Recovering Three-dimensional Shape Around a Corner using Ultrafast Time-of-Flight Imaging

Andreas Velten[1], Thomas Willwacher[2], Otkrist Gupta[1],
Ashok Veeraraghavan[3], Moungi G. Bawendi[4], Ramesh Raskar[1*]

[*]Corresponding Author: raskar@media.mit.edu
[1]MIT Media Lab, 75 Amherst St., Cambridge, MA 02139
[2]Harvard University, Department of Mathematics,
One Oxford Street, Cambridge, MA 02138
[3]Rice University, Department of Electrical and
Computer Engineering, 6100 Main Street, Houston, TX 77005
[4]MIT, Department of Chemistry,
77 Massachusetts Ave., Cambridge, MA 02139

Tuesday 13th March, 2012

Abstract

The recovery of objects obscured by scattering is an important goal in imaging and has been approached by exploiting coherence properties, ballistic photons or penetrating wavelengths, for example. Common methods use scattered light transmitted through an occluding material, although these fail if the occluder is opaque. Light is scattered by transmission through objects, but also by multiple reflection from diffuse surfaces in a scene. This reflected light contains information about the scene that becomes mixed by the diffuse reflections before reaching the image sensor. This mixing is difficult to decode using traditional cameras. Here we show the combination of a time-of-flight technique and computational reconstruction algorithms to untangle image information mixed by diffuse reflection. We demonstrate a 3D range camera able to look around a corner using diffusely reflected light that achieves sub-millimeter depth precision and centimeter lateral precision over 40 cm x 40 cm x 40 cm of hidden space.

1 Introduction

The light detected on an image sensor is composed of light from objects that are in the line of sight, reflected light from objects that are not within the line of sight, and background light. Light from objects outside the line of sight reaches the sensor via multiple reflections (or bounces). In conventional imaging, it is difficult to exploit this non-line-of-sight light if the reflections or bounces are diffuse.

Line-of-sight time of flight information is commonly used in LIDAR (LIght Detection And Ranging) [1] and two dimensional gated viewing [2] to determine the object distance or to reject unwanted scattered light. By considering only the early ballistic photons from a sample, these methods can image through turbid media or fog [3]. Other methods, like coherent LIDAR [4], exploit the coherence of light to determine the time of flight. However, light that has undergone multiple diffuse reflections, has diminished coherence.

Recent methods in computer vision and inverse light transport study multiple diffuse reflections in free-space. Dual Photography [5] shows one can exploit scattered light to recover 2D images of objects illuminated by a structured dynamic light source and hidden from the camera. Time gated viewing using mirror reflections allows imaging around corners, for example from a glass window [6, 7, 8]. Three bounce analysis of a time-of-flight camera can recover hidden 1-0-1 planar barcodes [9, 10] but the technique assumes well separated isolated hidden patches with known correspondence between hidden patches and recorded pulses. Similar to these and other inverse light transport approaches [11], we use a light source to illuminate one scene spot at a time and record the reflected light after its interaction with the scene.

We demonstrate an incoherent ultrafast imaging technique to recover 3D shapes of non-line-of-sight objects using this diffusely reflected light. We illuminate the scene with a short pulse and use the time of flight of returning light as a means to analyze direct and scattered light from the scene. We show that the extra temporal dimension of the observations under very high temporal sampling rates makes the hidden 3D structure observable. With a single or a few isolated hidden patches, pulses recorded after reflections are distinct and can be easily used to find 3D position of the hidden patches. However, with multiple hidden scene points, the reflected pulses may overlap in both space and time when they arrive at the detector. The loss of correspondence between 3D scene points and their contributions to the detected pulse stream is the main technical challenge. We present a computational algorithm based on back-projection to invert this process. Our main contributions are two-fold. We introduce the new problem of recovering the 3D structure of a hidden object and we show that the 3D information is retained in the temporal dimension after multi-bounce interactions between visible and occluded parts. We also present an experimental realization of the ability to recover the 3D structure of a hidden object, thereby demonstrating a 3D range camera able to look around a corner. The ability to record 3D shapes beyond the line of sight can potentially be applied in industrial inspection, endoscopy, disaster relief scenarios or more generally in situations where direct imaging of a scene is impossible.

