doi:10.1038/nature25489

Supplementary Methods

Equipment details The time-resolved detector is a PDM series single photon avalanche diode (SPAD) from Micro Photon Devices with a $100\,\mu\mathrm{m} \times 100\,\mu\mathrm{m}$ active area, a reported $27\,\mathrm{ps}$ timing jitter ($100\,\mathrm{kHz}$ laser at $675\,\mathrm{nm}$), and 40.9 dark counts per second. Detection events are time stamped with $4\,\mathrm{ps}$ temporal resolution using a PicoHarp 300 Time-Correlated Single Photon Counting (TCSPC) module. The detected light is focused by a $75\,\mathrm{mm}$ achromatic doublet (Thorlabs AC254-075-A-ML), and filtered using a laser line filter (Thorlabs FL670-10). The detection optics are co-axially aligned with an active light source using a polarized beamsplitter (Thorlabs PBS251). The light source (ALPHALAS PICOPOWER-LD-670-50) consists of a $670\,\mathrm{nm}$ wavelength pulsed laser diode with a reported pulse width of $30.6\,\mathrm{ps}$ at a $10\,\mathrm{MHz}$ repetition rate and $0.11\,\mathrm{mW}$ average power. A 2-axis scanning galvonometer (Thorlabs GVS012) raster scans the illumination and detection spots across a wall located approximately $2\,\mathrm{m}$ from the system at an oblique angle. The measured jitter of the entire system is approximately $60\,\mathrm{ps}$ without the laser line filter, and $200\,\mathrm{ps}$ with the filter on. We only use the filter for the outdoor experiments to reduce ambient light. The increase in jitter is due to the fact that the spectral transmission properties of the filter affect the temporal characteristics of the imaging system via the time-bandwidth product.[16] An appropriate choice of spectral line filter and pulsed laser could mitigate this effect.

Challenges with confocal scanning Unlike conventional approaches, a confocalized NLOS system exposes the detector to direct reflections. This can be problematic, because the overwhelmingly bright contribution of direct light reduces the SNR of the indirect signal in two ways. First, after detecting a photon, the SPAD sensor becomes inactive for approximately $75\,\mathrm{ns}$ and ignores any photons that strike the detector for this period of time (commonly referred to as the dead time of the device). If the contribution of direct light is too strong, this reduces the detection probability of indirect photons. Second, approximately 0.1% of detected photon events produce a secondary event due to an effect known as afterpulsing. The contribution of direct photons therefore increases the number of spurious photons detected by the SPAD, further reducing the SNR of the indirect signal.

To avoid the negative effects of a strong direct signal, a time-gated SPAD could be used for detection to gate out photons due to direct light.[16] Given that our SPAD operates in free-running mode, we instead illuminate and image two slightly different points on a wall to reduce the contribution of direct light. The distance between these points should be sufficiently small so as to not affect the confocal image formation model.

System calibration The first calibration step involves aligning the detector and light source by adjusting the position of the beamsplitter to maximize the photon count rate. When perfectly aligned, the strong direct signal significantly reduces the number of indirect photons detected by the SPAD. Therefore, the second step involves slightly adjusting the position of the beamsplitter to decrease the direct photon counts and increase the indirect

photon counts originating from a hidden retroreflector placed within the scene (e.g., the exit sign). The SPAD detected between 0.29 and 1 million counts per second for all experiments (i.e., the number of detected events ranged between 2.9% and 10% of the total number of pulses). For the final step, the system scans a 6×6 grid of points on the wall and uses the time of arrival of direct photons and known galvanometer mirror angles to compute the relative position and orientation of the wall relative to the system.

Acquisition procedure The system scans 64×64 equidistant points on a wall for indoor experiments, and 32×32 points for the outdoor experiment. At a $10\,\mathrm{MHz}$ repetition rate, the PicoHarp 300 returns unprocessed histograms containing $25,000$ bins with a temporal resolution of $4\,\mathrm{ps}$ per bin. The acquired histograms are temporally aligned such that the direct pulses appear at time $t = 0$. Histograms are then trimmed to either 2048 bins for indoor experiments or 4096 bins for the outdoor experiment. The first 600 bins are set to 0 to remove the direct component. The histograms are then downsampled by a factor of 4 to either 512 or 1024 bins before processing, where each bin now has a temporal resolution of $16\,\mathrm{ps}$. The acquisition time for each histogram is either $0.1\,\mathrm{s}$ or $1\,\mathrm{s}$, as indicated for each respective experiment.

Validating radiometric falloff To verify the radiometric intensity falloff in the proposed image formation model, we measure the intensity of a small patch behind the wall while varying the distance between the NLOS patch and the sampled wall. Figure 1 shows the intensity response for several different materials: a diffuse patch and retroreflective patches of different grades ("engineering" grade and "diamond" grade). For the diffuse patch, measurements (blue circles) closely match the predicted $\frac{1}{r^4}$ falloff (blue line), where r is the distance between patch and wall. Similarly, the diamond grade retroreflective material (red circles) closely matches the predicted $\frac{1}{r^2}$ falloff (red line). Lower-quality retroreflectors, such as the engineering grade retroreflective material or most retroreflective paints, exhibit a falloff that is somewhere between diffuse and perfectly retroreflective. The engineering grade retroreflective material (green circles), for example, can be modeled by a falloff term of $\frac{1}{r^{2.3}}$ (green line).

Note that the intensity falloff only includes the propagation distance between the visible wall and the hidden surface as well as the distance back along the same path. The additional travel distance between the laser/detector and the wall is fixed, and does not affect the intensity falloff rate.

Image formation We briefly review the conventional NLOS formulation that models the indirect light transport that occurs between two different points on a wall. Several simplifying assumptions are made in the derivation of this image formation model (see below), resulting in an approximation of the physical light transport process. We then introduce the confocal NLOS image formation model, followed by the light cone transform and a discretization of the proposed model.

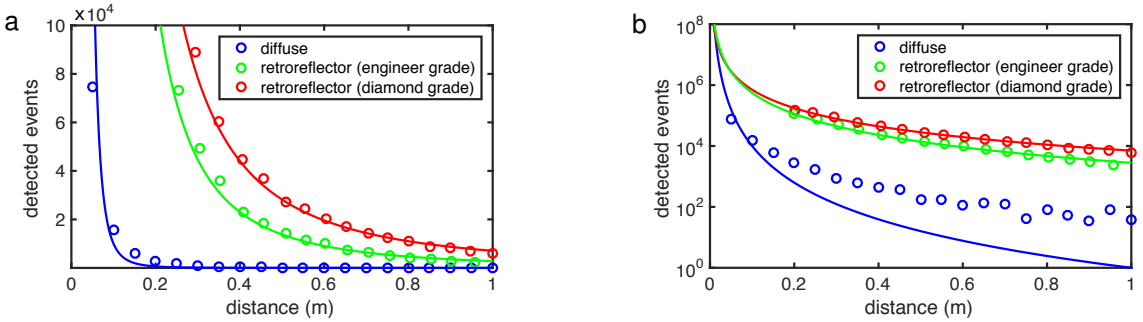

Figure 1: (a) A linear plot and (b) a semi-logarithmic plot of the number of indirect photons detected as a function of an object's distance from the wall. The target objects are $5\,\mathrm{cm} \times 5\,\mathrm{cm}$ squares with different reflectance properties. The number of events is approximated by the function $\frac{1}{r^4}$ for the diffuse object (blue line), $\frac{1}{r^{2.3}}$ for the engineer grade retroreflector (green line), and $\frac{1}{r^2}$ for the diamond grade retroreflector (red line).

Conventional non-line-of-sight imaging

As illustrated in Figure 2, conventional NLOS imaging records a transient image[14,25–27] of a flat wall with a time-resolved detector while sequentially illuminating points on the wall with an ultra-short laser pulse.[14–19] The geometry and albedo of the wall is assumed to be known or it can be scanned in a pre-processing step. Without loss of generality, we model the wall as a reference plane at position $z = 0$. The recorded transient image τ is

$$\tau\left(x', y', t\right) = \iiint_{\Omega} \frac{1}{r_l^2 r^2}\, \rho\left(x, y, z\right) \times \delta\left(\sqrt{\left(x'-x\right)^2 + \left(y'-y\right)^2 + z^2} + \sqrt{\left(x_l-x\right)^2 + \left(y_l-y\right)^2 + z^2} - tc\right)\, dx\, dy\, dz. \tag{5}$$

Here, ρ is the sought-after albedo of the hidden scene at each point in the three-dimensional half-space Ω satisfying $z > 0$. The transient image is recorded while the light source illuminates position x_l, y_l on the wall with an ultra-short pulse. This pulse is diffusely reflected off the wall and then scattered by the hidden scene back towards the wall. The radiometric term $1/r_l^2 r^2$ models the square distance falloff using the distance r_l between x_l, y_l and some hidden scene point x, y, z, as well as the distance r from that point to the sampled detector position on the wall x', y'. This equation can be discretized as $\boldsymbol{\tau} = \mathbf{A}\boldsymbol{\rho}$ and solved with an iterative numerical approach that does not require $\mathbf{A}$ to be directly inverted.

The conventional NLOS image formation model makes the following assumptions: there is only single scattering behind the wall (i.e., no inter-reflections in the hidden scene parts), and there are no occlusions between hidden scene parts. We also assume surfaces reflect light isotropically (i.e., the ratio of reflected radiance is independent of both the incident light direction and outgoing view direction), in order to avoid the added complexity of introducing Lambert's cosine terms to model diffuse surface reflections.

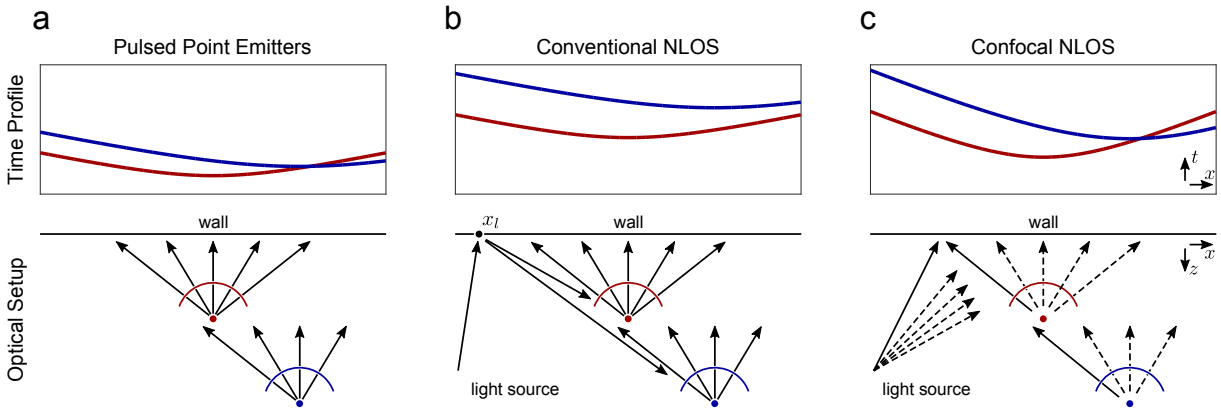

Figure 2: Illustration of time profiles incident on a wall from two scatterers that emit or reflect ultra-short pulses of light. (a) The two points emit a single pulse at the same time. (b) Conventional NLOS imaging scans several combinations of light source positions x_l, y_l and detector positions on the wall; here, this is illustrated for one light source position. (c) Confocal NLOS scans over the wall once with a co-located illuminator and detector. Note that the time profiles of the confocal setup are identical to those of point emitters after scaling the arrival times by a factor of two.

Confocal non-line-of-sight imaging

Instead of exhaustively scanning different combinations of light source positions x_l, y_l and detector positions x', y' on the wall, confocal NLOS imaging is a sequential scanning approach where the light source and a single detector are co-axial, or *"confocalized"*. Data recorded with a confocal NLOS setup thus represents a subset of the samples required by conventional NLOS imaging. One of the primary benefits of the confocal setup is that it is consistent with existing scanned LIDAR systems that often use avalanche photodiodes (APDs) or single photon avalanche diodes (SPADs). The proposed signal processing approach to NLOS imaging may therefore be compatible with many existing scanners.

Here, the transient image on the wall is given by

$$\tau\left(x', y', t\right) = \iiint_{\Omega} \frac{1}{r^4}\, \rho\left(x, y, z\right)\, \delta\left(2\sqrt{\left(x'-x\right)^2 + \left(y'-y\right)^2 + z^2} - tc\right)\, dx\, dy\, dz. \tag{6}$$

This image formation model shares the same assumptions as Equation (5) (i.e., no multi-bounce transport, no occlusions, and isotropic scattering).

Equation (6) is a laterally (i.e., in x and y) shift-invariant convolution. The convolution kernel is the surface of a spatio-temporal 4D hypercone

$$x^2 + y^2 + z^2 - \left(\frac{tc}{2}\right)^2 = 0. \tag{7}$$

This formulation for light propagation is similar to Minkowski's light cone used in special relativity,[28] except that it models a spherical wavefront propagating at half the speed of light.

Dirac delta identity

The image formation model of Equation (6) can be rewritten in a more convenient form by squaring and scaling the arguments of the Dirac delta with the following identity:

$$\int_{\Omega} f(\mathbf{x})\ \delta\left(2\|\mathbf{x}'-\mathbf{x}\|_2 - tc\right)\ d\mathbf{x} = \int_{\Omega} f(\mathbf{x})\ \|\mathbf{x}'-\mathbf{x}\|_2\ \delta\quad \|\mathbf{x}'-\mathbf{x}\|_2^2 - \left(\frac{tc}{2}\right)^2\Bigg)\ d\mathbf{x} \tag{8}$$

where we denote $\mathbf{x} = (x, y, z)$ and $\mathbf{x}' = (x', y', 0)$ for simplicity.

