LETTER

Confocal non-line-of-sight imaging based on the light-cone transform

Matthew O'Toole[1], David B. Lindell[1] & Gordon Wetzstein[1]

How to image objects that are hidden from a camera's view is a problem of fundamental importance to many fields of research[1–20], with applications in robotic vision, defence, remote sensing, medical imaging and autonomous vehicles. Non-line-of-sight (NLOS) imaging at macroscopic scales has been demonstrated by scanning a visible surface with a pulsed laser and a time-resolved detector[14–19]. Whereas light detection and ranging (LIDAR) systems use such measurements to recover the shape of visible objects from direct reflections[21–24], NLOS imaging reconstructs the shape and albedo of hidden objects from multiply scattered light. Despite recent advances, NLOS imaging has remained impractical owing to the prohibitive memory and processing requirements of existing reconstruction algorithms, and the extremely weak signal of multiply scattered light. Here we show that a confocal scanning procedure can address these challenges by facilitating the derivation of the light-cone transform to solve the NLOS reconstruction problem. This method requires much smaller computational and memory resources than previous reconstruction methods do and images hidden objects at unprecedented resolution. Confocal scanning also provides a sizeable increase in signal and range when imaging retroreflective objects. We quantify the resolution bounds of NLOS imaging, demonstrate its potential for real-time tracking and derive efficient algorithms that incorporate image priors and a physically accurate noise model. Additionally, we describe successful outdoor experiments of NLOS imaging under indirect sunlight.

LIDAR systems use time-resolved sensors to scan the three-dimensional (3D) geometry of objects[21–24]. Such systems acquire range measurements by recording the time required for light to travel along a direct path from a source to a point on the object and back to a sensor. Recently, these types of sensors have also been used to perform NLOS tracking[12,13] or imaging[14–20] of objects 'hidden around corners', where the position and shape of the objects are computed from indirect light paths. The light travelling along indirect paths scatters multiple times before reaching a sensor and may scatter off objects outside a camera's direct line of sight (Fig. 1). Recovering images of hidden objects from indirect light paths involves a challenging inverse problem because there are infinitely many such paths to consider. With applications in remote sensing and machine vision, NLOS imaging could enable capabilities for a variety of imaging systems.

The challenging task of imaging objects that are partially or fully obscured from view has been tackled with approaches based on time-gated imaging[2], coherence gating[3], speckle correlation[4,5], wavefront shaping[6], ghost imaging[7,8], structured illumination[9] and intensity imaging[10,11]. At macroscopic scales, the most promising NLOS imaging systems rely on time-resolved detectors[12–20]. However, NLOS imaging with time-resolved systems remains a hard problem for three main reasons. First, the reconstruction step is prohibitively computationally demanding, in terms of both memory requirements and processing cycles. Second, the flux of multiply scattered light is extremely low, requiring either extensive acquisition times in dark environments or a sufficiently high-power laser to overcome the contribution of ambient light. Finally, NLOS imaging often requires a custom hardware system made with expensive components, thus preventing its widespread use.

Confocal NLOS (C-NLOS) imaging aims to overcome these challenges. Whereas previous NLOS acquisition setups exhaustively illuminate and image pairs of distinct points on a visible surface (such as a wall), the proposed system illuminates and images the same point (Fig. 1) and raster-scans this point across the wall to acquire a 3D transient (that is, time-resolved) image[14,25–27]. C-NLOS imaging offers several advantages over existing methods. First, it facilitates the derivation of a closed-form solution to the NLOS problem. The proposed NLOS reconstruction procedure is several orders of magnitude faster and more memory-efficient than previous approaches, and it also produces higher-quality reconstructions. Second, whereas indirectly scattered light remains extremely weak for diffuse objects, retroreflective objects (such as road signs, bicycle reflectors and high-visibility safety apparel) considerably increase the indirect signal by reflecting light back to its source with minimal scattering. This retroreflectance property can only be exploited by confocalized systems that simultaneously illuminate and image a common point and may be the enabling factor towards making NLOS imaging practical in certain applications (such as autonomous driving). Third, LIDAR systems already perform confocal scanning to acquire point clouds from direct light paths. Our prototype system was built from the ground up, but commercial LIDAR systems may be capable of supporting the algorithms developed here with minimal hardware modifications.