2 Results

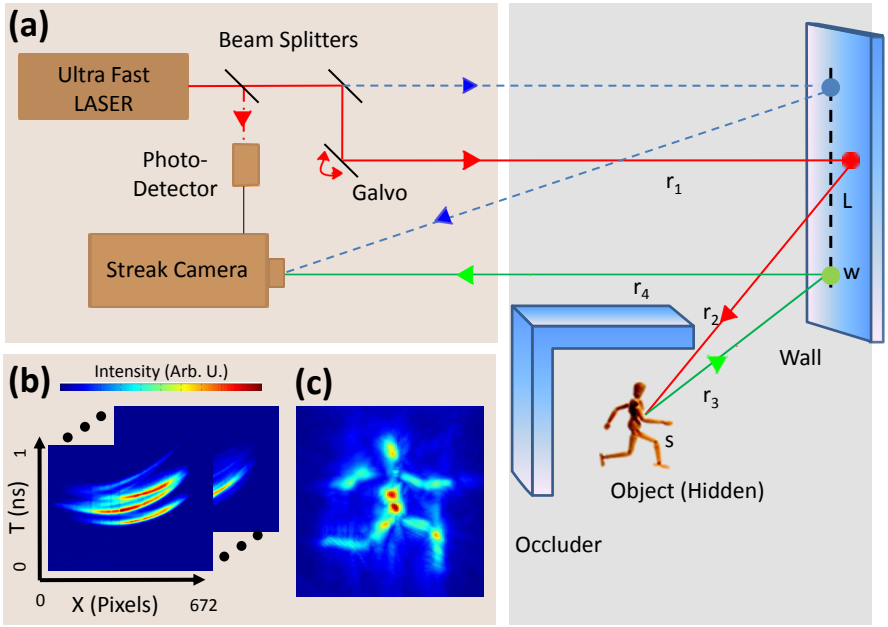


Figure 1: Experimental Setup *(a) The capture process. We capture a series of images by sequentially illuminating a single spot on the wall with a pulsed laser and recording an image of the dashed line segment on the wall with a streak camera. The laser pulse travels a distance r_1 to strike the wall at a point L, some of the diffusely scattered light strikes the hidden object (for example at s after traveling a distance r_2), returns to the wall (for example at w, after traveling over r_3) and is collected by the camera after traveling the final distance r_4 from w to the camera center of projection. The position of the laser beam on the wall is changed by a set of galvanometer actuated mirrors. (b) An example of streak images sequentially collected. Intensities are normalized against a calibration signal. Red corresponds to the maximum, blue to the minimum intensities. (c) The 2D projected view of the 3D shape of the hidden object, as recovered by the reconstruction algorithm. Here the same color map corresponds to backprojected filtered intensities or confidence values of finding an object surface at the corresponding voxel.*

Imaging Process The experimental setup is shown in Figure 1. Our scene consists of a 40 cm high and 25 cm wide wall referred to as diffuser wall. We use an ultrafast laser and a streak camera and both are directed at this wall. As a time reference we also