Proof The Dirac delta function can be expressed as the limit of a sequence of normalized functions[30]

$$\delta(x) = \lim_{\epsilon\to 0^+} \kappa_\epsilon(x) = \lim_{\epsilon\to 0^+} \frac{1}{\epsilon}\kappa\left(\frac{x}{\epsilon}\right)\ . \tag{9}$$

For example, the limit of a sequence of normalized hat functions (i.e., $\kappa(x) = \max(1-|x|, 0)$) is the Dirac delta function. The following uses this definition and a change of variables, $\epsilon = \epsilon' \frac{4}{2\|\mathbf{x}'-\mathbf{x}\|_2 + tc}$, to derive Equation (8):

$$\begin{aligned}
\int_{\Omega} f(\mathbf{x})\ \delta\left(2\|\mathbf{x}'-\mathbf{x}\|_2 - tc\right)\ d\mathbf{x} &= \int_{\Omega} f(\mathbf{x})\ \lim_{\epsilon\to 0^+}\frac{1}{\epsilon}\kappa\left(\frac{2\|\mathbf{x}'-\mathbf{x}\|_2 - tc}{\epsilon}\right)\ d\mathbf{x} \\
&= \int_{\Omega} f(\mathbf{x})\ \lim_{\epsilon'\to 0^+}\frac{2\|\mathbf{x}'-\mathbf{x}\|_2 + tc}{4\epsilon'}\ \kappa\left(\frac{4\|\mathbf{x}'-\mathbf{x}\|_2^2 - (tc)^2}{4\epsilon'}\right)\ d\mathbf{x} \\
&= \int_{\Omega} f(\mathbf{x})\ \left(\frac{2\|\mathbf{x}'-\mathbf{x}\|_2 + tc}{4}\right)\ \lim_{\epsilon'\to 0^+}\frac{1}{\epsilon'}\ \kappa\left(\frac{\|\mathbf{x}'-\mathbf{x}\|_2^2 - \left(\frac{tc}{2}\right)^2}{\epsilon'}\right)\ d\mathbf{x} \\
&= \int_{\Omega} f(\mathbf{x})\ \left(\frac{2\|\mathbf{x}'-\mathbf{x}\|_2 + tc}{4}\right)\ \delta\left(\|\mathbf{x}'-\mathbf{x}\|_2^2 - \left(\frac{tc}{2}\right)^2\right)\ d\mathbf{x} \\
&= \int_{\Omega} f(\mathbf{x})\ \|\mathbf{x}'-\mathbf{x}\|_2\ \delta\quad \|\mathbf{x}'-\mathbf{x}\|_2^2 - \left(\frac{tc}{2}\right)^2\Bigg)\ d\mathbf{x}\quad \blacksquare
\end{aligned} \tag{10}$$

Radiometric considerations

One of several interesting and unique properties of confocal NLOS imaging is that the distance function r is directly related to the measured time-of-flight as

$$\frac{2}{c}\underbrace{\sqrt{(x'-x)^2 + (y'-y)^2 + z^2}}_{r} = t \quad \Leftrightarrow \quad r = \frac{tc}{2}. \tag{11}$$

Therefore, the corresponding radiometric term, $\frac{1}{r^4}$, can be pulled out of the triple integral of Equation (6). For conventional NLOS imaging, we only know the combined distance $r_l + r = tc$, but we cannot easily use this information to replace the radiometric falloff term $1/\left(r_l^2 r^2\right)$ in Equation (5).

Another important property is that retroreflective materials can be modeled by replacing the radiometric falloff term $\frac{1}{r^4}$ with $\frac{1}{r^2}$, signifying a drastic increase in the indirect light signal as a function of distance r. Retroreflective materials cannot be handled appropriately by existing non-confocal NLOS methods.

The light cone transform

We propose the *light cone transform* (LCT) that expresses the confocal NLOS image formation model as a shift-invariant 3D convolution in the transform domain. The LCT is a computationally efficient way for computing the forward model and, more importantly, leads to a closed-form expression for the inverse problem.

We start by using the Dirac delta identity from Equation (8) to rewrite the image formation model in Equation (6) as

$$\tau\left(x', y', t\right) = \iiint_{\Omega} \frac{1}{r^3}\, \rho\left(x, y, z\right)\, \delta\left(\left(x'-x\right)^2 + \left(y'-y\right)^2 + z^2 - \left(\frac{tc}{2}\right)^2\right) dx\, dy\, dz. \tag{12}$$

Next, we pull out the radiometric term from the integral and perform a change of variables by letting $z = \sqrt{u}$, $\frac{dz}{du} = \frac{1}{2\sqrt{u}}$ such that

$$\tau\left(x', y', t\right) = \left(\frac{2}{tc}\right)^3 \iiint_{\Omega} \rho\left(x, y, \sqrt{u}\right)\, \delta\left(\left(x'-x\right)^2 + \left(y'-y\right)^2 + u - \left(\frac{tc}{2}\right)^2\right) \frac{1}{2\sqrt{u}}\, dx\, dy\, du. \tag{13}$$

We also introduce a second change of variables using $v = \left(\frac{tc}{2}\right)^2$, such that

$$\underbrace{v^{\frac{3}{2}}\, \tau\left(x', y', \frac{2}{c}\sqrt{v}\right)}_{\mathcal{R}_t\{\tau\}(x',y',v)} = \iiint_{\Omega} \underbrace{\frac{1}{2\sqrt{u}}\, \rho\left(x, y, \sqrt{u}\right)}_{\mathcal{R}_z\{\rho\}(x,y,u)}\, \underbrace{\delta\left(\left(x'-x\right)^2 + \left(y'-y\right)^2 + u - v\right)}_{h(x'-x, y'-y, v-u)}\, dx\, dy\, du. \tag{14}$$

The image formation model is a 3D convolution, which can alternatively be written as

$$\mathcal{R}_t\left\{\tau\right\} = h * \mathcal{R}_z\left\{\rho\right\}, \tag{15}$$

where $*$ is the 3D convolution operator, h is the shift-invariant convolution kernel, $\mathcal{R}_z\left\{\cdot\right\}$ resamples ρ along the z-axis and attenuates the result by $1/2\sqrt{u}$, and $\mathcal{R}_t\left\{\cdot\right\}$ resamples τ along the time axis and scales the result by $v^{\frac{3}{2}}$. Note that a similar transform is not known to exist for the conventional NLOS problem, and that the LCT is specific to the confocal case.

Discretizing the image formation

The transforms introduced with the continuous image formation model (Equation (15)) are implemented as discrete operations in practice. For example, the operation $\mathcal{R}_t\left\{\tau\right\}$ can be represented as an integral transform

$$\mathcal{R}_t\left\{\tau\right\}\left(x', y', v\right) = v^{\frac{3}{2}}\, \tau\left(x', y', \frac{2}{c}\sqrt{v}\right) = \int_{t>0} v^{\frac{3}{2}}\, \delta\left(t - \frac{2}{c}\sqrt{v}\right)\, \tau(x', y', t)\, dt\,. \tag{16}$$

Note that this transformation applies to all points (x', y') independently.

The discrete analog of this transform is given by a matrix-vector multiplication $\mathbf{R}_t\boldsymbol{\tau}$, between the vectorized representation of the transient image $\boldsymbol{\tau} \in \mathbb{R}_+^{n_x n_y n_t}$ and matrix $\mathbf{R}_t \in \mathbb{R}_+^{n_x n_y n_h \times n_x n_y n_t}$. Consider the case of a

single measurement on the wall (i.e., $n_x = 1$ and $n_y = 1$), where Ω_{xy} is the region sampled by the detector. The individual elements of the vectorized transient image and corresponding transform matrix are then given by

$$(\boldsymbol{\tau})_j = \iint_{\Omega_{xy}} \int_{t_{j-1}}^{t_j} \tau(x', y', t)\; dt\, dx'\, dy'\,, \quad (\mathbf{R}_t)_{ij} = \int_{h_{i-1}}^{h_i} \int_{t_{j-1}}^{t_j} v^{\frac{3}{2}}\; \delta\left(t - \frac{2}{c}\sqrt{v}\right)\; dt\, dv\,, \tag{17}$$

where $1 \leq i \leq n_h$ and $1 \leq j \leq n_t$. Here, the transient image is defined over a range of time values $[a, b] \subset (0, \infty)$, which is uniformly discretized into n_t equal spaces such that $a = t_0 < t_1 < \cdots < t_{n_t} = b$. Similarly, the matrix $\mathbf{R}_t$ resamples the transient image into n_h elements where $\left(\frac{ca}{2}\right)^2 = h_0 < h_1 < \cdots < h_{n_h} = \left(\frac{cb}{2}\right)^2$.

The corresponding discrete analog of the transformation $\mathcal{R}_z\left\{\rho\right\}$ is similarly defined as a matrix-vector product $\mathbf{R}_z\boldsymbol{\rho}$, between the vectorized representation of unknown surface albedos $\boldsymbol{\rho} \in \mathbb{R}_+^{n_x n_y n_z}$ and matrix $\mathbf{R}_z \in \mathbb{R}_+^{n_x n_y n_h \times n_x n_y n_z}$. The elements are defined as

$$(\boldsymbol{\rho})_k = \iint_{\Omega_{xy}} \int_{z_{k-1}}^{z_k} \rho(x, y, z)\; dz\, dx\, dy\,, \quad (\mathbf{R}_z)_{ik} = \int_{h_{i-1}}^{h_i} \int_{z_{k-1}}^{z_k} \frac{1}{2\sqrt{u}}\; \delta\left(z - \sqrt{u}\right)\; dz\, du\,, \tag{18}$$

where $1 \leq k \leq n_z$ and $\frac{ca}{2} = z_0 < z_1 < \cdots < z_{n_z} = \frac{cb}{2}$.

The full discrete image formation model is therefore

$$\boldsymbol{\tau} = \mathbf{A}\boldsymbol{\rho} = \mathbf{R}_t^{-1}\mathbf{H}\mathbf{R}_z\boldsymbol{\rho}\,. \tag{19}$$

where the matrix $\mathbf{A} = \mathbf{R}_t^{-1}\mathbf{H}\mathbf{R}_z$ is referred to as the light transport matrix. Note that each of these matrices is independently applied to the respective dimension and can therefore be applied to large-scale datasets in a memory efficient way. The matrix $\mathbf{H} \in \mathbb{R}_+^{n_x n_y n_h \times n_x n_y n_h}$ represents the shift-invariant 3D convolution with the 4D hypercone (i.e., a convolution with a discretized version of the kernel h), which models light transport in free space in the transform domain. Together, these matrices represent the *discrete light cone transform*.

The discrete light cone transform provides a fast and memory efficient approach to computing both forward light transport (i.e., $\mathbf{A}\boldsymbol{\rho}$) and inverse light transport (i.e., $\mathbf{A}^{-1}\boldsymbol{\rho}$) without forming any of the matrices explicitly. Computational efficiency is largely achieved using the convolution theorem to compute matrix-vector multiplications with $\mathbf{H}$ as element-wise multiplications in the Fourier domain. Similarly, matrix-vector multiplications with their inverses are computed as element-wise divisions in the Fourier domain.

Inverse methods Here, we derive a closed-form solution for the discrete NLOS problem. We first assume that the noise model associated with the discrete light transform model in Equation (19) satisfies

$$\tilde{\boldsymbol{\tau}} = \mathbf{H}\tilde{\boldsymbol{\rho}} + \boldsymbol{\eta}, \tag{20}$$

where $\tilde{\boldsymbol{\tau}} = \mathbf{R}_t\boldsymbol{\tau}$, $\tilde{\boldsymbol{\rho}} = \mathbf{R}_z\boldsymbol{\rho}$, and $\boldsymbol{\eta} \in \mathbb{R}^{n_x n_y n_h}$ is white noise.

The solution $\tilde{\boldsymbol{\rho}}_*$ that minimizes the mean square error with respect to the ground truth solution $\tilde{\boldsymbol{\rho}}$ is well known to be given by the Wiener deconvolution filter:[29]

$$\tilde{\boldsymbol{\rho}}_* = \mathbf{F}^{-1}\left[\frac{1}{\widehat{\mathbf{H}}}\cdot\frac{|\widehat{\mathbf{H}}|^2}{|\widehat{\mathbf{H}}|^2+\frac{1}{\alpha}}\right]\mathbf{F}\tilde{\boldsymbol{\tau}}, \tag{21}$$

where the matrix $\mathbf{F}$ represents the 3D discrete Fourier transform and $\widehat{\mathbf{H}}$ is a diagonal matrix containing the Fourier transform of the shift-invariant 3D convolution kernel. α is a frequency-dependent term representing the signal-to-noise ratio (SNR).

Expanding Equation (21) results in the closed-form solution for the confocal NLOS problem:

$$\boldsymbol{\rho}_* = \mathbf{R}_{\tilde{z}}^{-1}\mathbf{F}^{-1}\left[\frac{1}{\widehat{\mathbf{H}}}\cdot\frac{|\widehat{\mathbf{H}}|^2}{|\widehat{\mathbf{H}}|^2+\frac{1}{\alpha}}\right]\mathbf{F}\mathbf{R}_t\boldsymbol{\tau}. \tag{22}$$

In the Supplementary Derivations, we also outline a maximum a posteriori estimator for the reconstruction problem that lifts these assumptions on the noise model and that also allows image priors to be imposed on the reconstructed volume.

Supplementary Discussion

Resolution limits The accuracy of the reconstructed shape depends on several factors, including the system jitter, the scanning area $2w \times 2w$, and the distance of the hidden object from the wall z (Figure 3). Here, the resolution limit is defined as the minimum resolvable distance of two scatterers Δd. Specifically, two scattering points, $\mathbf{q}_1$ and $\mathbf{q}_2$, are resolvable in space only if their indirect signals are resolvable in time:

$$\Delta d = \text{abs}\left(\|\mathbf{p}-\mathbf{q}_1\|_2 - \|\mathbf{p}-\mathbf{q}_2\|_2\right) \geq \frac{c\Delta t}{2} \tag{23}$$

Using this formula, a bound on the minimum axial distance, i.e., along the z-axis, can be defined as $\Delta z \geq c\Delta t/2$. In practice, this bound is related to the full width at half maximum (FWHM) of the temporal jitter of the detection system, represented by the scalar γ, as

$$\Delta z \geq \frac{c\gamma}{2} \tag{24}$$

As illustrated in Figure 3, the FWHM is a practical means to derive resolution limits of a NLOS imaging system and it is closely related to diffraction-limited resolution limits in microscopy. Alternative resolution criteria include the Rayleigh criterion and the Sparrow limit. However, due to the fact that the temporal point spread functions in NLOS imaging are not Airy disks, the FWHM criterion is an appropriate formulation for this application. If the temporal jitter follows a Gaussian shape, the FWHM is directly proportional to the standard deviation σ of the jitter as $\gamma = 2\sqrt{2\ln 2}\sigma$.