Similarly to other NLOS imaging approaches, our image formation model makes the following assumptions: there is only single scattering behind the wall (that is, no inter-reflections in the hidden part of the scene), light scatters isotropically (that is, the model ignores Lambert's cosine terms), and no occlusions occur within the hidden scene. Our approach also supports retroreflective materials through a minor modification of the image formation model.

C-NLOS measurements consist of a two-dimensional set of temporal histograms, acquired by confocally scanning points x', y' on a planar wall at position $z' = 0$. This 3D volume of measurements, τ, is given by

$$\tau(x', y', t) = \iiint_{\Omega} \frac{1}{r^4}\, \rho(x, y, z)\, \delta\big(2\sqrt{(x'-x)^2 + (y'-y)^2 + z^2} - tc\big)\, \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \tag{1}$$

where c is the speed of light. Every measurement sample $\tau(x', y', t)$ captures the photon flux at point (x', y') and time t relative to an incident pulse scattered by the same point at time $t = 0$. Here, the function ρ is the albedo of the hidden scene at each point (x, y, z) with $z > 0$ in the 3D half-space Ω. The Dirac delta function δ represents the surface of a spatio-temporal four-dimensional hypercone given by $x^2 + y^2 + z^2 - (tc/2)^2 = 0$, which models light propagation from the wall to the object and back to the wall. It is also closely related to Minkowski's light cone[28], which is a geometric

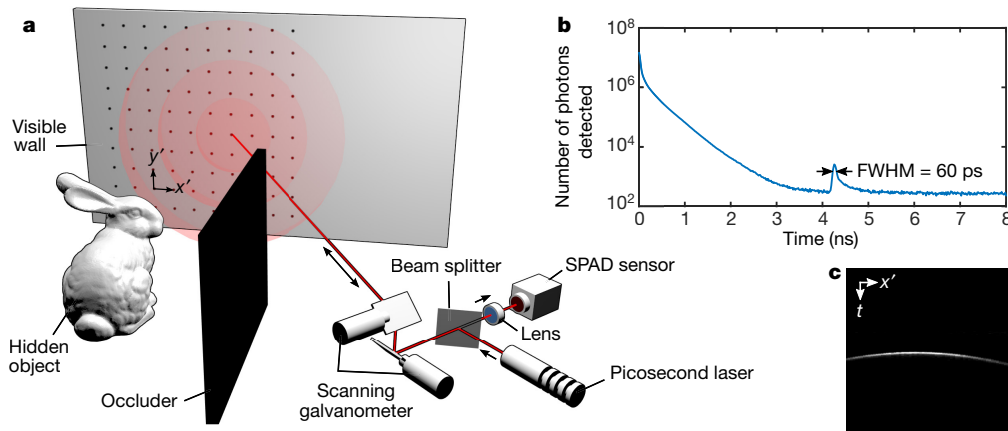[1]Department of Electrical Engineering, Stanford University, Stanford, California 94305 USA.

Figure 1 | Overview of confocal imaging hardware and measurements.
a, A pulsed laser and time-resolved detector raster-scan a wall to record both the direct light reflecting off the wall and the indirect light from a hidden object. **b**, A histogram measured at a scanned point on the visible wall indicates the temporal precision of the detector. In this experiment, the hidden object is a 5 cm × 5 cm square made from retroreflective tape. The detection time of the indirect signal ($t = 4.27$ ns) relative to the direct

signal ($t = 0$ ns) corresponds to twice the distance of the hidden object from the scanned point ($r = 0.64$ m). FWHM, full-width at half-maximum. **c**, Scanning a sequence of points along the wall produces a 'streak image' that captures the spatio-temporal geometry of indirect light transport. Each column in this image represents the histogram measured at a discrete point ($x',0$) on a wall and contains the indirect light from the hidden square.