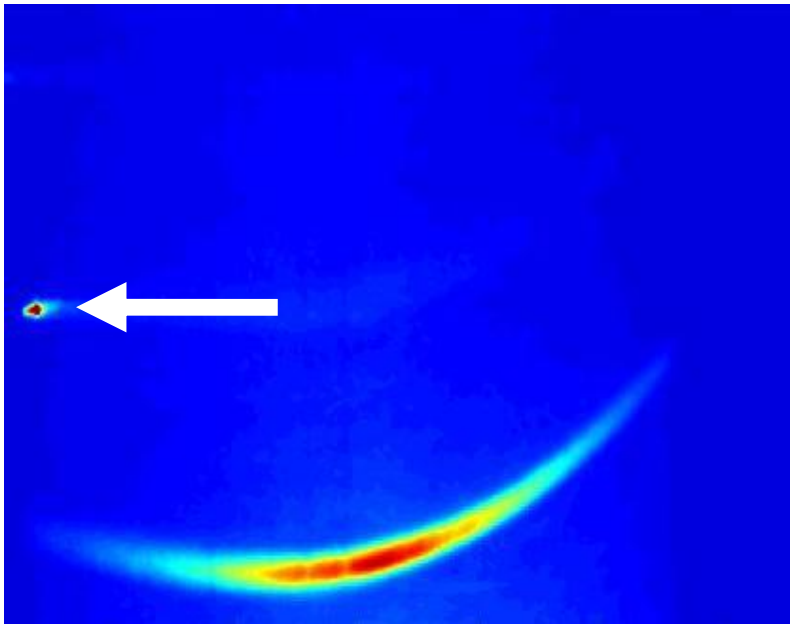


Figure 2: Streak image with calibration spot *The calibration spot in a streak image (highlighted with an arrow). The calibration spot is created by an attenuated beam split off the laser beam that strikes the wall in the field of view of the camera. It allows to monitor the long term stability of the system and calibrate for drifts in timing synchronization.*

direct an attenuated portion of the laser beam into the field of view of the streak camera (see Figure 2). The target object is hidden in the scene (mannequin in Figure 1) so that direct light paths between the object and the laser or the camera are blocked. Our goal is to produce three dimensional range data of the target object.

The streak camera records a streak image with one spatial dimension and one temporal dimension. We focus the camera on the dashed line segment on the diffuser wall shown in Figure 1a. We arrange the scanning laser to hit spots on the wall above or below this line segment so that single bounce light does not enter the camera. Though the target object is occluded, light from the laser beam is diffusely reflected by the wall, reaches the target object, is reflected by multiple surface patches and returns back to the diffuser wall, where it is reflected again and captured by the camera. In a traditional camera, this image would contain little or no information about the occluded target object (see also Supplementary Figure S6 and Supplementary Methods).

In our experimental setup, the laser emits 50 fs long pulses. The camera digitizes information in time intervals of 2 ps. We assume the geometry of the directly visible part of the setup is known. Hence the only unknown distances in the path of the laser pulses

are those from the diffuser wall to the different points on the occluded target object and back (paths r_2 and r_3 in Fig. 1). The 3D geometry of the occluded target is thus encoded in the streak images acquired by the camera and decoded using our reconstruction algorithm. The recorded streak images lack correspondence information, i.e., we do not know which pulses received by the camera came from which surface point on the target object. Hence, a straightforward triangulation or trilateration to determine the hidden geometry is not possible.

Consider a simple scenario with a small hidden patch as illustrated in Figure 2a. It provides intuition on how the geometry and location of the target object are encoded in the streak images. The reflected spherical pulse propagating from the hidden patch arrives at the points on the diffuser wall with different time delays and creates a hyperbolic curve in the space-time streak image. We scan and successively change the position of the laser spot on the diffuser wall. The shape and position of the recorded hyperbolic curve varies accordingly. Each pixel in a streak image corresponds to a finite area on the wall and a 2 ps time interval, a discretized space-time bin. However, the effective time resolution of the system is 15 ps due to a finite temporal point spread function of the camera. The detailed description of image formation is included in the Supplementary Methods Section.

The inverse process to recover the position of the small hidden patch from the streak images is illustrated in Figure 4b–e. Consider three pixels p, q and r in the streak image at which non zero light intensity is measured (Figure 2a). The possible locations in the world which could have contributed to a given pixel lie on an ellipsoid in Cartesian space. The foci of this ellipsoid are the laser spot on the diffuser wall and the point on the wall observed by that pixel. For illustration, we draw only a 2D slice of the ellipsoid, i.e., an ellipse, in Figure 2b. The individual ellipses from each of the three pixels p, q and r intersect at a single point. In the absence of noise, the intersection of three ellipses uniquely determines the location of the hidden surface patch that contributed intensity to the three camera pixels. In practice we lack correspondence, i.e., we do not know whether or not light detected at two pixels came from the same 3D surface point.