Similarly, we can derive an approximate bound on the lateral resolution Δx. For this purpose, we assume that Δx is much smaller than the area of the wall being sampled, such that $\cos(\theta) \approx c\Delta t/\left(2\Delta x\right)$ (see Figure 3(d)).

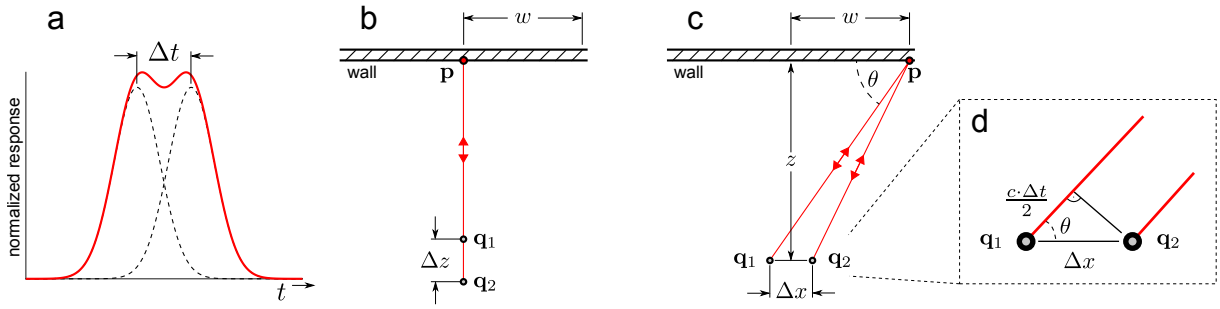

Figure 3: (a) Illustration of the temporal response of two hidden points measured at one location on a visible surface. Assuming that the jitter of the detection system is well-described by a Gaussian function, the difference in time of arrival Δt of the response of two points has to be farther apart than the FWHM γ of the Gaussian jitter, i.e., $\Delta t \geq \gamma$. Panel (a) illustrates the case where $\Delta t = \gamma$, which we consider the minimum resolvable difference. To derive bounds on (b) axial and (c, d) lateral resolvable feature size, Δt is related to the axial (Δz) and lateral (Δx) difference in position of the two points $\mathbf{q}_{1/2}$. In this illustration, $\mathbf{p}$ indicates the scanned point on the visible wall that carries the most amount of information for resolving the difference between the hidden points $\mathbf{q}_{1/2}$. As shown in panels (c) and (d), the lateral resolution is mainly determined by the size of the sampled wall $2w$ and the distance of the point to the wall z.

Then,

$$\frac{c\Delta t}{2} \approx \Delta x \cos(\theta) \approx \Delta x \cos\left(\tan^{-1}\left(\frac{z}{w}\right)\right) \quad (25)$$

$$= \Delta x \frac{w}{\sqrt{w^2+z^2}} \quad (26)$$

Using the FWHM criterion, the minimum lateral distance between two points is

$$\Delta x \geq \frac{c\sqrt{w^2+z^2}}{2w}\gamma \quad (27)$$

The formulation above allows us to make several important insights. First, the smallest resolvable axial feature size Δz represents the lowest resolution bound for all NLOS imaging applications. Second, the smallest resolvable lateral feature size Δx is depth dependent and only in the limit of $z = 0$ equal to Δz. For sufficiently large depth values, i.e., $z \gg w$, Δx increases linearly with z.

Note that these bounds not only apply to the confocal NLOS acquisition setup introduced in this paper, but also to all conventional NLOS approaches that scan different combinations of imaged and illuminated positions on the wall. The fundamental limit of resolvable feature size is dictated by the temporal resolution of the detection system and the size of the sampled area on the wall, which is also noted in prior work.[16] These insights are not surprising, because they directly relate to characteristics of diffraction-limited, coherent imaging systems, where resolution is governed by wavelength and numerical aperture.

We verify these theoretical bounds with two experiments using a planar resolution chart and a set of individual reflectors. Both targets are custom-made from retroreflective material and shown in Figure 4 (left column). For both experiments, the confocal laser scanner samples the wall at 64×64 uniformly-spaced locations over a total area of $2w \times 2w = 40\,\mathrm{cm} \times 40\,\mathrm{cm}$. The acquisition time per sampled location is constant (approx. $0.1\,\mathrm{s}$) for all experiments. The temporal uncertainty of the imaging system is $60\,\mathrm{ps}$, including SPAD jitter and finite laser pulse width. Using Equation (27), we predict lateral resolutions of approx. $2\,\mathrm{cm}$ and $3.1\,\mathrm{cm}$ when the NLOS target is $z = 40\,\mathrm{cm}$ and $z = 65\,\mathrm{cm}$ away from the wall, respectively.

The resolution chart consists of 5 groups of lines, and the width of the lines in each group is $3\,\mathrm{cm}$, $2\,\mathrm{cm}$, $1.5\,\mathrm{cm}$, $1\,\mathrm{cm}$, and $0.5\,\mathrm{cm}$. Based on our prediction, the first group should be just resolvable at a distance of $z = 65\,\mathrm{cm}$ and the second group should be just resolvable at $z = 40\,\mathrm{cm}$. Columns 2-5 of Figure 4 show maximum intensity projections of reconstructed volumes via front and side views at two different target distances. As predicted, group 1 is resolvable at the farther distance while moving the target closer to the wall allows for group 2 to be recovered as well.

The second experiment shows a different resolution target that consists of 3×3 circular retroreflectors, each with a diameter of $3.2\,\mathrm{cm}$, spaced apart by $10\,\mathrm{cm}$. As we move this target away from the wall, the resolution of the reconstructed target slightly degrades and it becomes more noisy due to the deceased signal-to-noise ratio in the measurements.

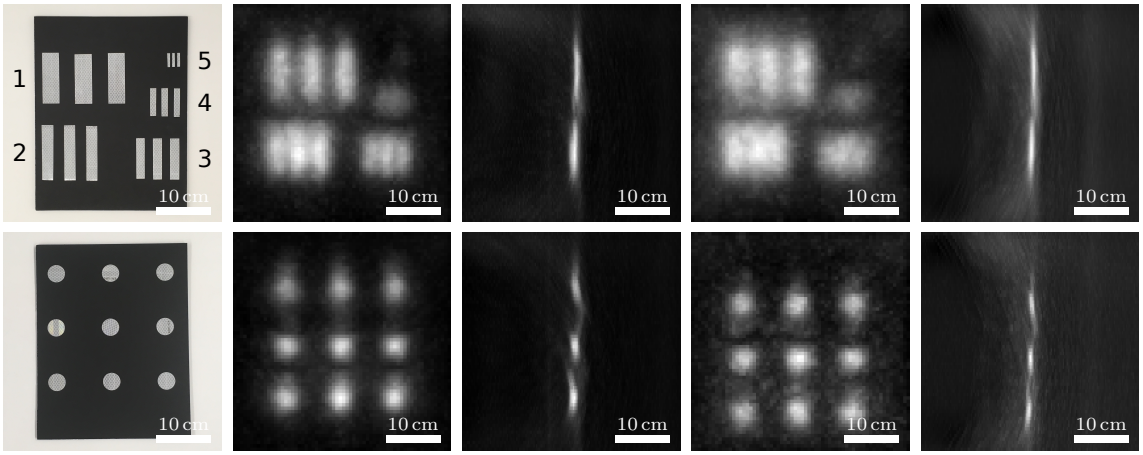

Figure 4: Experimental resolution analysis of confocal NLOS imaging. For this setup, the theoretical lateral resolution is $\Delta x \approx 2.0\,\mathrm{cm}$ for a plane at depth $z = 40\,\mathrm{cm}$, and $\Delta x \approx 3.1\,\mathrm{cm}$ at depth $z = 65\,\mathrm{cm}$. Images share the same scale. Row 1: This resolution chart has multiple groups of lines. According to the theoretical predictions, group 1 should be just resolvable at $z = 65\,\mathrm{cm}$ and group 2 at $z = 40\,\mathrm{cm}$, which is observed in practice. Row 2: The lateral resolution of a set of circular reflectors degrades slightly as the target moves away from the wall. Moreover, the signal-to-noise ratio of the measurements decreases and results in noisier reconstructions at farther distances.

Compute time and memory requirements Here, we list expected runtimes and memory requirements for several non-line-of-sight imaging approaches. The backprojection algorithm, a direct inverse of the light transport matrix, and iterative approaches to solving the inverse problem have been proposed in previous work. The discrete light cone transform introduced in this paper allows a direct inverse of the light transport matrix to be computed with orders of magnitude lower computational complexity and less memory requirements than existing methods. In the following, we assume for simplicity that the recorded histograms sample the hidden volume of resolution $N \times N \times N$ at $N \times N$ locations on the wall. Each of the measured locations is represented as a temporal histogram of length N. Thus, the light transport matrix $\mathbf{A}$ is of size $N^3 \times N^3$ and it contains N^5 non-zero elements that would have to be stored in a sparse matrix representation.

The backprojection algorithm is a matrix-vector product of the transpose light transport matrix and the measurement vector, which has a computational complexity of $O\left(N^5\right)$. The backprojection algorithm can be computed without explicitly constructing the matrix, and only requires $O\left(N^3\right)$ memory.

Computing the inverse light transport matrix, for example via the singular value decomposition, includes $O\left(N^9\right)$ computational complexity for computing the SVD as well as two additional matrix-vector multiplications and a vector-vector multiplication. The overall runtime is still on the order of $O\left(N^9\right)$ and the memory requirements just for storing the decompositions is $O\left(N^6\right)$. An attractive alternative to the SVD would be an iterative, large scale solver, such as the conjugate gradient method. In this case, the computational complexity for

each iteration would be in the same order as the backprojection algorithm. In practice, it is computationally more efficient to discretize the light transport matrix for use in iterative inverse procedures,[20] at the cost of increasing memory requirements to the order of $O\left(N^5\right)$.

The discrete light cone transform (LCT) applies a transformation operation along the time axis ($O\left(N^3\right)$), followed by a 3D Fourier transform ($O\left(N^3\log N\right)$), an element-wise multiplication ($O\left(N^3\right)$), an inverse 3D Fourier transform ($O\left(N^3\log N\right)$), and another transformation operation ($O\left(N^3\right)$). The combined computational complexity is $O\left(N^3\log N\right)$. The alternating-direction method of multipliers (ADMM) and its linearized version (L-ADMM) are iterative algorithms that apply the LCT sequentially along with other, computationally less complex, proximal operators (see Supplementary Derivations). Therefore, the order of the computational complexity per iteration is similar to that of the LCT. The LCT only requires a single volume to be stored in memory, so the memory requirements are several orders of magnitude better than existing methods. ADMM and L-ADMM require several intermediate variables of the same size as the volume to be stored but the order of memory required remains the same as the LCT.

	Backprojection	Direct Inverse	Iterative Inverse*	Direct Inverse with LCT	ADMM with LCT*	L-ADMM with LCT*
Runtime	$O\left(N^5\right)$	$O\left(N^9\right)$	$O\left(N^5\right)$	$O\left(N^3\log N\right)$	$O\left(N^3\log N\right)$	$O\left(N^3\log N\right)$
Memory	$O\left(N^3\right)$	$O\left(N^6\right)$	$O\left(N^5\right)$	$O\left(N^3\right)$	$O\left(N^3\right)$	$O\left(N^3\right)$

* The complexity listed for iterative methods represent the computation and memory requirements per iteration.

Supplementary Results

In support of the results shown in the primary text, we show several additional results of C-NLOS imaging in Figures 5–12. Here, Figures 5–9 are physical experiments. Figures 10–12 are simulations that explore the impact of tradeoffs necessary to achieve lower acquisition times by either reducing exposure time or by subsampling the wall.

Figure 5 shows a scene with a flat retroreflective traffic sign. The volume of recorded measurements is shown on the upper left and other columns in the first row show two $x-t$ slices of the measurements as indicated by dashed, yellow lines. These represent only indirect parts of the recorded light transport; the direct reflections off the wall are digitally gated out by the proposed acquisition procedure. Compared with a reconstruction via backprojection (left column) or filtered backprojection (center left column), the proposed LCT-based reconstruction (center right column) reveals significantly more details, making the text and arrow legible. Even with the Laplacian filter proposed by Velten et al.[15] (center left column), the filtered backprojection method is not able to adequately recover the hidden scene, because the solution to the inverse system is only approximated but not actually computed.

Note that for conciseness, we have thus far assumed that the volume of unknown albedos is reconstructed directly from the transient image captured of the visible wall. In practice, however, the histograms measured by SPADs are not necessarily the same as the transient image. A single photon that hits a SPAD has some probability of creating an electron avalanche that is then time-stamped by the time-to-digital converter. After a detected event, the SPAD must be quenched before another event can be recorded. Assuming that the length of the emitted laser pulse is significantly smaller than the SPAD's dead time, at most one of the photons in that pulse can trigger an event, which is not necessarily the first arriving photon. The presence or absence of a detected event within a short window is thus a Bernoulli trial, which is repeated many times to accumulate the histogram. A sequence of Bernoulli trials with constant probabilities results in measured histograms being observations of an inhomogeneous Poisson process. In the Supplementary Derivations (next section), we develop a comprehensive signal processing framework that incorporates Poisson noise in the measurements as well as scene priors on the recovered volume of albedos. At the core of this signal processing framework is the light cone transform, but it is embedded in an iterative optimization framework. Please refer to prior work for more details on the connection between transient images and measured SPAD histograms.[27]

We show results of this iterative approach in the right column of Figure 5 and impose priors on the Poissonian nature of the measurements, a nonnegativity prior on the recovered albedos, as well as a combination of sparsity and isotropic 3D total variation (TV) priors on the albedos. Related priors are commonly applied in various computational imaging applications, such as LIDAR,[23] and they have also been proposed for NLOS imaging.[17,20] We are the first to develop appropriate formulations for incorporating these priors in NLOS applications that are dominated by Poisson noise. Please refer to the mathematical derivations section for more details.