representation of light propagation through space and time. We note that the function is shift-invariant in the x and y axes, but not in the z axis. A feature of this formulation is that the distance function $r = \sqrt{(x'-x)^2 + (y'-y)^2 + z^2} = tc/2$ can be expressed in terms of the arrival time t; the radiometric term $1/r^4$ can thus be pulled out of the triple integral. Equation (1) can also be modified to model retro-reflective materials by replacing $1/r^4$ with $1/r^2$, which represents a large increase in the flux of the indirect light (see Supplementary Information for details).

The most remarkable property of equation (1) is the fact that a change of variables in the integral by $z = \sqrt{u}$, $dz/du = 1/(2\sqrt{u})$ and $v = (tc/2)^2$ results in

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

which can be expressed as a straightforward 3D convolution, where $\mathcal{R}_t\{\tau\} = h * \mathcal{R}_z\{\rho\}$. Here, the function h is a shift-invariant 3D convolution kernel, the transform $\mathcal{R}_z$ nonuniformly resamples and attenuates the elements of volume ρ along the z axis, and the transform $\mathcal{R}_t$ non-uniformly resamples and attenuates the measurements τ along the time axis. The inverses of both $\mathcal{R}_z$ and $\mathcal{R}_t$ also have closed-form expressions. We refer to equation (2) as the light-cone transform (LCT).

The image formation model can be discretized as $R_t\tau = HR_z\rho$, where $\tau \in \mathbb{R}_+^{n_x n_y n_t}$ is the vectorized representation of the measurements, and $\rho \in \mathbb{R}_+^{n_x n_y n_z}$ is the vectorized volume of the albedos of the hidden

surface. The process of discretizing each function involves defining a finite grid and integrating the function over each cell in the grid. The matrix $H \in \mathbb{R}_+^{n_x n_y n_h \times n_x n_y n_h}$ represents the shift-invariant 3D convolution operation, and the matrices $R_t \in \mathbb{R}_+^{n_x n_y n_h \times n_x n_y n_t}$ and $R_z \in \mathbb{R}_+^{n_x n_y n_h \times n_x n_y n_z}$ represent the transformation operations applied to the temporal and spatial dimensions, respectively. We note that both transformation matrices are independently applied to their respective dimension and can therefore be applied to large-scale datasets in a computationally and memory-efficient way. Similarly, the 3D convolution operation H can be computed efficiently in the Fourier domain. Together, these matrices represent the discrete LCT.

By treating NLOS imaging as a spatially invariant 3D deconvolution problem, a closed-form solution can be derived from the convolution theorem. The convolution operation is expressed as an element-wise multiplication in the Fourier domain and inverted according to

$$\boldsymbol{\rho}_* = R_z^{-1} F^{-1}\left(\frac{1}{\widehat{H}}\frac{|\widehat{H}|^2}{|\widehat{H}|^2 + \frac{1}{\alpha}}\right) F R_t \boldsymbol{\tau} \quad (3)$$

where F is the 3D discrete Fourier transform, $\boldsymbol{\rho}_*$ is the estimated volume of the albedos of the hidden surface, $\widehat{H}$ is a diagonal matrix containing the Fourier coefficients of the 3D convolution kernel, and α represents the frequency-dependent signal-to-noise ratio of the measurements. This approach is based on Wiener filtering[29], which minimizes the mean squared error between the reconstructed volume and the ground truth. As α approaches infinity, the formulation above becomes an inverse filter (that is, the filter applied in the frequency domain is $1/\widehat{H}$). Similarly, the Fourier-domain filter in equation (3) could be replaced by $\widehat{H}^*$ to implement a backprojection reconstruction procedure.
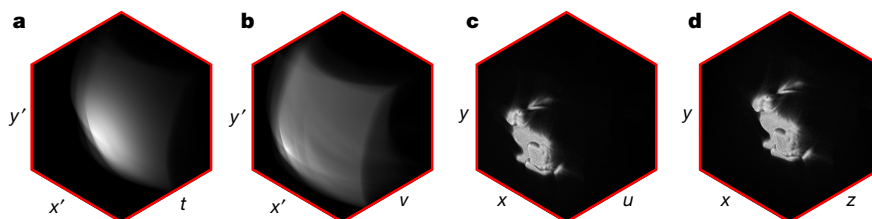