Therefore, we discretize the Cartesian space into voxels and compute the likelihood of the voxel being on a hidden surface. For each voxel, we determine all streak image pixels that could potentially have received contributions of this voxel based on the time of flight $r_1 + r_2 + r_3 + r_4$ and sum up the measured intensity values in these pixels. In effect, we let each pixel vote for all points on the corresponding ellipsoid. The signal energy contributed by each pixel is amplified by a factor to compensate for the distance attenuation. If the distance attenuation factor were not accounted for, the scene points that are far away from the wall would be attenuated by a factor of $(r_2 r_3)^2$ and would be lost during the reconstruction. Therefore, we amplify the contribution of each pixel to a particular voxel by a factor of $(r_2 r_3)^\alpha$ before backprojection. Reconstruction quality depends weakly on the value of α. We experimented with various values of α and found that $\alpha = 1$ is a good choice for reduced computation time. This process of computing likelihood by summing up weighted intensities is called backprojection [12]. We call the resulting 3D scalar function on voxels a heatmap.

The summation of weighted intensities from all pixels in a single streak image cre-

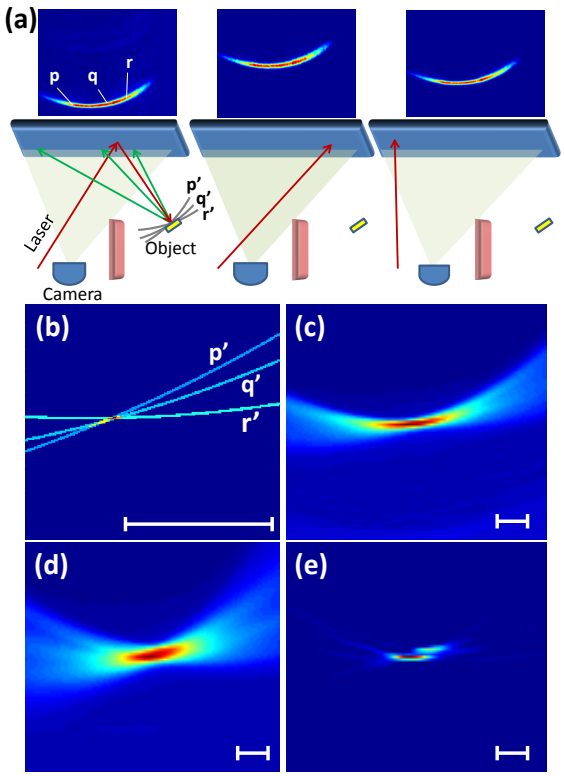


Figure 3: Reconstruction Algorithm *An illustrative example of geometric reconstruction using streak camera images. (a) Data capture. The object to be recovered consists of a 2 cm $\times$ 2 cm size square white patch beyond the line of sight (i.e. hidden). The patch is mounted in the scene and data is collected for different laser positions. The captured streak images corresponding to three different laser positions are displayed in the top row. Shapes and timings of the recorded response vary with laser positions and encode the position and shape of the hidden patch. (b) Contributing voxels in Cartesian space. For recovery of hidden position, consider the choices of contributing locations. The possible locations in Cartesian space that could have contributed intensity to the streak image pixels p, q, r are the ellipses p', q', r' (ellipsoids in 3D). For illustration, these three ellipse sections are also shown in (a) bottom left in Cartesian coordinates. If there is a single world point contributing intensity to all 3 pixels, the corresponding ellipses intersect, as is the case here. The white bar corresponds to 2 cm in all subfigures. (c) Backprojection and heatmap. We use a back-projection algorithm that finds overlayed ellipses corresponding to all pixels, Here we show summation of elliptical curves from all pixels in the first streak image. (d) Backprojection using all pixels in a set of 59 streak images. (e) Filtering. After filtering with a second derivative, the patch location and 2 cm lateral size are recovered.*