In Figure 6, we show another scene that contains two retroreflective objects that partly occlude each other. Even though this scene violates one of the assumptions made by our image formation model, the reconstructions in this experiment are relatively robust to these occlusions. The two characters are well-separated, as seen in the $x - z$ slices in the bottom row, and a small amount of residual blur in the LCT results (center right column) is removed by adding a combination of sparsity and isotropic 3D TV priors to the unknown volume (right column).

Figure 7 shows another, more complex 3D scene. The front of the mannequin torso, head, and limbs are coated with a retroreflective paint before recording this object. The 3D structure of this hidden scene is well-recovered. Similar to Figures 5 and 6, the LCT reconstruction shows significant improvements over backprojection-type approaches. Adding additional priors further improves our results.

A purely diffuse scene is tested in Figure 8. The SNR for diffuse NLOS scenes is significantly lower than that of retroreflective scenes, as seen in the noisy measurements (top row). The decreased SNR results in much noisier measurements for the LCT (center right column). Again, these can be mitigated by applying priors on the recovered albedos, here promoting contiguous surfaces, but the overall reconstruction quality is slightly lower than for retroreflective objects.

A scene captured outdoors in indirect sunlight is shown in Figure 9. The reconstructed retroreflective letter is identifiable in the reconstruction despite the high levels of ambient sunlight. The letter is placed $115\,\mathrm{cm}$ away from the wall. The wall itself consists of a light and dark stone, and so there is an observable variation in the intensity of the measurements. The reconstruction method is robust in this challenging scenario.

All physical experiments, except for the "Outside S", are recorded by sampling the wall at 64×64 locations with 512 time bins for the temporal histograms, each with a precision of $16\,\mathrm{ps}$. The "Outside S" scene is sampled at 32×32 locations on the wall. The exposure time per sample is $0.1\,\mathrm{s}$ for the "Exit Sign", the "SU", the "Outside S" and the resolution chart scenes. The exposure time for the "Mannequin" and "Diffuse S" scenes is $1\,\mathrm{s}$ per sampled location. Thus, the total exposure time therefore ranges from $1.7\,\mathrm{min}$ for the "Outside S" scene and $68\,\mathrm{min}$ for the "Diffuse S" scene. The sampled area on the wall is $80\,\mathrm{cm} \times 80\,\mathrm{cm}$ for the "Exit Sign", $70\,\mathrm{cm} \times 70\,\mathrm{cm}$ for the "SU", the "Mannequin", and the "Diffuse S", $1\,\mathrm{m} \times 1\,\mathrm{m}$ for the "Outside S", and $40\,\mathrm{cm} \times 40\,\mathrm{cm}$ for the resolution chart. The simulations in Figures 10 & 11 have a resolution of 512×512 spatial samples, and Figure 12 has a resolution of 256×256. Reconstruction times with the LCT approach are below $1\,\mathrm{s}$ for all scenes with a resolution up to 64×64 samples and approx. 3–5 min for the simulations with a resolution of 512×512. Our unoptimized implementation of the backprojection method and the filtered backprojection methods took $8.5\,\mathrm{min}$ for the scenes with 64×64 samples, which makes our LCT-based reconstruction more than $500\times$ faster than the inferior backprojection-type methods. Higher-resolution reconstructions with priors took significantly longer, up to 8 hours for the simulations with a resolution of 512×512. See Table 1 for a summary of experimental details, including the detected photon count for each experiment.

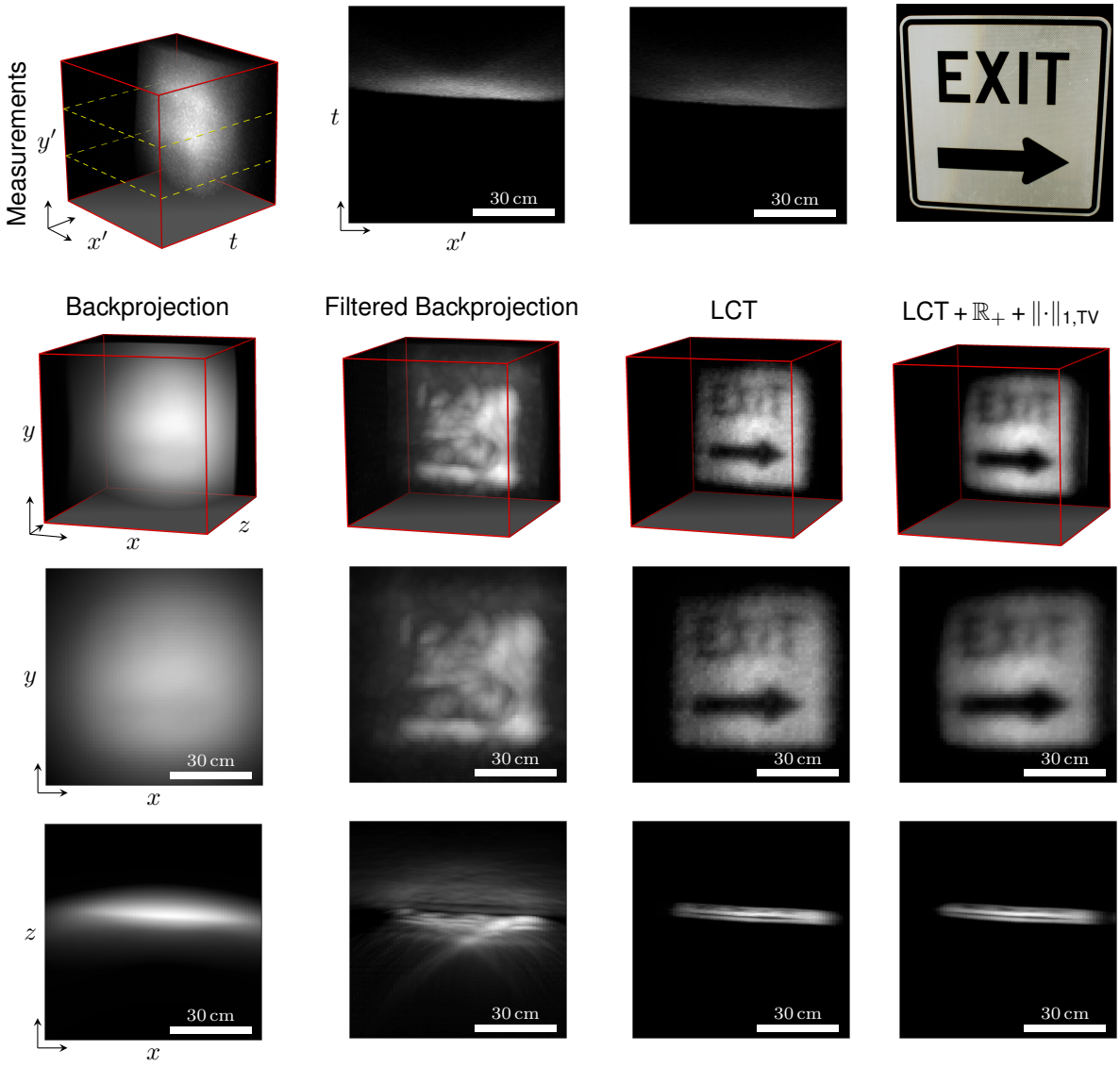

Figure 5: Experimental result: "Exit Sign". A planar, retroreflective traffic sign is imaged outside the direct line of the sight of the detector and light source. The volume of time-resolved measurements is shown in the top row along with two $x - t$ slices or "streak images" indicated by dashed, yellow lines and a photograph of the traffic sign. Backprojection-type methods (left and center left columns) fail to recover fine details and are usually blob-like. The light cone transform (LCT) developed in this paper is a direct inverse method that unlocks significantly faster and higher-quality reconstructions with a smaller memory footprint than other methods (center right column). The LCT also allows for additional priors, such as nonnegativity, sparsity, or total variation, to be imposed on the unknown albedos (right column).

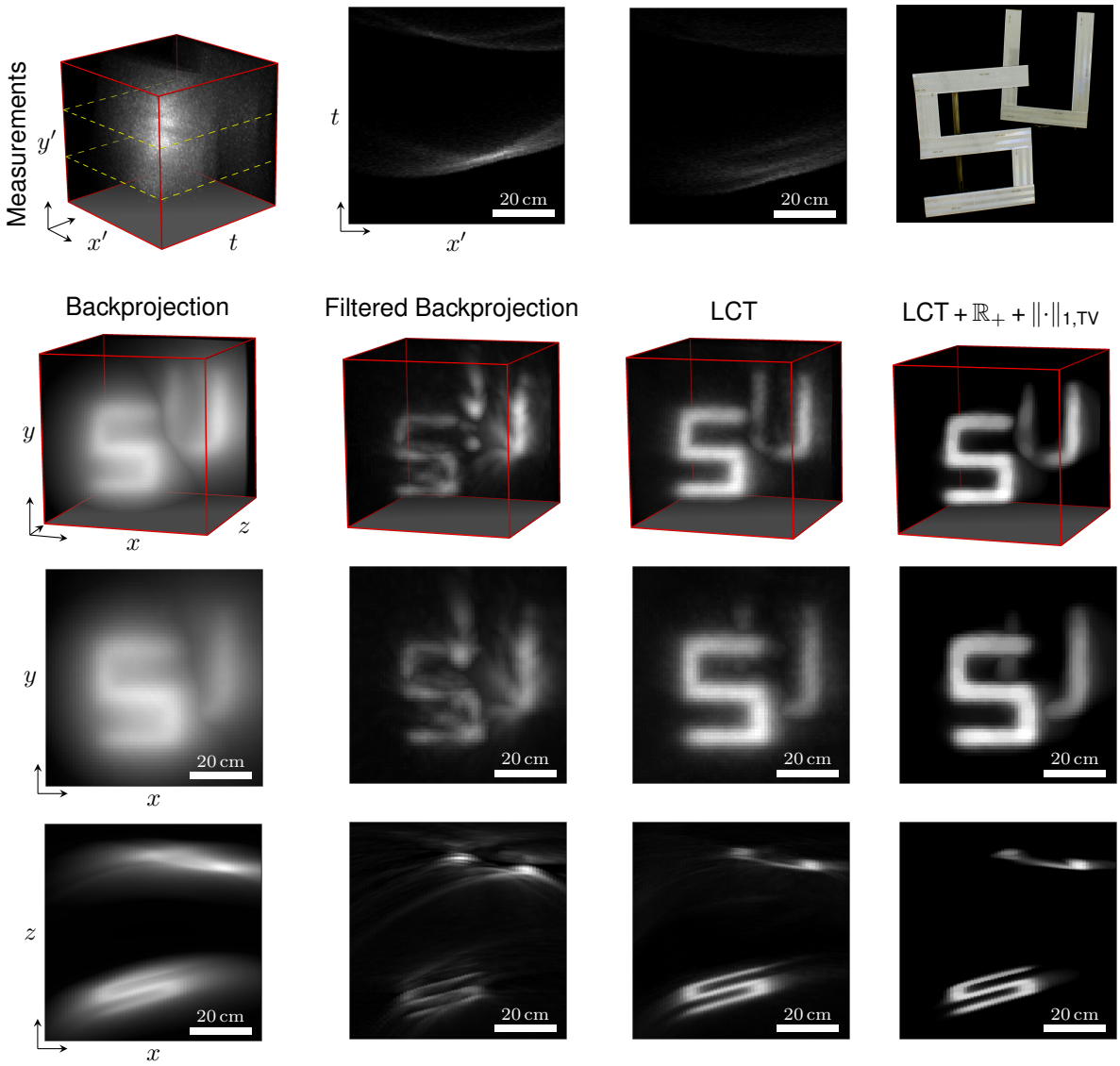

Figure 6: Experimental result: "SU". This scene contains two planar, retroreflective objects that partly occlude each other. Even though occlusions violate one of the assumptions of the proposed image formation, the reconstructions are relatively insensitive to this violation. Again, backprojection-type methods (left and center left columns) fail to recover fine details whereas a direct inverse via the LCT (center right column) or via the LCT and scene priors (right column) allows for significantly better and also faster reconstructions.