Figure 2 | Overview of the reconstruction procedure. The confocal measurements of the wall τ (**a**) are resampled and attenuated along the time axis, yielding $R_t\tau$ (**b**). These measurements are then convolved with

a Wiener filter to produce the volume $R_z\rho_*$ (**c**), and the result is resampled and attenuated along the depth dimension to produce the hidden volume ρ_* (**d**). Bunny model from the Stanford Computer Graphics Laboratory.
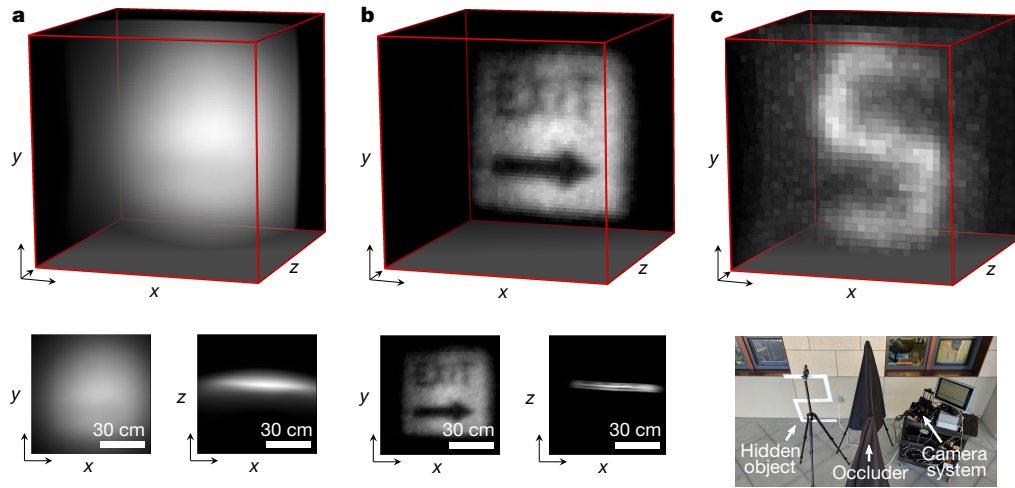
Figure 3 | NLOS reconstructions from SPAD measurements. a, Result for a hidden 'Exit' sign, obtained using the backprojection method. **b,** Result of the proposed LCT reconstruction procedure. **c,** The proposed method can also reconstruct the shape and albedo of objects outdoors, under indirect sunlight. The bottom right panel is a photograph of the experimental setup, which consists of a hidden 'S'-shaped object, black cloth acting as an occluder and the confocal scanning prototype.

Wiener filtering with a constant α inaccurately assumes that the transformed measurements contain white noise. Therefore, we also derive an iterative reconstruction procedure that combines the LCT with a physically accurate Poisson noise model (Supplementary Information).

Figure 2 illustrates the inverse LCT applied to indirect measurements of a bunny model simulated with a physically based ray tracer[30]. The process involves evaluating equation (3) in three steps: (i) resampling and attenuating the measurements τ with the transform R_t, (ii) applying the Wiener filter to the result, and (iii) applying the inverse transform R_z^{-1} to recover ρ. These three steps are efficient in terms of memory and number of operations required. The most costly step is the application of the Wiener filter, which requires $O(N^3 \log N)$ operations for the 3D fast-Fourier transforms and has memory requirements of $O(N^3)$, where N is the maximum number of elements across all dimensions in space-time. In comparison, existing backprojection-type reconstructions[15–17] require $O(N^5)$ operations, and methods based on inversion are much more costly both in their memory and processing requirements[17,18,20].