ates an approximate heatmap for the target patch (Figure 2c). Repeating the process for many laser positions on the diffuser wall, and using pixels from the corresponding streak images provides a better approximation (Figure 2d). In practice, we use approximately 60 laser positions. Traditional backprojection requires a high-pass filtering step. We use the second derivative of the data along the z direction of the voxel grid and approximately perpendicular to the wall as an effective filter and recover the hidden surface patch in Figure 2e. Because values at the voxels in the heatmap are the result of summing a large number of streak image pixels, the heatmap contains low noise and the noise amplification associated with a second derivative filter is acceptable.

Algorithm The first step of our imaging algorithm is data acquisition. We sequentially illuminate a single spot on the diffuser wall with a pulsed laser and record an image of the line segment of the wall with a streak camera. Then we estimate an oriented bounding box for the working volume to set up a voxel grid in Cartesian space (see Methods). In the backprojection, for each voxel, we record the summation of weighted intensities of all streak image pixels that could potentially have received contributions of this voxel based on the time of flight. We store the resulting three dimensional heatmap of voxels. The backprojection is followed by filtering. We compute a second derivative of the heatmap along the direction of the voxel grid facing away from the wall. In an optional post processing step, we compute a confidence value for each voxel by computing local contrast with respect to the voxel neighborhood in the filtered heatmap. To compute contrast, we divide each voxel heatmap value by the maximum in the local neighborhood. For better visualization, we apply a soft threshold on the voxel confidence value.

We estimate the oriented bounding box of the object in the second step by running the above algorithm at low spatial target resolution and with down sampled input data. Details of the reconstruction process and the algorithm can be found in the Methods Section as well as in the Supplementary Methods Section.

Reconstructions We show results of the 3D reconstruction for multi-part objects in Figures 5 and 3. The mannequin in Figure 5 contains non-planar surfaces with variations in depth and occlusions. We accurately recover all major geometrical features of the object. Figure 5i shows the reconstruction of the same object in slightly different poses to demonstrate the reproducibility and stability of the method as well as the consistency in the captured data. Note that the sporadic inaccuracies in the reconstruction are consistent across poses and are confined to the same 3D locations. The stop-motion animation in Supplementary Movie 1 shows the local nature of the missing or phantom voxels more clearly. Hence, the persistent inaccuracies are not due to signal noise or random measurement errors. This is promising as the voxel confidence errors are primarily due to limitations in the reconstruction algorithm and instrument calibration. These limitations can be overcome with more sophistication. However, our current method is limited to diffuse reflection from near-Lambertian opaque surfaces. Note also, that parts of the object that are occluded from the diffuser wall or facing away from it are not reconstructed.

