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

B Derivation of the phasor field

C LOS template functions

D Implementation details of the RSD solvers

A Overview of the Method

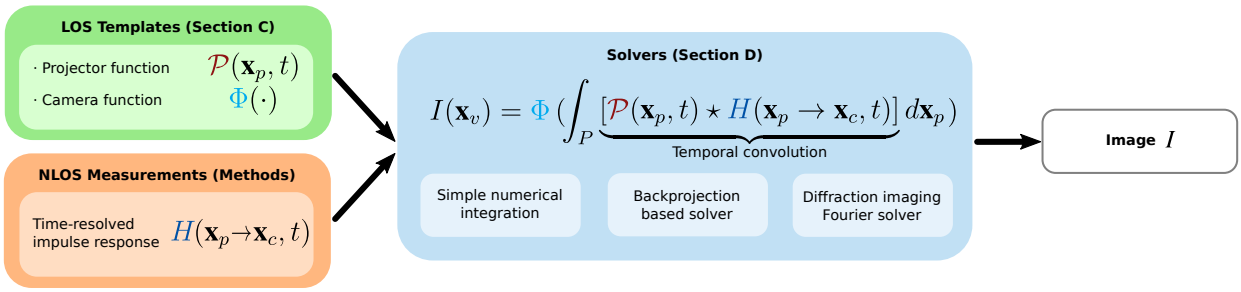

Figure S.1: Our imaging framework: The flowchart describes how to image an NLOS scene from a chosen template LOS system. Please refer to the text for details.

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.

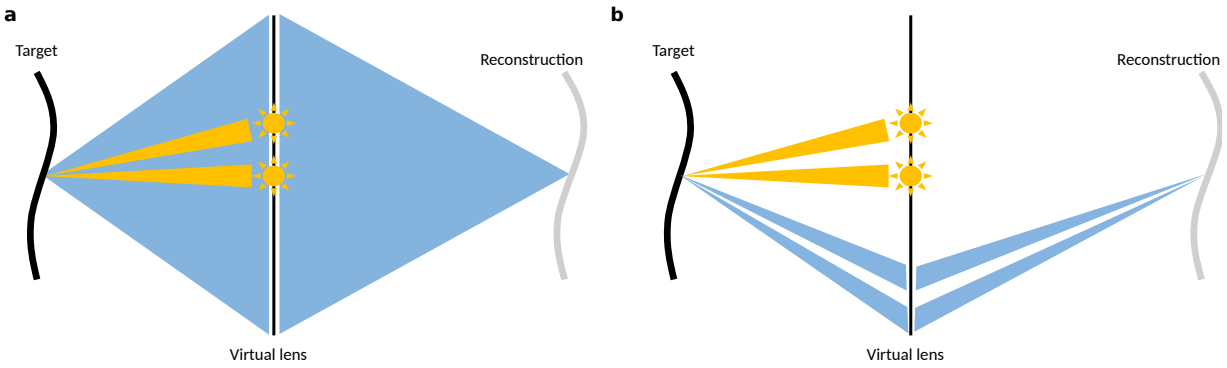

Figure S.2: Light hitting the relay wall (in yellow) illuminates the target scene, and is scattered and bounced back by unknown objects in the scene (in blue). **a)** When the target object is perfectly Lambertian, it bounces light back to the entire captured surface of the relay (our virtual lens); 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).