In addition to improved runtime and memory efficiency, a primary benefit of the LCT over backprojection-based approaches is that the inverted solution is accurate. In Fig. 3, we compare the reconstruction quality of the backprojection algorithm and the LCT for a retroreflective traffic sign. The dimensions of the hidden sign are 0.61 m × 0.61 m and the diffuse wall is sampled at 64 × 64 locations over a 0.8 m × 0.8 m region. The total exposure time is 6.8 min (that is, 0.1 s per sample) and the runtime for MATLAB to recover a volume of 64 × 64 × 512 voxels is 1 s on a MacBook Pro (3.1-GHz Intel Core i7). To compare the reconstruction quality of the two methods, we compute the backprojection result using the LCT, which is just as

efficient as inverting the problem with the LCT. Even though unfiltered backprojection could be slightly sharpened by linear filters, such as a Laplacian[15], backprojection methods do not solve the inverse problem (see Supplementary Information for detailed comparisons). In Supplementary Information, we also show a variety of reconstructed example scenes, as well as results for NLOS tracking[11–13] of retroreflective objects in real time.

Applying NLOS imaging outdoors requires the indirect light from the hidden object to be detected in the presence of strong ambient illumination. To accomplish this, C-NLOS imaging takes advantage of the high light throughput associated with retroreflective objects. Figure 3 presents an outdoor NLOS experiment under indirect sunlight (approximately 100 lx). The dimensions of the hidden retroreflective object are 0.76 m × 0.51 m, with 32 × 32 sampled locations over a 1 m × 1 m area. The exposure is 0.1 s per sample, with a total exposure time of 1.7 min. MATLAB reconstructs a volume of 32 × 32 × 1,024 voxels in 0.5 s.

The fundamental bounds on the resolution of NLOS imaging approaches couple the full-width at half-maximum of the temporal resolution of the imaging system, represented by the scalar γ, to the smallest resolvable axial Δz and lateral Δx spatial feature size as follows

$$\Delta z \geq \frac{c\gamma}{2} \quad \text{and} \quad \Delta x \geq \frac{c\sqrt{w^2+z^2}}{2w}\gamma \tag{4}$$

where $2w$ is the sampled width or height of the visible wall (see Supplementary Information for details).

To evaluate the limits of the reconstruction procedure, we simulate the acquisition of 1,024 × 1,024 points sampled over a 1 m × 1 m area
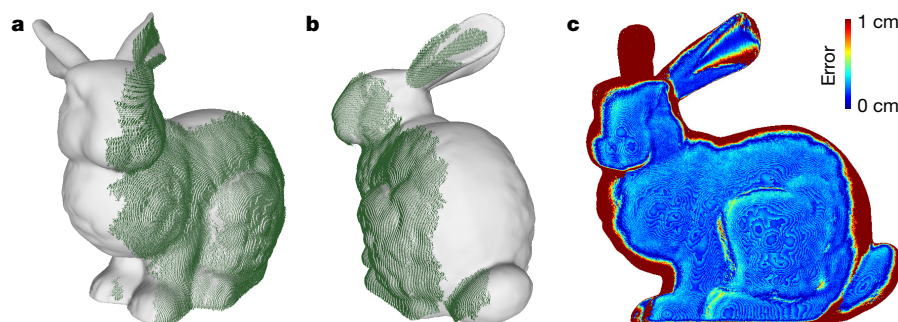

Figure 4 | Comparison between simulated C-NLOS reconstruction and ground-truth geometry. a, b, Rendered point clouds reconstructed with the LCT (green) over the ground-truth geometry (grey). **c,** Pointwise difference between the surfaces along the z axis.