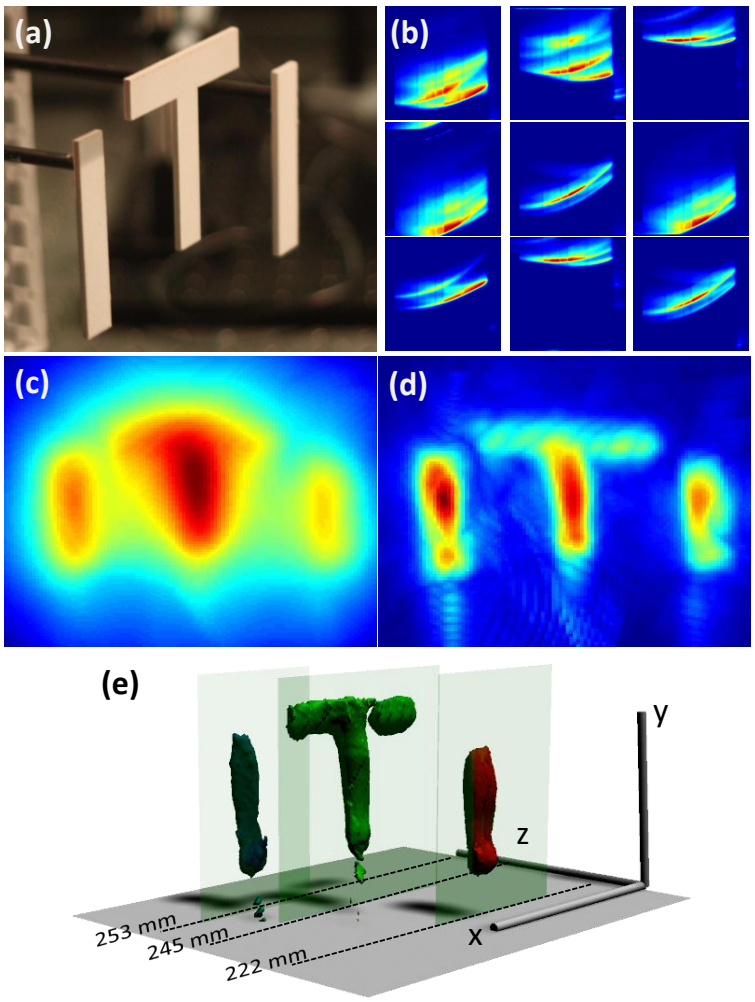


Figure 4: Depth in Reconstructions *Demonstration of the depth and lateral resolution. (a) The hidden objects to be recovered are three letters, I, T, I at varying depths. The "I" is 1.5 cm in wide and all letters are 8.2 cm high. (b) 9 of 60 images collected by the streak camera. (c) Projection of the heatmap on the x-y plane created by the back projection algorithm. (d) Filtering after computing second derivative along depth (z). The color in these images represents the confidence of finding an object at the pixel position. (e) A rendering of the reconstructed 3D shape. Depth is color coded and semi-transparent planes are inserted to indicate the ground truth. The depth axis is scaled to aid visualization of the depth resolution.*