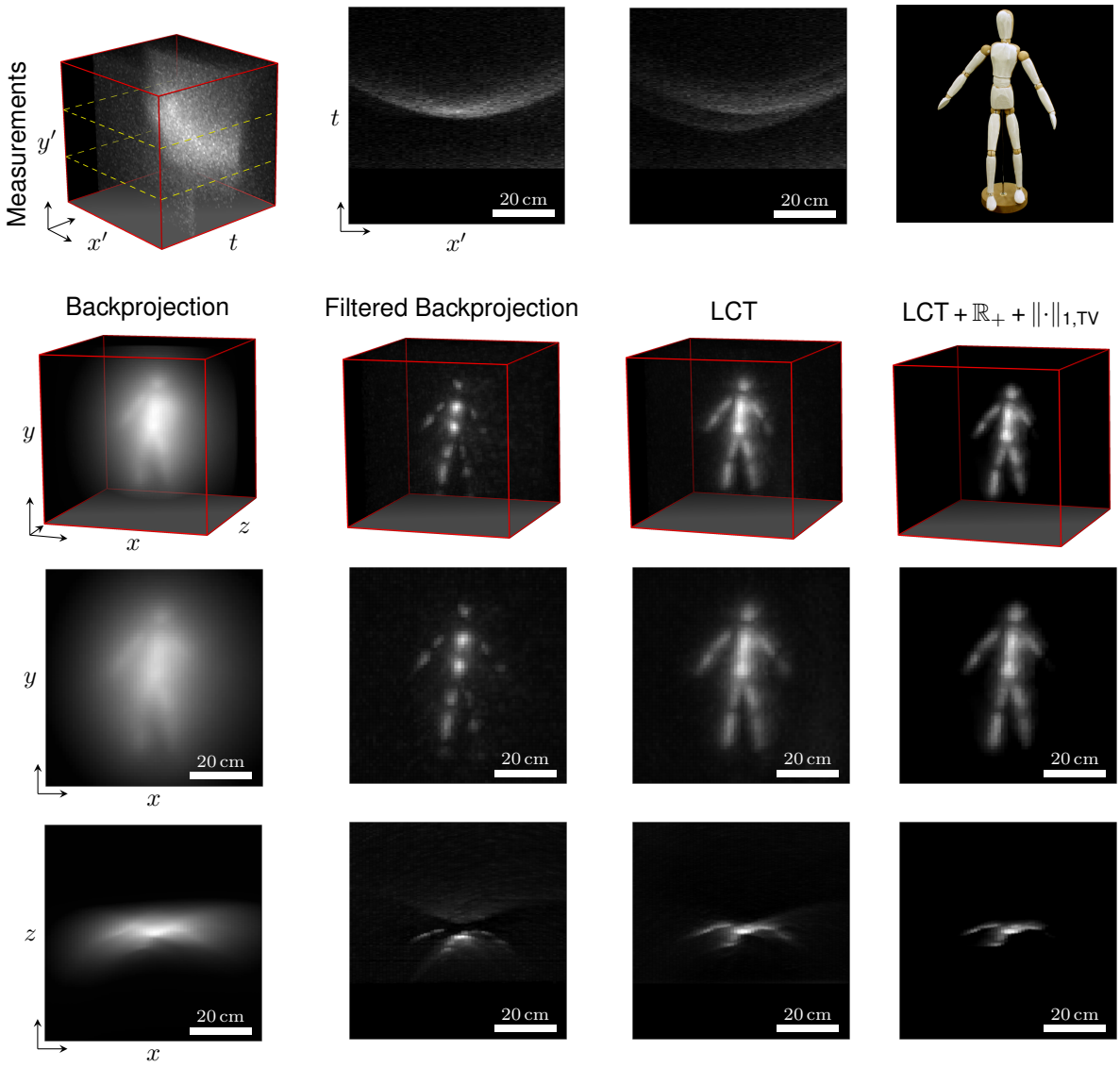

Figure 7: Experimental result: "Mannequin". This scene has a more complex 3D structure than the planar objects shown so far. The mannequin is coated with retroreflective paint. Despite its complex 3D structure, the mannequin is composed of surface patches that are spatially well-separated from each other. The hands, arms, legs, head, and torso do not contain fine geometric detail and have sufficient spacing between them for backprojection-type reconstruction approaches (left and center left columns) to generate a reasonable approximation of the 3D shape. Nevertheless, as with the previous examples, LCT-type reconstructions (center right and right columns) are successful in recovering more details of the mannequin.

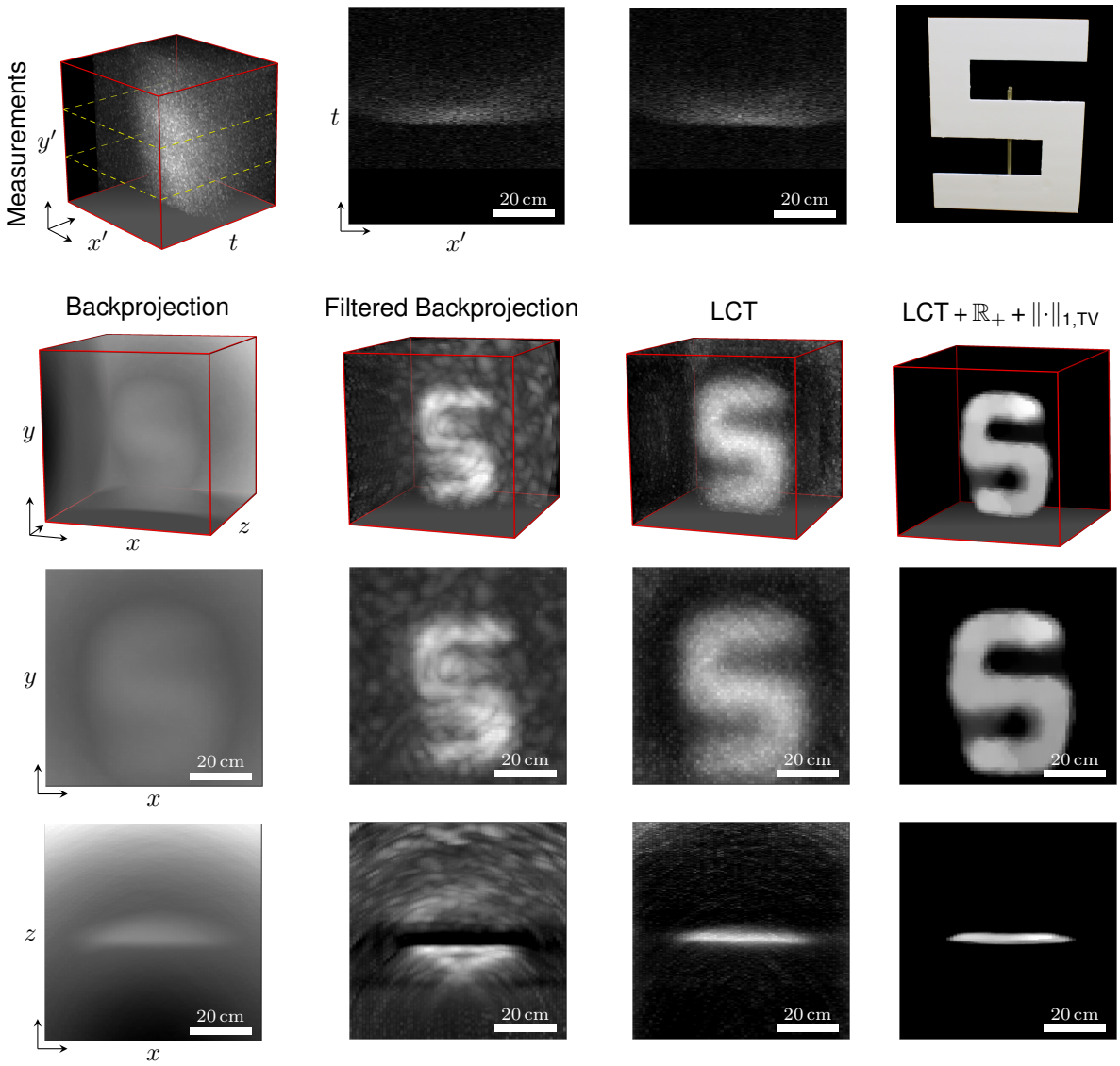

Figure 8: Experimental result: "Diffuse S". As opposed to the previous examples, this scene shows a planar diffuse object. Even with an exposure time of 1 second per sampled location, the measurements of this diffuse scene (top row) are significantly noisier than those of retroreflective scenes because the intensity reduces with distance at a faster rate for diffuse objects than for retroreflective objects. The noisy measurements propagate into the reconstructions (center right column), but can be mitigated by a total variation (TV) prior (right column).

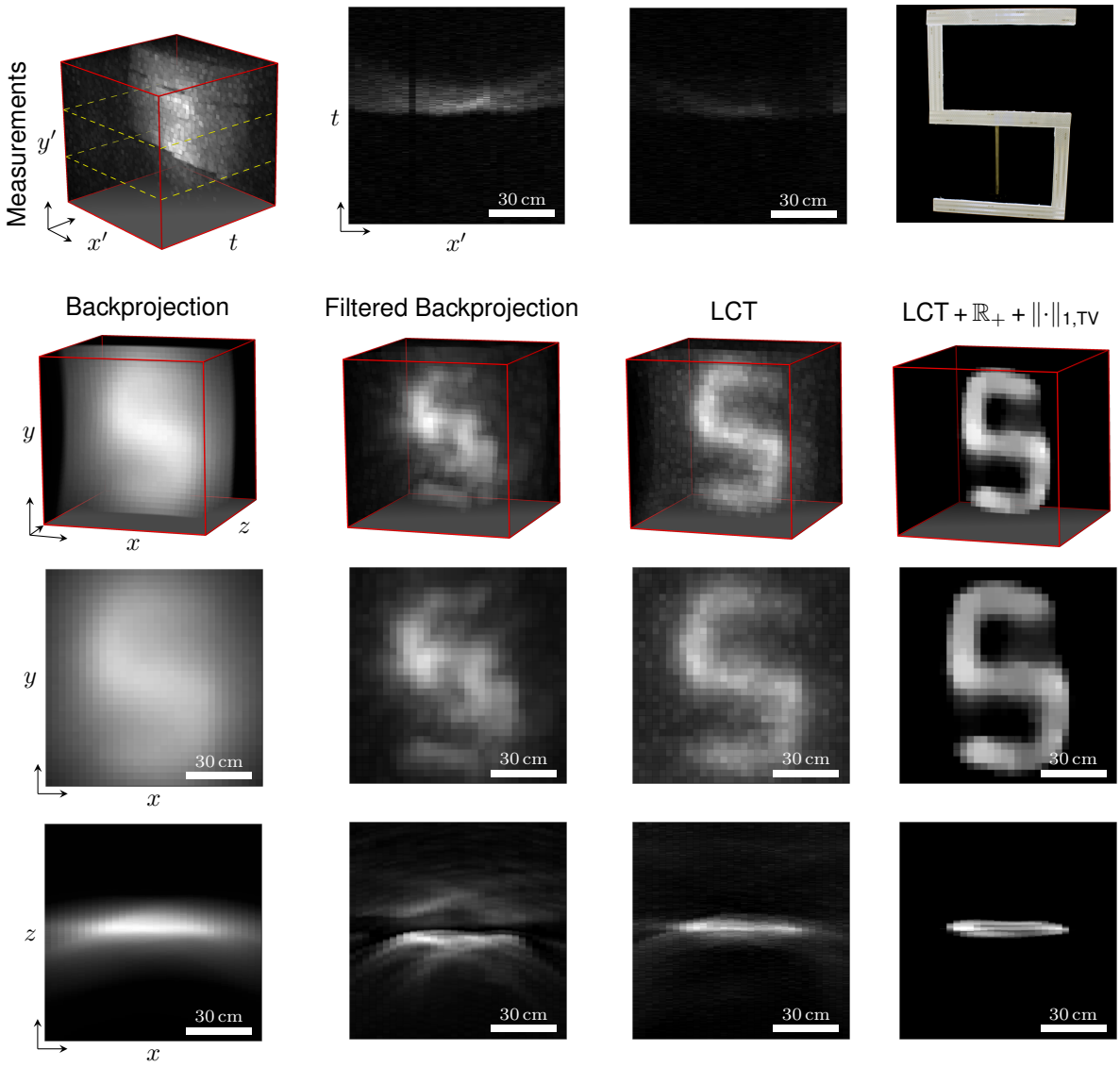

Figure 9: Experimental result: "Outside S". A planar, retroreflective object is imaged outdoors in indirect sunlight. The object is placed outside the direct line of sight of the detector, and approximately 115 cm away from a building wall. The wall is composed of light and dark-colored stone, and so measurements captured on the darker stone are noticeably lower in intensity. Despite the challenging nature of the outdoor environment, the reconstructed letter is recognizable in the LCT reconstruction (center right column), where details are more accurately represented than with filtered backprojection. Adding priors to the reconstruction further improves the quality of the reconstructed result (right column).

Experiment name	Samples	Bins	Detection area	Exposure	Retro.	Counts[1]	Counts[2]	Counts[3]
Res. Chart (40 cm)	64×64	512	$0.4\,\mathrm{m} \times 0.4\,\mathrm{m}$	0.1 s	Yes	1.0×10^5	5.9×10^4	2.9×10^3
Res. Chart (65 cm)	64×64	512	$0.4\,\mathrm{m} \times 0.4\,\mathrm{m}$	0.1 s	Yes	1.0×10^5	5.5×10^4	1.8×10^3
Dot Chart (40 cm)	64×64	512	$0.4\,\mathrm{m} \times 0.4\,\mathrm{m}$	0.1 s	Yes	1.0×10^5	5.8×10^4	1.4×10^3
Dot Chart (65 cm)	64×64	512	$0.4\,\mathrm{m} \times 0.4\,\mathrm{m}$	0.1 s	Yes	1.0×10^5	5.9×10^4	0.8×10^3
Exit Sign	64×64	512	$0.8\,\mathrm{m} \times 0.8\,\mathrm{m}$	0.1 s	Yes	4.1×10^4	2.0×10^4	5.6×10^3
SU	64×64	512	$0.7\,\mathrm{m} \times 0.7\,\mathrm{m}$	0.1 s	Yes	2.9×10^4	1.3×10^4	1.2×10^3
Mannequin	64×64	512	$0.7\,\mathrm{m} \times 0.7\,\mathrm{m}$	1.0 s	Yes	5.1×10^5	2.6×10^5	1.8×10^3
Diffuse S	64×64	512	$0.7\,\mathrm{m} \times 0.7\,\mathrm{m}$	1.0 s	No	3.6×10^5	1.8×10^4	1.1×10^3
Outside S	32×32	1024	$1.0\,\mathrm{m} \times 1.0\,\mathrm{m}$	0.1 s	Yes	8.5×10^4	3.3×10^4	2.7×10^3

Table 1: Experimental details for Figures 4 to 9. Samples: number of points sampled on wall. Bins: number of histogram bins used to generate result. Detection area: physical size of region sampled on wall. Exposure: exposure time per sample. Retro.: boolean for hidden object's retroreflectivity. Counts[1]: average number of counts in unprocessed histogram. Counts[2]: average number of counts in cropped/downsampled histogram (without removal of direct component). Counts[3]: average number of counts in cropped/downsampled histogram (with removal of direct component).

We further investigate the impact of sensor noise with simulations in Figures 10 and 11.