and 1,024 time bins with a temporal resolution of 8 ps per bin. We recover a volume containing $1{,}024 \times 1{,}024 \times 1{,}024$ voxels. Figure 4 shows the target geometry in grey and the recovered shape overlaid in green. The error map indicates a median absolute reconstruction error of 2.5 mm (mean absolute error 15.1 mm, mean square error 2.7 mm). Occlusions and higher-order bounces of indirect illumination are not modelled by any existing NLOS imaging method, including ours, which may lead to violations in the image formation model and errors in the reconstructed volume. For example, the right ear of the bunny is not accurately recovered owing to self-occlusions by the left ear in the measurements. We note that the conventional approach of discretizing and inverting the image formation model at this resolution would require an excess of 9 petabytes of memory just to store a sparse representation of the linear system.

The co-design of a confocal scanning technique and a computationally efficient inverse method facilitates fast, high-quality reconstructions of hidden objects. To achieve real-time frame rates with C-NLOS imaging, three improvements to our current prototype are required. First, to reduce acquisition time, a more powerful laser is needed. For eye-safe operation, this laser may need to operate in the short-wave infrared regime[11,12,22]. Second, for retroreflective objects, the measurement of multiple histograms can be performed in parallel, with minimal crosstalk. This property could enable a single-photon avalanche diode (SPAD) array and a diffused laser source to acquire the full C-NLOS image in a single shot. Third, to improve the computation time, our highly parallelizable algorithm could be implemented in a graphics processing unit or a field-programmable gate array.

The proposed technique thus enables NLOS imaging with conventional hardware at much higher speeds, with a smaller memory footprint and lower power consumption, over a longer range, under ambient lighting and at higher resolution than any existing approach of which we are aware.