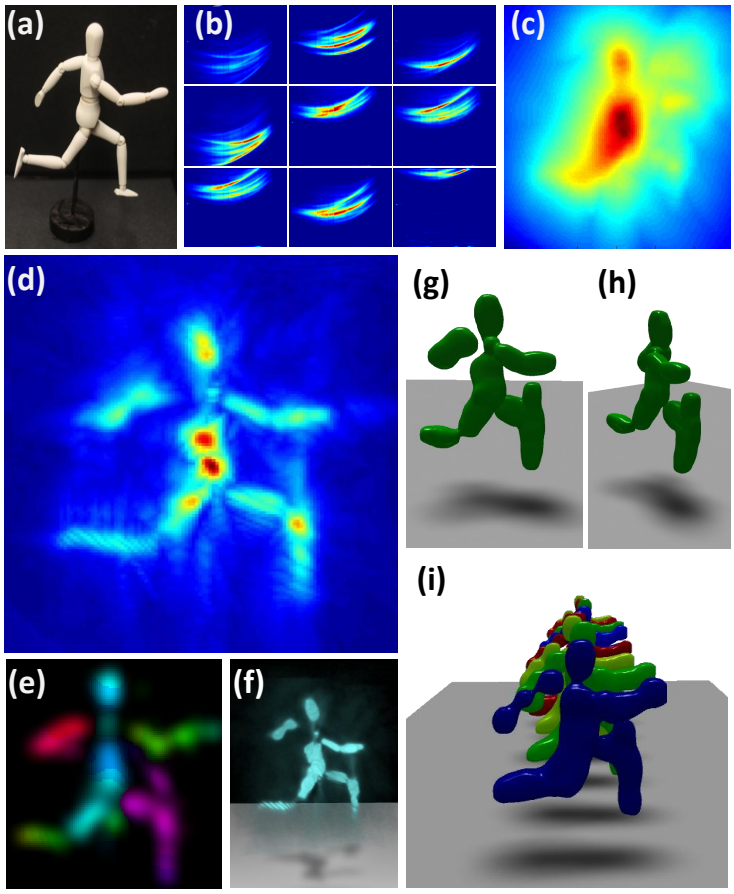


Figure 5: Complex Object Reconstruction *(a) Photo of the object. The mannequin is approximately 20 cm tall and is placed about 25 cm from the diffuser wall. (b) Nine of the 60 raw streak images. (c) Heatmap. Visualization of the heatmap after backprojection. The maximum value along the z direction for each x-y coordinate in Cartesian space. The hidden shape is barely discernible. (d) Filtering. The second derivative of the heatmap along depth (z) projected on the x-y plane reveals the hidden shape contour. (e) Depth map. Color encoded depth (distance from the diffuser wall) shows the left leg and right arm closer in depth compared to the torso and other leg and arm. (f) Confidence map. A rendered point cloud of confidence values after soft threshold. Images (g-h) show the object from different viewpoints after application of a volumetric blurring filter. (i) The stop-motion animation frames from multiple poses to demonstrate reproducability. See Supplementary Movie 1 for an animation. Shadows and the ground plane in images (f-i) have been added to aid visualization.*

Figure 3 shows a reconstruction of multiple planar objects at different unknown depths. The object planes and boundaries are reproduced accurately to demonstrate depth and lateral resolution. The reconstruction is affected by several factors such as calibration, sensor noise, scene size and time resolution. Below, we consider them individually.

The sources of calibration errors are lens distortions on the streak camera that lead to a warping of the collected streak image, measurement inaccuracies in the visible geometry, and measurement inaccuracies of the center of projection of the camera and the origin of the laser. For larger scenes, the impact of static calibration errors would be reduced.

The sensor introduces intensity noise and timing uncertainty, i.e., jitter. The reconstruction of 3D shapes is more dependent on the accuracy of the time of arrival than the signal to noise ratio (SNR) in received intensity. Jitter correction, as described in the Methods section, is essential, but does not remove all uncertainties. Improving the SNR is desirable because it yields faster capture times. Similar to many commercial systems, e.g., LIDAR, the SNR could be significantly improved by using an amplified laser with more energetic pulses and a repetition rate in the kHz range and a triggered camera. The overall light power would not change, but fewer measurements for light collection could significantly reduce signal independent noise such as background and shot noise.

We could increase the scale of the system for larger distances and bigger target objects. By using a longer pulse, with proportionally reduced target resolution and increased aperture size one could build systems without any change in the ratio of received and emitted energy, i.e., the link budget. When the distance r_2 between diffuser wall and the hidden object (see Fig. 1) is increased without increasing the size of the object, the signal strength drops dramatically ($\propto 1/(r_2 r_3)^2$) and the size of the hidden scene is therefore limited. A configuration where laser and camera are very far from the rest of the scene is, however, plausible. A loss in received energy can be reduced in two ways. The laser beam can be kept collimated over relatively long distances and the aperture size of the camera can be increased to counterbalance a larger distance between camera and diffuser wall.

The timing resolution, along with spatial diversity in the positions of spots illuminated and viewed by the laser and the camera affects the resolution of 3D reconstructions. Additional factors include the position of the voxel in Cartesian space and the overall scene complexity. The performance evaluation subsection of the Supplementary Methods section describes depth and lateral resolution. In our system, translation along the direction perpendicular to the diffuser wall can be resolved with a resolution of 400 μm – better than the full width half maximum (FWHM) time resolution of the imaging system. Lateral resolution in a plane parallel to the wall is lower and is limited to 0.5-1 cm depending on proximity to the wall.