Figure 10 evaluates the case where temporal histograms (row 1, column 2) with a reasonably high signal-to-noise ratio (SNR) are simulated for an NLOS object in the shape of a bunny (row 1, column 1). Measurements are simulated with a physically-based ray tracer (PBRT) and Poisson noise is added subsequently. Neither the backprojection method (row 1, column 3) nor the filtered backprojection method[15] (row 1, column 4) achieve good results. The LCT allows for significantly higher-resolution reconstructions (rows 2-4, column 1). Here we show the maximum intensity projection (MIP, row 2) as well as the recovered shape (row 3) and also the error of the recovered shape with respect to ground truth (row 4). The LCT achieves a high-quality reconstruction at a resolutions of 512^3 voxels, which is well beyond the scope of what conventional algorithms can handle.[1] Fine details on the bunny's surface are faithfully recovered. Accounting for the Poissonian nature of the measurement noise and applying a nonnegativity prior (rows 2-4, column 2), a nonnegativity prior combined with a sparsity prior on the 3D albedo volume (rows 2-4, column 3), or a nonnegativity prior combined with a 3D TV prior on the 3D albedo volume (rows 2-4, column 4) does not noticeably improve the reconstruction quality in this experiment.

None of the approaches evaluated in Figure 10 are able to recover the right ear of the bunny. This is due to self occlusions in the measurements. Occlusions are not modeled by any existing NLOS algorithm and the resulting violation of the image formation model results in reconstruction errors, such as those observed in the right ear. Moreover, we see that the error of the recovered shape is higher around the outline of the bunny than in the interior. This is due to the fact that the image formation model in our and other NLOS approaches does not take surface normals and the resulting foreshortening of surface patches into account, which results in intensity

[1]Even a sparse representation of the light transport matrix corresponding to a 512^3 volume would require 0.3 petabytes of memory and that of a volume with 1024^3 voxels (see bunny experiment in primary text) approx. 9 petabytes when represented with 64 bit double-precision floating point values.

variations along the measured "streaks". Both occlusions and hidden surface normals are not modeled adequately by existing NLOS approaches and could further improve NLOS reconstructions in future work. However, a direct inverse such as the LCT, which is enabled by a shift-invariant image formation model, may not be applicable in that case.

Figure 11 shows another experiment. Here, we significantly reduce the measured signal (row 1, column 2), which results in more Poisson noise than the experiment in Figure 10. Neither the backprojection method (row 1, column 3) nor the filtered backprojection method (row 1, column 4) generates a meaningful result. The LCT reconstruction is also heavily corrupted by noise, although most of these artifacts accumulate at the far end of the hidden volume and the shape of the bunny (row 3, column 1) can still be made out clearly. Using the iterative reconstruction framework developed in the Supplementary Derivations, we again account for Poisson noise and add a nonnegativity prior (rows 2-4, column 2), a nonnegativity prior combined with a sparsity prior on the 3D albedo volume (rows 2-4, column 3), and a nonnegativity prior combined with a 3D TV prior on the 3D albedo volume (rows 2-4, column 4). In this experiment, the quality of the reconstructions is significantly improved and especially the sparsity prior (rows 2-4, column 3) suppresses reconstruction noise. Thus, priors on the hidden volume can improve reconstructions that are heavily corrupted by noise and modeling the noise statistics adequately seems crucial. This can be achieved by embedding the LCT in an iterative signal processing framework, as derived in the following, but comes at the cost of increased reconstruction times.

Finally, we try to answer the following question with the experiment shown in Figure 12: if one would like to reduce the acquisition time of C-NLOS imaging, would it be better to reduce the exposure time of a fixed number of samples on the wall or to subsample the wall while keeping the exposure time per sample fixed? This experiment suggests that it may be beneficial to reduce the number of samples rather that the exposure time per sampled location. This experiment further argues for the importance of high SNR in measurements which can be optimized using retroreflective materials, as facilitated by C-NLOS imaging.

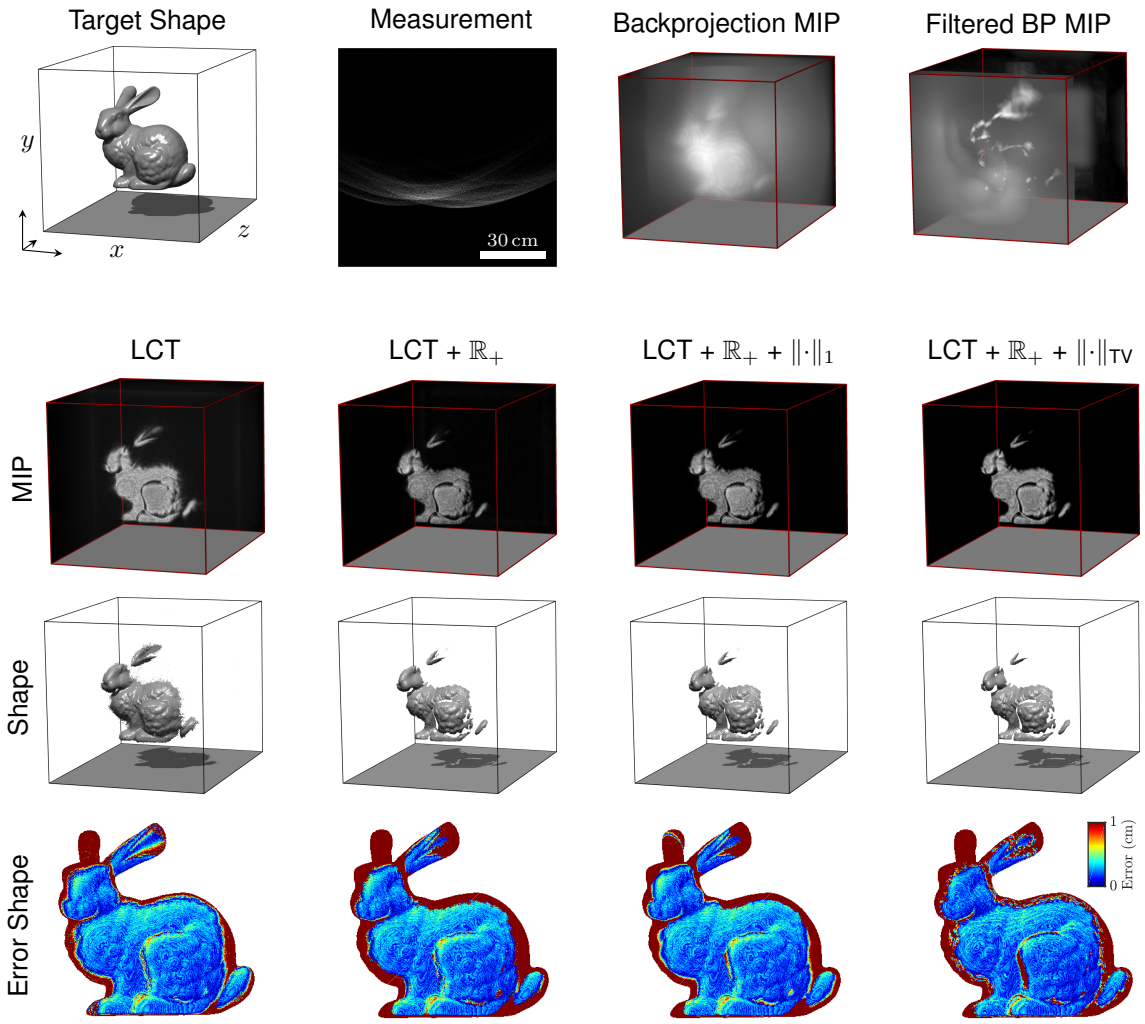

Figure 10: Evaluation of various reconstruction approaches for weak Poisson noise. The top row shows the target shape of the hidden object, one $x - t$ slice of the measurements as well as maximum intensity projections (MIPs) of reconstructions using the backprojection and filtered backprojection methods. Rows 2-4 also show MIPs, reconstructed shapes, and the error in shape of the light cone transform (LCT, column 1) as well as iterative solutions that use the LCT and a maximum a priori estimation accounting for Poisson noise along with a nonegativity prior (column 2), a nonnegativity prior together with a sparsity prior (column 3), and a nonnegativity prior together with a total variation prior (column 4). Here, the LCT achieves high-quality results which are not significantly improved by additional priors.

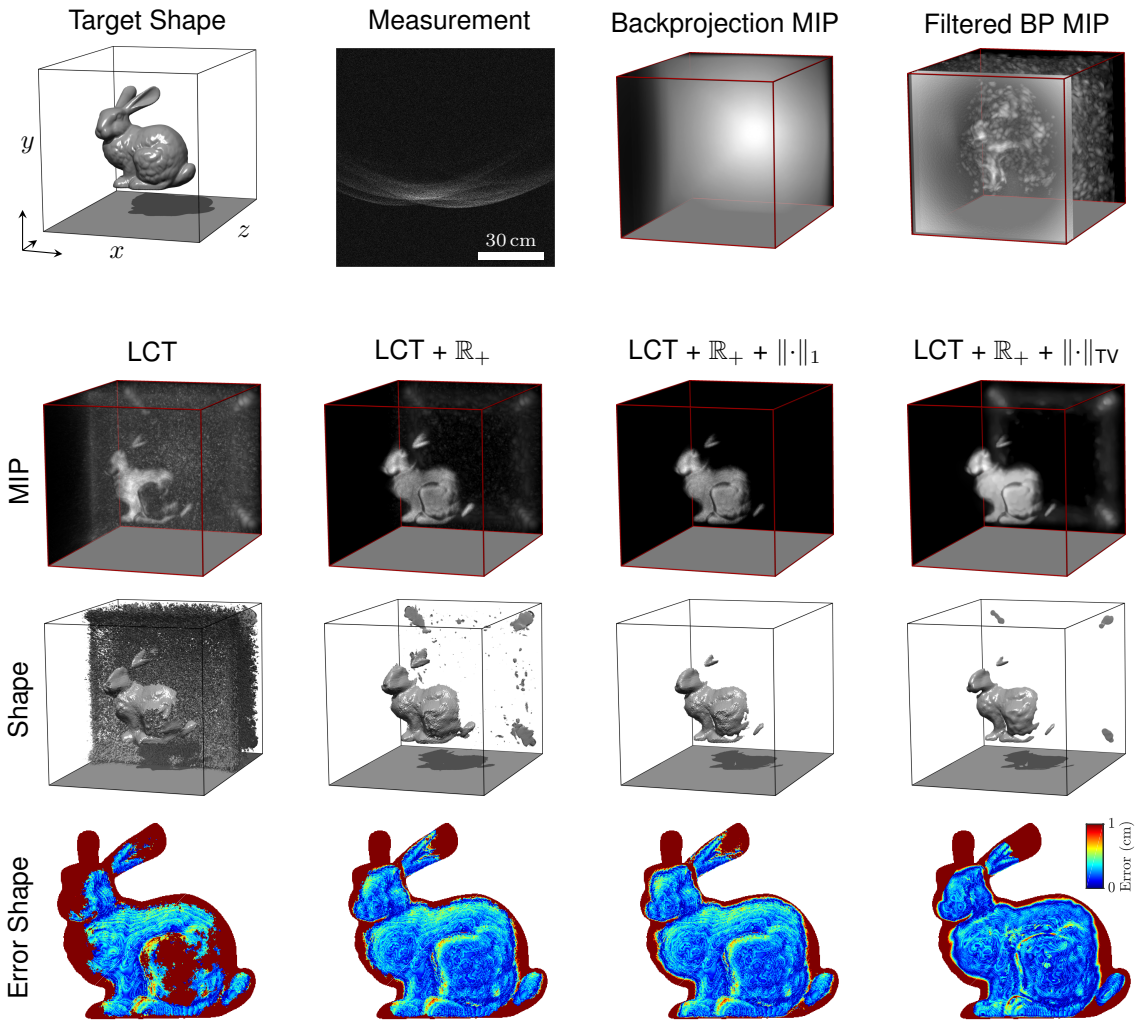

Figure 11: Evaluation of various reconstruction approaches for strong Poisson noise. The top row shows the target shape of the hidden object, one $x - t$ slice of the measurements as well as maximum intensity projections (MIPs) of reconstructions using the backprojection and the filtered backprojection methods. Rows 2-4 also show MIPs, reconstructed shapes, and the error in shape of the light cone transform (LCT, column 1) as well as iterative solutions that use the LCT and a maximum a priori estimation accounting for Poisson noise along with a nonnegativity prior (column 2), a nonnegativity prior together with a sparsity prior (column 3), and a nonnegativity prior together with a total variation prior (column 4). Here, backprojection-type methods fail to recover a meaningful result. The LCT reconstruction is corrupted by noise, especially at the far end of the volume. Additional priors, in particular those promoting sparsity, significantly improve the reconstruction.

