at distance z can be approximated as

$$\mathcal{P}(u_d, v_d, z) \approx \gamma \frac{e^{ikz}}{z} \iint_S \mathcal{P}(u_s, v_s) e^{ik\frac{(u_d-u_s)^2+(v_d-v_s)^2}{2z}} du_s dv_s, \tag{S.17}$$

where u and v are plane coordinates, and subscripts s and d refer to the coordinates in the source and destination planes, so that $\mathbf{x}_s = [u_s, v_s, 0] \in S$ and $\mathbf{x}_d = [u_d, v_d, z] \in D$. This can be interpreted as a 2D spatial convolution with a kernel $K(u, v) = \gamma \frac{e^{ikz}}{z} e^{ik\frac{u^2+v^2}{2z}}$.

This approximation can be used for the RSD propagator in all our camera operators. The criteria for the validity of the Fresnel approximation is well known[4] and given by

$$\frac{d^4}{4L^3\lambda} << 1, \tag{S.18}$$

where d is the effective aperture radius of the virtual camera, L is the distance between the aperture and the focal plane, and λ is the wavelength.

D.2 Confocal Time-gated System using Backprojection

Plugging in the corresponding terms from Tables S.2, and S.3 for the confocal time-gated imaging system in Equation S.15 we obtain

$$\begin{aligned} I(\mathbf{x}_v) &= \Phi\left(\int_P \int_{-\infty}^{+\infty} e^{i\omega(t-\tau-\frac{1}{c}|\mathbf{x}_v-\mathbf{x}_p|)} e^{-\frac{(t-\tau-t_0-\frac{1}{c}|\mathbf{x}_v-\mathbf{x}_p|)^2}{2\sigma^2}} H(\mathbf{x}_p \rightarrow \mathbf{x}_c, \tau) d\tau \, d\mathbf{x}_p\right) \\ &= \left|\int_C \int_P \int_{-\infty}^{+\infty} e^{i\omega(-\frac{1}{c}|\mathbf{x}_v-\mathbf{x}_c|-\tau-\frac{1}{c}|\mathbf{x}_v-\mathbf{x}_p|)} e^{-\frac{(-\frac{1}{c}|\mathbf{x}_v-\mathbf{x}_c|-\tau-t_0-\frac{1}{c}|\mathbf{x}_v-\mathbf{x}_p|)^2}{2\sigma^2}} H(\mathbf{x}_p \rightarrow \mathbf{x}_c, \tau) d\tau \, d\mathbf{x}_p \, d\mathbf{x}_c\right|^2. \end{aligned} \tag{S.19}$$

There are multiple ways of solving this expression. We can simply numerically compute the integrals, or we can re-write the expression to include backprojection or diffraction operators. This is desirable since fast methods to execute these operators exist.

Let us first write the expression as a backprojection. We introduce a shifted time $t_s = -\frac{1}{c}|\mathbf{x}_v - \mathbf{x}_c| - \frac{1}{c}|\mathbf{x}_v - \mathbf{x}_p|$ to obtain

$$\Phi(\mathcal{P}(\mathbf{x}_c, t)) = \left|\int_C \int_P \int_{-\infty}^{+\infty} e^{i\omega(t_s-\tau)} e^{-\frac{(t_s-\tau-t_0)^2}{2\sigma^2}} H(\mathbf{x}_p \rightarrow \mathbf{x}_c, \tau) d\tau \, d\mathbf{x}_p \, d\mathbf{x}_c\right|^2. \tag{S.20}$$

We break up this expression into two steps. First we perform a convolution on all the collected time responses in H to obtain an intermediate result $H'(\mathbf{x}_p \rightarrow \mathbf{x}_c, t)$

$$H'(\mathbf{x}_p \rightarrow \mathbf{x}_c, t) = (e^{i\omega t} e^{-\frac{(t'-t_0)^2}{2\sigma^2}}) \star H(\mathbf{x}_p \rightarrow \mathbf{x}_c, t'), \tag{S.21}$$

followed by shifting and summing the results:

$$\begin{aligned} I(\mathbf{x}_v) &= \left|\int_C \int_P H'(\mathbf{x}_p \rightarrow \mathbf{x}_c, -\frac{1}{c}|\mathbf{x}_v - \mathbf{x}_c| - \frac{1}{c}|\mathbf{x}_v - \mathbf{x}_p|) d\mathbf{x}_p \, d\mathbf{x}_c\right|^2 \\ &\approx \left|\sum_{\mathbf{x}_c \in C} \sum_{\mathbf{x}_p \in P} H'(\mathbf{x}_p \rightarrow \mathbf{x}_c, -\frac{1}{c}|\mathbf{x}_v - \mathbf{x}_c| - \frac{1}{c}|\mathbf{x}_v - \mathbf{x}_p|)\right|^2. \end{aligned} \tag{S.22}$$

where the second term is a backprojection, for which efficient implementations exist[1].

D.3 Transient Camera using Backprojection

Last, we derive our NLOS transient system. Operating similarly to the confocal time-gated system, by plugging in the corresponding terms from Tables S.2, and S.3 in Equation S.15 we obtain:

$$H'(\mathbf{x}_p \rightarrow \mathbf{x}_c, t) = (e^{i\omega t}\delta(\mathbf{x}_p - \mathbf{x}_{ls})e^{-\frac{(t'-t_0)^2}{2\sigma^2}}) \star H(\mathbf{x}_p \rightarrow \mathbf{x}_c, t') \tag{S.23}$$

and

$$I(\mathbf{x}_v, t) = \left| \int_C H'(\mathbf{x}_p \rightarrow \mathbf{x}_c, t - \frac{1}{c}|\mathbf{x}_v - \mathbf{x}_c| - \frac{1}{c}|\mathbf{x}_v - \mathbf{x}_{ls}|)d\mathbf{x}_c \right|^2 . \tag{S.24}$$

Besides the use of only one illumination point $\mathbf{x}_p = \mathbf{x}_{ls}$, this reconstruction differs from the confocal system in that it depends on time t. The reconstruction is 4-dimensional, resulting in a video of the light propagation in the 3D reconstruction space. To reduce computational cost, we can optionally locate empty voxels by first using our confocal imaging functions.

References

1. Arellano, V., Gutierrez, D., & Jarabo, A.. Fast back-projection for non-line of sight reconstruction. *Optics Express* **25**, 11574-11583 (2017).

2. O'Toole, M., Lindell, D. B., & Wetzstein, G. Confocal non-line-of-sight imaging based on the light-cone transform. *Nature* **555**, 338 (2018).

3. Heide, F., O'Toole, M., Zang, K., Lindell, D., Diamond, S., & Wetzstein, G. Non-line-of-sight Imaging with Partial Occluders and Surface Normals. *ACM Transactions on Graphics* (2019).

4. Goodman, J. *Introduction to Fourier optics* 3rd edn (Roberts and Company Publishers, 2005).

5. Velten, A. et al. Femto-photography: capturing and visualizing the propagation of light. *ACM Transactions on Graphics* **32**, 44 (2013).

6. Wu, D. et al. Decomposing Global Light Transport Using Time of Flight Imaging. *International Journal of Computer Vision* **107(2)** 123-138 (2014).

7. Shen, F. & Wang, A. Fast-fourier-transform based numerical integration method for the Rayleigh-Sommerfeld diffraction formula. *Applied optics* **45**, 6. (2006).