1. Freund, I. Looking through walls and around corners. *Physica A* **168,** 49–65 (1990).
2. Wang, L., Ho, P. P., Liu, C., Zhang, G. & Alfano, R. R. Ballistic 2-D imaging through scattering walls using an ultrafast optical Kerr gate. *Science* **253,** 769–771 (1991).
3. Huang, D. *et al.* Optical coherence tomography. *Science* **254,** 1178–1181 (1991).
4. Bertolotti, J. *et al.* Non-invasive imaging through opaque scattering layers. *Nature* **491,** 232–234 (2012).
5. Katz, O., Heidmann, P., Fink, M. & Gigan, S. Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations. *Nat. Photon.* **8,** 784–790 (2014).
6. Katz, O., Small, E. & Silberberg, Y. Looking around corners and through thin turbid layers in real time with scattered incoherent light. *Nat. Photon.* **6,** 549–553 (2012).
7. Strekalov, D. V., Sergienko, A. V., Klyshko, D. N. & Shih, Y. H. Observation of two-photon "ghost" interference and diffraction. *Phys. Rev. Lett.* **74,** 3600–3603 (1995).
8. Bennink, R. S., Bentley, S. J. & Boyd, R. W. "Two-photon" coincidence imaging with a classical source. *Phys. Rev. Lett.* **89,** 113601 (2002).
9. Sen, P. *et al.* Dual photography. *ACM Trans. Graph.* **24,** 745–755 (2005).
10. Bouman, K. L. *et al.* In *IEEE 16th Int. Conference on Computer Vision* 2270–2278 (IEEE, 2017); http://openaccess.thecvf.com/content_iccv_2017/html/Bouman_Turning_Corners_Into_ICCV_2017_paper.html.
11. Klein, J., Peters, C., Martin, J., Laurenzis, M. & Hullin, M. B. Tracking objects outside the line of sight using 2D intensity images. *Sci. Rep.* **6,** 32491 (2016).
12. Chan, S., Warburton, R. E., Gariepy, G., Leach, J. & Faccio, D. Non-line-of-sight tracking of people at long range. *Opt. Express* **25,** 10109–10117 (2017).
13. Gariepy, G., Tonolini, F., Henderson, R., Leach, J. & Faccio, D. Detection and tracking of moving objects hidden from view. *Nat. Photon.* **10,** 23–26 (2016).
14. Kirmani, A., Hutchison, T., Davis, J. & Raskar, R. In *IEEE 12th Int. Conference on Computer Vision* 159–166 (IEEE, 2009); http://ieeexplore.ieee.org/document/5459160/.
15. Velten, A. *et al.* Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nat. Commun.* **3,** 745 (2012).
16. Buttafava, M., Zeman, J., Tosi, A., Eliceiri, K. & Velten, A. Non-line-of-sight imaging using a time-gated single photon avalanche diode. *Opt. Express* **23,** 20997–21011 (2015).
17. Gupta, O., Willwacher, T., Velten, A., Veeraraghavan, A. & Raskar, R. Reconstruction of hidden 3D shapes using diffuse reflections. *Opt. Express* **20,** 19096–19108 (2012).
18. Wu, D. *et al.* In *Computer Vision – ECCV 2012* (eds Fitzgibbon, A., *et al.*) 542–555 (Springer, 2012); https://link.springer.com/chapter/10.1007/978-3-642-33718-5_39.
19. Tsai, C.-Y., Kutulakos, K. N., Narasimhan, S. G. & Sankaranarayanan, A. C. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 7216–7224 (IEEE, 2017); http://openaccess.thecvf.com/content_cvpr_2017/html/Tsai_The_Geometry_of_CVPR_2017_paper.html.
20. Heide, F., Xiao, L., Heidrich, W. & Hullin, M. B. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 3222–3229 (IEEE, 2014); http://openaccess.thecvf.com/content_cvpr_2014/html/Heide_Diffuse_Mirrors_3D_2014_CVPR_paper.html.
21. Schwarz, B. LIDAR: mapping the world in 3D. *Nat. Photon.* **4,** 429–430 (2010).
22. McCarthy, A. *et al.* Kilometre-range, high resolution depth imaging via 1560 nm wavelength single-photon detection. *Opt. Express* **21,** 8904–8915 (2013).
23. Kirmani, A. *et al.* First-photon imaging. *Science* **343,** 58–61 (2014).
24. Shin, D. *et al.* Photon-efficient imaging with a single-photon camera. *Nat. Commun.* **7,** 12046 (2016).
25. Abramson, N. Light-in-flight recording by holography. *Opt. Lett.* **3,** 121–123 (1978).
26. Velten, A. *et al.* Femto-photography: capturing and visualizing the propagation of light. *ACM Trans. Graph.* **32,** 44 (2013).
27. O'Toole, M. *et al.* In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 2289–2297 (IEEE, 2017); http://openaccess.thecvf.com/content_cvpr_2017/html/OToole_Reconstructing_Transient_Images_CVPR_2017_paper.html.
28. Minkowski, H. Raum und zeit. *Phys. Z.* **10,** 104–111 (1909).
29. Wiener, N. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series* Vol. 7 (MIT Press, 1949).
30. Pharr, M., Jakob, W. & Humphreys, G. *Physically Based Rendering: From Theory to Implementation* 3rd edn (Morgan Kaufmann, 2017).

Author Contributions M.O'T. conceived the method, developed the experimental setup, performed the indoor measurements and implemented the LCT reconstruction procedure. M.O'T. and D.B.L. performed the outdoor measurements. D.B.L. applied the iterative LCT reconstruction procedures shown in Supplementary Information. G.W. supervised all aspects of the project. All authors took part in designing the experiments and writing the paper and Supplementary Information.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to M.O'T. (matthew.otoole@gmail.com) and G.W. (gordon.wetzstein@stanford.edu).

Reviewer Information *Nature* thanks D. Faccio, V. Goyal and M. Laurenzis for their contribution to the peer review of this work.