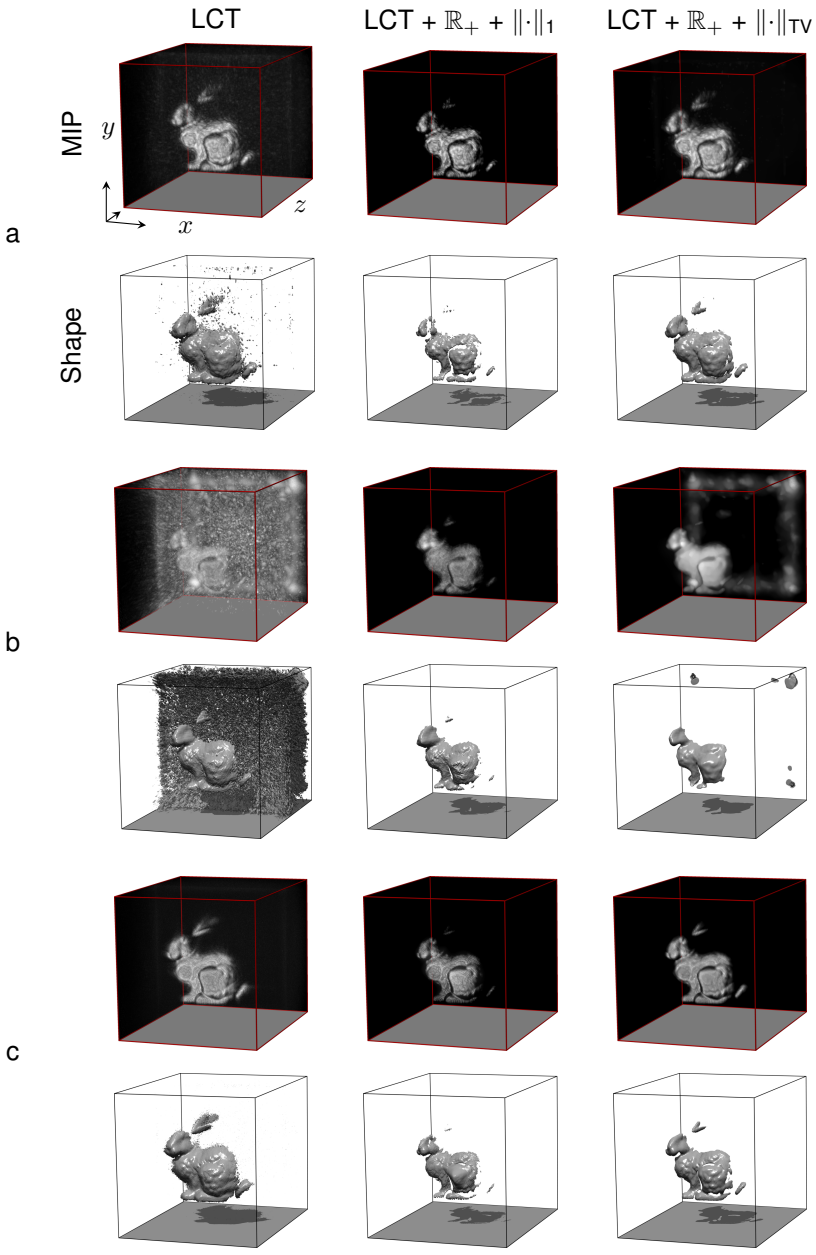

Figure 12: Evaluation of variation in acquisition time and sampling density for reconstructions with Poisson noise for a fixed-size volume in simulation. Rows 5-6 (panel (c)) can be interpreted as the "ground truth" of this scene. The two rows of panel (c) show maximum intensity projections (MIP) and shape visualizations. These images show reconstructions for densely sampled measurements with long acquisition times. To minimize total acquisition time, one can either record fewer samples with the same exposure time as the "ground truth" (panel (a)) or keep the number of samples comparable while reducing the exposure time (panel (b)). In panel (a), we simulate a reduced number of total samples. For the direct reconstruction via LCT, the sparse measurements are upsampled using nearest neighbor interpolation before applying the LCT (panel (a), left column). For the LCT + prior reconstructions, a sampling operator is incorporated into the iterative reconstructions which accounts for subsampling as described in the next section (panel (a), center and right columns). Panel (b) shows reconstructions for densely sampled measurements with short acquisition times, but equivalent total acquisition time as in (a). The reconstruction quality of panel (a) is better than that of panel (b), suggesting that longer exposure times for fewer samples would be a preferred tradeoff when trying to minimizing acquisition times.

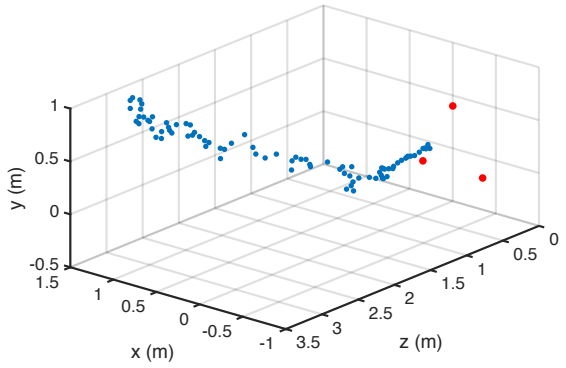

Figure 13: Tracking results for a non-line-of-sight retroreflective sign moving across a room. The system records measurements at three different points (shown in red) on a wall as position $z = 0$, and uses the indirect signal to compute the distance of the sign from the wall. The intersection of three spheres produces 3D coordinates of the sign (shown in blue). The exposure period for each point is $0.1\,\mathrm{s}$, producing 3D coordinates at a rate of approximately $3\,\mathrm{Hz}$. Measurements become noisier as the object moves away from the wall.

Real-time non-line-of-sight tracking Instead of fully recovering the shape of non-line-of-sight objects, one may also be interested in detecting the presence of an object or roughly track its location.[11–13] The minimum number of required sample points to unambiguously track a single 3D object is three. The detector records the time of flight between the three sample points and the NLOS object. A single sample point is ambiguous because the manifold of possible 3D locations of the corresponding object is the surface of a hemisphere. With three measurements, we can compute the intersection of the three respective hemispheres to triangulate the position of the object.

Figure 13 shows the results of tracking the 3D location of a planar traffic sign outside the line of sight of the detector. Three locations on the wall are sampled with mutual distances of approx. 60 cm, as indicated by the red dots. This tracking procedure is performed at interactive framerates with $3\,\mathrm{Hz}$ and an exposure time of $0.1\,\mathrm{s}$ per sample. Please also see the Supplementary Video for additional real-time tracking results in outdoor applications.

Supplementary Derivations

In the following, we derive efficient C-NLOS estimation schemes for applications where Poisson noise dominates the image formation. A closed-form solution for the C-NLOS problem does not exist in this case, so we utilize the alternating direction method of multipliers (ADMM) as an iterative reconstruction method. Our ADMM algorithm is tailored to C-NLOS imaging and makes extensive use of the discrete light cone transform derived above.

Modeling single photon detection Photon counters, such as the single photon avalanche diodes used in our prototype, detect the time of arrival of a single photon event within a certain time window with some probability. In low-flux applications, such as non-line-of-sight imaging, this is a Bernoulli experiment, which is repeated many times.[23,24] Therefore, the image formation of confocal non-line-of-sight imaging is

$$\mathbf{h} \sim \mathcal{P}(\mathbf{A}\boldsymbol{\rho} + \mathbf{d}), \tag{28}$$

where $\mathbf{h} \in \mathbb{Z}_*^{n_x n_y n_t}$ is a vectorized representation of the measured photon count for $n_x \times n_y$ sampled locations on the wall, each containing n_t discrete histogram time bins. Ambient light and erroneous photon detection events known as dark count are modeled by the term $\mathbf{d} \in \mathbb{R}_+^{n_x n_y n_t}$. The unknown albedos of each voxel in the hidden volume of resolution $n_x \times n_y \times n_z$ is $\boldsymbol{\rho} \in \mathbb{R}_+^{n_x n_y n_z}$. The matrix $\mathbf{A} \in \mathbb{R}_+^{n_x n_y n_t \times n_x n_y n_z}$ encodes the discretized version of the image formation outlined by Equation (12).

The probability of having taken a measurement at one particular SPAD and histogram bin i is thus given as

$$p\left(\mathbf{h}_i | \mathbf{A}\boldsymbol{\rho}, \mathbf{d}\right) = \frac{\left(\mathbf{A}\boldsymbol{\rho} + \mathbf{d}\right)_i^{\mathbf{h}_i} \, e^{-(\mathbf{A}\boldsymbol{\rho}+\mathbf{d})_i}}{\mathbf{h}_i!}. \tag{29}$$

Maximum a posteriori C-NLOS estimation via ADMM Here, we derive a maximum a posteriori (MAP) estimator for the inverse NLOS problem that minimizes the negative log-likelihood of Equation (29) and also imposes a prior $\Gamma\left(\cdot\right)$ on the volume as

$$\begin{aligned} \underset{\{\boldsymbol{\rho}\}}{\text{minimize}} \quad & -\log \; p\left(\mathbf{h}|\mathbf{A}\boldsymbol{\rho}, \mathbf{d}\right) + \Gamma(\boldsymbol{\rho}) \\ \text{subject to} \quad & 0 \leq \boldsymbol{\rho} \end{aligned} \tag{30}$$

Without loss of generality, we can bring the constraint into the objective function and write this as

$$\underset{\{\boldsymbol{\rho}\}}{\text{minimize}} \; -\log \; p\left(\mathbf{h}|\mathbf{A}\boldsymbol{\rho}, \mathbf{d}\right) + \mathcal{I}_{\mathbb{R}_+}(\boldsymbol{\rho}) + \Gamma(\boldsymbol{\rho}). \tag{31}$$

Here, $\mathcal{I}_{\mathbb{R}_+}\left(\cdot\right)$ is the indicator function that enforces the nonnegativity constraints

$$\mathcal{I}_{\mathbb{R}_+}\left(x\right) = \begin{cases} 0 & x \in \mathbb{R}_+ \\ +\infty & x \notin \mathbb{R}_+ \end{cases} \tag{32}$$

where $\mathbb{R}_+$ is the closed nonempty convex set representing nonnegative real-valued numbers.

Next, we follow the general approach of the alternating direction method of multipliers (ADMM)[31] and split the unknowns while enforcing consensus in the constraints

$$\begin{aligned} \underset{\{\boldsymbol{\rho}\}}{\text{minimize}} \quad & \underbrace{-\log\left(p\left(\mathbf{h}|\mathbf{z}_1,\mathbf{d}\right)\right)}_{g_1(\mathbf{z}_1)} + \underbrace{\mathcal{I}_{\mathbb{R}_+}\left(\mathbf{z}_2\right)}_{g_2(\mathbf{z}_2)} + \underbrace{\Gamma\left(\mathbf{z}_3\right)}_{g_3(\mathbf{z}_3)} \\ \text{subject to} \quad & \underbrace{\begin{bmatrix}\mathbf{A}\\ \mathbf{I}\\ \mathbf{I}\end{bmatrix}}_{\mathbf{K}}\boldsymbol{\rho} - \underbrace{\begin{bmatrix}\mathbf{z}_1\\ \mathbf{z}_2\\ \mathbf{z}_3\end{bmatrix}}_{\mathbf{z}} = 0 \end{aligned} \tag{33}$$

The Augmented Lagrangian of this objective is formulated as

$$L_\rho\left(\boldsymbol{\rho},\mathbf{z},\mathbf{y}\right) = \sum_{i=1}^{3} g_i\left(\mathbf{z}_i\right) + \mathbf{y}^T\left(\mathbf{K}\boldsymbol{\rho}-\mathbf{z}\right) + \frac{\rho}{2}\left\|\mathbf{K}\boldsymbol{\rho}-\mathbf{z}\right\|_2^2. \tag{34}$$

An iterative solver can now be constructed that sequentially minimizes the Augmented Lagrangian with respect to each of the variables $\boldsymbol{\rho}$ and $\mathbf{z}_i$. For convenience, the scaled form of ADMM uses the scaled dual variable $\mathbf{u} = (1/\rho)\mathbf{y}$ instead of the Lagrange multiplier $\mathbf{y}$. The variables $\mathbf{z}_i$ and $\mathbf{u}_i$ are initialized to zero, and $\boldsymbol{\rho}$ is initialized to $\boldsymbol{\rho}_0$, which can be zero or an initial guess (e.g. the closed-form LCT reconstruction). This leads to the following iterative updates:

$\mathbf{z}_i \leftarrow 0,\ \mathbf{u}_i \leftarrow 0,\ \boldsymbol{\rho} \quad \boldsymbol{\rho}_0$
for $k = 1$ **to** $maxiter$

$$\mathbf{z}_1 \quad \mathbf{prox}_{\mathcal{P},\rho}\left(\mathbf{v}\right) = \underset{\{\mathbf{z}_1\}}{\arg\min}\, L_\rho\left(\boldsymbol{\rho},\mathbf{z},\mathbf{y}\right) = \underset{\{\mathbf{z}_1\}}{\arg\min}\, g_1\left(\mathbf{z}_1\right) + \frac{\rho}{2}\left\|\mathbf{v}-\mathbf{z}_1\right\|_2^2, \quad \mathbf{v} = \mathbf{A}\boldsymbol{\rho}+\mathbf{u}_1 \tag{35}$$

$$\mathbf{z}_2 \quad \mathbf{prox}_{\mathcal{I},\rho}\left(\mathbf{v}\right) = \underset{\{\mathbf{z}_2\}}{\arg\min}\, L_\rho\left(\boldsymbol{\rho},\mathbf{z},\mathbf{y}\right) = \underset{\{\mathbf{z}_2\}}{\arg\min}\, g_2\left(\mathbf{z}_2\right) + \frac{\rho}{2}\left\|\mathbf{v}-\mathbf{z}_2\right\|_2^2, \quad \mathbf{v} = \boldsymbol{\rho}+\mathbf{u}_2 \tag{36}$$

$$\mathbf{z}_3 \quad \mathbf{prox}_{\Gamma,\rho}\left(\mathbf{v}\right) = \underset{\{\mathbf{z}_3\}}{\arg\min}\, L_\rho\left(\boldsymbol{\rho},\mathbf{z},\mathbf{y}\right) = \underset{\{\mathbf{z}_3\}}{\arg\min}\, g_3\left(\mathbf{z}_3\right) + \frac{\rho}{2}\left\|\mathbf{v}-\mathbf{z}_3\right\|_2^2, \quad \mathbf{v} = \boldsymbol{\rho}+\mathbf{u}_3 \tag{37}$$

$$\underbrace{\begin{bmatrix}\mathbf{u}_1\\ \mathbf{u}_2\\ \mathbf{u}_3\end{bmatrix}}_{\mathbf{u}} \quad \mathbf{u}+\mathbf{K}\boldsymbol{\rho}-\mathbf{z} \tag{38}$$

$$\boldsymbol{\rho} \quad \mathbf{prox}_{\|\cdot\|_2}\left(\mathbf{v}\right) = \underset{\{\boldsymbol{\rho}\}}{\arg\min}\, L_\rho\left(\boldsymbol{\rho},\mathbf{z},\mathbf{y}\right) = \underset{\{\boldsymbol{\rho}\}}{\arg\min}\, \frac{1}{2}\left\|\mathbf{K}\boldsymbol{\rho}-\mathbf{v}\right\|_2^2, \quad \mathbf{v} = \mathbf{z}-\mathbf{u} \tag{39}$$

end for

Note that both g_1 and g_2 are convex. As long as the prior Γ is convex, ADMM is guaranteed to find the global minimum of the objective. Even though ADMM is not guaranteed to converge for non-convex priors, many such priors have been shown to result in robust convergence in practice.[27]

Proximal Operator for Quadratic Term (Eq. (39))

The quadratic term requires special attention. An efficient closed-form solution can be found if the equality constraint for the prior $\Gamma(\cdot)$ can be formulated as in Equation (33) with $\boldsymbol{\rho} = \mathbf{z}_3$. For some priors (e.g. total variation), an additional matrix operator is incorporated into the equality constraint such that $\mathbf{D}\boldsymbol{\rho} = \mathbf{z}_3$, e.g. for the finite difference matrix $\mathbf{D}$. Unfortunately, a closed-form solution of the quadratic problem is not feasible in this case and we propose to follow the strategy of linearized ADMM (L-ADMM),[31] which approximates this quadratic problem with a linear one. We derive both approaches in the following.

Case 1: solving the quadratic subproblem with ADMM

For a closed-form solution of the proximal operator of the quadratic subproblem, we factor the radiometric term $\mathbf{R}$ from $\mathbf{R}_t^{-1}$ and incorporate it in the negative log-likelihood of Equation (33); we let $\tilde{\mathbf{R}}_t^{-1}$ denote the resampling operator, where $\mathbf{R}_t^{-1} = \mathbf{R}\tilde{\mathbf{R}}_t^{-1}$, such that

$$g_1(\mathbf{z}_1) = -\log\left(p\left(\mathbf{h}|\mathbf{R}\mathbf{z}_1, \mathbf{d}\right)\right), \quad \mathbf{K} = \begin{bmatrix} \tilde{\mathbf{R}}_t^{-1}\mathbf{H}\mathbf{R}_z \\ \mathbf{I} \\ \mathbf{I} \end{bmatrix}. \tag{40}$$

Then, the proximal operator of the quadratic problem becomes

$$\begin{aligned}\mathbf{prox}_{\|\cdot\|_2}(\mathbf{v}) &= \underset{\{\boldsymbol{\rho}\}}{\arg\min}\ \frac{1}{2}\left\| \begin{bmatrix} \tilde{\mathbf{R}}_t^{-1}\mathbf{H}\mathbf{R}_z \\ \mathbf{I} \\ \mathbf{I} \end{bmatrix} \boldsymbol{\rho} - \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \mathbf{v}_3 \end{bmatrix} \right\|_2^2 = \left(\mathbf{R}_z^T\mathbf{H}^T \underbrace{(\tilde{\mathbf{R}}_t^{-1})^T\tilde{\mathbf{R}}_t^{-1}}_{=I}\mathbf{H}\mathbf{R}_z + 2\mathbf{I}\right)^{-1}\left(\tilde{\mathbf{R}}_t^{-1}\mathbf{H}\mathbf{R}_z\right)^T\mathbf{v} \\ &= \left(\mathbf{R}_z^T\mathbf{H}^T\mathbf{H}\mathbf{R}_z + 2\mathbf{I}\right)^{-1}\left(\left(\tilde{\mathbf{R}}_t^{-1}\mathbf{H}\mathbf{R}_z\right)^T\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3\right). \end{aligned} \tag{41}$$

An expression for the closed-form inverse matrix can be derived as

$$\left(\mathbf{R}_z^T\mathbf{H}^T\mathbf{H}\mathbf{R}_z + 2\mathbf{I}\right)^{-1} \tag{42}$$

$$= \frac{1}{2}\mathbf{I} - \frac{1}{4}\mathbf{R}_z^T\left(\left(\mathbf{H}^T\mathbf{H}\right)^{-1} + \frac{1}{2}\underbrace{\mathbf{R}_z\mathbf{R}_z^T}_{=\mathbf{I}}\right)^{-1}\mathbf{R}_z \qquad \text{Woodbury Identity} \tag{43}$$

$$= \frac{1}{2}\mathbf{I} - \frac{1}{4}\mathbf{R}_z^T\left(\mathbf{F}^{-1}\left(\frac{1}{\widehat{\mathbf{H}}^*\widehat{\mathbf{H}}}\right)\mathbf{F} + \frac{1}{2}\underbrace{\mathbf{F}^{-1}\mathbf{F}}_{=\mathbf{I}}\right)^{-1}\mathbf{R}_z \tag{44}$$

$$= \frac{1}{2}\mathbf{I} - \frac{1}{4}\mathbf{R}_z^T\mathbf{F}^{-1}\left(\frac{1}{\widehat{\mathbf{H}}^*\widehat{\mathbf{H}}} + \frac{1}{2}\mathbf{I}\right)^{-1}\mathbf{F}\mathbf{R}_z \tag{45}$$

$$= \frac{1}{2}\mathbf{I} - \frac{1}{4}\mathbf{R}_z^T\mathbf{F}^{-1}\left(\frac{\widehat{\mathbf{H}}^*\widehat{\mathbf{H}}}{1 + \frac{1}{2}\widehat{\mathbf{H}}^*\widehat{\mathbf{H}}}\right)\mathbf{F}\mathbf{R}_z. \tag{46}$$

Again, $\mathbf{F}$ is the discrete Fourier transform along the spatial dimensions x, y and $\widehat{\mathbf{H}}$ is the optical transfer function of the light cone, a diagonal matrix. Thus, the closed-form solution of the proximal operator of the quadratic problem

is

$$\mathbf{prox}_{\|\cdot\|_2}(\mathbf{v}) = \frac{1}{2}\mathbf{I} - \frac{1}{4}\mathbf{R}_z^T\mathbf{F}^{-1}\left(\frac{\widehat{\mathbf{H}}^*\widehat{\mathbf{H}}}{1+\frac{1}{2}\widehat{\mathbf{H}}^*\widehat{\mathbf{H}}}\right)\mathbf{F}\mathbf{R}_z\left(\left(\tilde{\mathbf{R}}_t^{-1}\mathbf{H}\mathbf{R}_z\right)^T\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3\right). \tag{47}$$

This proximal operator can be efficiently evaluated for large-scale problems without explicitly forming any of the matrices.

Case 2: approximating the quadratic subproblem with L-ADMM

If the choice of prior alters the equality constraints given by Equation (33), a closed-form solution of the proximal operator of the quadratic subproblem may not exist. In this case, Equation (39) could be solved iteratively, for example using the conjugate gradient (CG) method. However, this is computationally expensive because we have to solve an iterative method (CG) embedded within another iterative method (ADMM). Instead, we propose to use linearized ADMM (L-ADMM),[31] where the quadratic subproblem is approximated with a linear one, resulting in the following update in Equation (39):

$$\boldsymbol{\rho} \leftarrow \mathbf{prox}_{\|\cdot\|_2}(\mathbf{v}) \approx \boldsymbol{\rho} - \frac{\rho}{\mu}\mathbf{K}^T(\mathbf{K}\boldsymbol{\rho} - \mathbf{v}). \tag{48}$$

This approach can be interpreted as a single step of a gradient descent approach with a step length of ρ/μ. Consequently, one would expect the linearized ADMM approach to converge slower than solving the quadratic subproblem with the closed-form solution outlined above. Nevertheless, for certain priors which are incompatible with the closed-form solution, linearized ADMM may be a good alternative. A suggested choice for the step length is $\rho/\mu = 1/\|\mathbf{K}\|_2^2$ with $\rho = 1/\|\mathbf{K}\|_2$.

Proximal Operator for Poisson Term (Eq. 35)

The specific splitting approach of Equation (40) requires the proximal operator of the Poisson term to include the radiometric term $\mathbf{R}$. Recall, this proximal operator is defined as

$$\mathbf{prox}_{\mathcal{P},\rho}(\mathbf{v}) = \underset{\{\mathbf{z}_1\}}{\arg\min}\,\mathcal{O}(\mathbf{z}_1) = \underset{\{\mathbf{z}_1\}}{\arg\min} -\log\left(p\left(\mathbf{h}|\mathbf{R}\mathbf{z}_1,\mathbf{d}\right)\right) + \frac{\rho}{2}\|\mathbf{z}_1 - \mathbf{v}\|_2^2. \tag{49}$$

Due to the fact that $\mathbf{R}$ is diagonal, the noisy observations are *independent* of each other, so we do not need to account for the *joint probability* of all measurements. We write the i$^{\text{th}}$ element of the objective function $\mathcal{O}$ as

$$\begin{aligned}\mathcal{O}_i(\mathbf{z}_1) &= -\log\left(\left(\mathbf{r}_i\mathbf{z}_{1_i} + \mathbf{d}_i\right)^{\mathbf{h}_i} e^{-\left(\mathbf{r}_i\mathbf{z}_{1_i}+\mathbf{d}_i\right)}\frac{1}{\mathbf{h}_i!}\right) + \frac{\rho}{2}\left(\mathbf{z}_{1_i} - \mathbf{v}_i\right)^2 \\ &= -\log\left(\mathbf{r}_i\mathbf{z}_{1_i} + \mathbf{d}_i\right)\mathbf{h}_i + \left(\mathbf{r}_i\mathbf{z}_{1_i} + \mathbf{d}_i\right) - \log\left(\frac{1}{\mathbf{h}_i!}\right) + \frac{\rho}{2}\left(\mathbf{z}_{1_i} - \mathbf{v}_i\right)^2\end{aligned} \tag{50}$$

and its gradient as

$$\frac{\partial\mathcal{O}_i}{\partial\mathbf{z}_{1_i}} = -\mathbf{r}_i\frac{\mathbf{h}_i}{\mathbf{r}_i\mathbf{z}_{1_i} + \mathbf{d}_i} + \mathbf{r}_i + \rho\left(\mathbf{z}_{1_i} - \mathbf{v}_i\right), \qquad \frac{\partial\mathcal{O}_i}{\partial\mathbf{z}_{1_j}} = 0 \quad \forall i \neq j. \tag{51}$$

Equating this gradient to zero results in a classical root-finding problem of a quadratic, which can be solved independently for each $\mathbf{z}_{1_i}$. The quadratic has two roots, but due to the nonnegativity constraints in our objective function we are only interested in the positive one. Thus, we can define the proximal operator as

$$\mathbf{prox}_{\mathcal{P},\rho}(\mathbf{v}) = -\frac{\mathbf{r}^2 + \rho\mathbf{d} - \rho\mathbf{v}\mathbf{r}}{2\rho\mathbf{r}} + \sqrt{\left(\frac{\mathbf{r}^2 + \rho\mathbf{d} - \rho\mathbf{v}\mathbf{r}}{2\rho\mathbf{r}}\right)^2 + \frac{\mathbf{r}\mathbf{h} + \rho\mathbf{d}\mathbf{v} - \mathbf{r}\mathbf{d}}{\rho\mathbf{r}}}. \tag{52}$$

This operator is a closed-form solution that is independently applied to each pixel, i.e. all vector-vector multiplications, divisions, and exponentials in Equation (52) are performed element-wise. Note that we still need to choose a parameter ρ for the ADMM updates. Heuristically, small values (e.g., $1e^{-5}$) work best for large signals with little noise, but as the Poisson noise starts to dominate the image formation, ρ should be higher. For more details on similar types of proximal operators, please refer to Dupe et al.[32]

Proximal Operator for Indicator Function (Eq. (36))

The proximal operator for the indicator function representing the constraints is straightforward

$$\mathbf{prox}_{\mathcal{I},\rho}(\mathbf{v}) = \underset{\{\mathbf{z}_2\}}{\arg\min}\, \mathcal{I}_{\mathbb{R}_+}(\mathbf{z}_2) + \frac{\rho}{2}\|\mathbf{v} - \mathbf{z}_2\|_2^2 = \underset{\{\mathbf{z}_2 \in \mathbb{R}_+\}}{\arg\min}\, \|\mathbf{z}_2 - \mathbf{v}\|_2^2 = \Pi_{\mathbb{R}_+}(\mathbf{v}), \tag{53}$$

where $\Pi_{\mathbb{R}_+}(\cdot)$ is the element-wise projection operator onto the convex set $\mathbb{R}_+$

$$\Pi_{\mathbb{R}_+}(\mathbf{v}_j) = \begin{cases} 0 & \mathbf{v}_j < 0 \\ \mathbf{v}_j & \mathbf{v}_j \geq 0 \end{cases} \tag{54}$$

Proximal Operator for the Prior Γ (Eq. (37))

The objective function of Equation (37) is

$$\mathbf{prox}_{\Gamma,\rho}(\mathbf{v}) = \underset{\{\mathbf{z}_3\}}{\arg\min}\, \Gamma(\mathbf{z}_3) + \frac{\rho}{2}\|\mathbf{v} - \mathbf{z}_3\|_2^2. \tag{55}$$

A variety of priors, many with closed-form solutions, have been proposed.[31] For example, total variation (TV) is one of the most widely used priors in image reconstruction and has recently been proposed for denoising range data captured with SPADs.[24] For this choice of prior, solving Equation (55) amounts to a soft-thresholding operator[31] where we impose the constraint $\mathbf{z}_3 = \mathbf{D}\boldsymbol{\rho}$, with $\mathbf{D}$ implementing a finite difference operator. However, the addition of the matrix operator in the constraint complicates the closed-form solution of the quadratic ADMM subproblem described in Case 1. Thus for the TV prior, or other priors which bring in additional terms to the constraints, the L-ADMM approach can be used.

Other choices of priors include sparsity priors on the volume or self-similarity-type priors, such as non-local means or BM3D. Many of these priors were recently studied in the context of denoising and deconvolution of transient images;[27] all of these priors or weighted combinations of them are equally applicable for NLOS imaging.

References

[30] Arfken, G. *Mathematical Methods for Physicists* (Academic Press, 1985), third edn.

[31] Boyd, S., Parikh, N., Chu, E., Peleato, B. & Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning* **3**, 1–122 (2011).

[32] Dupe, F.-X., Fadili, M. & Starck, J.-L. Inverse problems with Poisson noise: primal and primal-dual splitting. *Int. Conference on Image Processing* 1901–1904 (2011).