3 Discussion

This paper's goals are twofold: to introduce the new challenging problem of recovering the 3D shape of a hidden object and to demonstrate the results using a novel co-design of an electro-optic hardware platform and a reconstruction algorithm. Designing and implementing a prototype for a specific application will provide further, more specific data about the performance of our approach in real world scenarios. We have demonstrated the 3D imaging of a non-trivial hidden three dimensional geometry from scattered light in free space. We compensate for the loss of information in the spatial light distribution caused by the scattering process by capturing ultrafast time of flight information.

Our reconstruction method assumes that light is only reflected once by a discrete surface on the hidden object without inter-reflections within the object and without subsurface scattering. We further assume that light travels in a straight line between reflections. Light that does not follow these assumptions will appear as time-delayed background in our heatmap and will complicate, but not necessarily prevent reconstruction.

The application of imaging beyond the line of sight is of interest for sensing in hazardous environments such as inside machinery with moving parts, for monitoring highly contaminated areas such as the sites of chemical or radioactive leaks where even robots can not operate or need to be discarded after use [13]. Disaster response and search and rescue planning, as well as autonomous robot navigation can benefit from the ability obtain complete information about the scene quickly [14, 15]

A promising theoretical direction is in inference and inversion techniques that exploit scene priors, sparsity, rank, meaningful transforms and achieve bounded approximations. Adaptive sampling can decide the next-best laser direction based on a current estimate of the 3D shape. Further analysis will include coded sampling using compressive techniques and noise models for SNR and effective bandwidth. Our current demonstration assumes friendly reflectance and planarity of the diffuse wall.

The reconstruction of an image from diffusely scattered light is of interest in a variety of fields. Change in spatial light distribution due to the propagation through a turbid medium is in principle reversible [16] and allows imaging through turbid media via computational imaging techniques [17, 18, 19]. Careful modulation of light can shape or focus pulses in space and time inside a scattering medium [20, 21]. Images of objects behind a diffuse screen, such as a shower curtain, can be recovered by exploiting the spatial frequency domain properties of direct and global components of scattered light in free space [22]. Our treatment of scattering is different but could be combined with many of these approaches.

In the future, emerging integrated solid state lasers, new sensors and non-linear optics should provide practical and more sensitive imaging devices. Beyond 3D shape, new techniques should allow us to recover reflectance, refraction and scattering properties and achieve wavelength resolved spectroscopy beyond the line of sight. The formulation could also be extended to shorter wavelengths (e.g., x-rays) or to ultrasound and

sonar frequencies. The new goal of hidden 3D shape recovery may inspire new research in the design of future ultrafast imaging systems and novel algorithms for hidden scene reconstruction.

4 Methods

Capture Setup The light source is a Kerr lens mode-locked Ti:Sapphire laser. It delivers pulses of about 50 fs length at a repetition rate of 75 MHz. Dispersion in the optical path of the pulse does not stretch the pulse beyond the resolution of the camera of 2 ps and therefore can be neglected. The laser wavelength is centered at 795 nm. The main laser beam is focused on the diffuser wall with a 1 m focal length lens. The spot created on the wall is about 1 mm in diameter and is scanned across the diffuser wall via a system of two galvanometer actuated mirrors. A small portion of the laser beam is split off with a glass plate and is used to synchronize the laser and streak camera as shown in Figure 1. The diffuser wall is placed 62 cm from the camera. The mannequin object (Figure 3) is placed at a distance of about 25 cm to the wall, the letters (Figure 4) are placed at 22.2 cm to 25.3 cm.

For time jitter correction, another portion of the beam is split off, attenuated and directed at the wall as the calibration spot. The calibration spot is in the direct field of view of the camera and can be seen in Figure 2. The calibration spot serves as a time and intensity reference to compensate for drifts in the synchronization between laser and camera as well as changes in laser output power. It also helps in detecting occasional shifts in the laser direction due to, for example, beam pointing instabilities in the laser. If a positional shift is detected, the data is discarded and the system is recalibrated. The streak camera's photocathode tube, much like an oscillocope, has time decayed burn out and local gain variations. We use a reference background photo to divide and compensate.

The camera is a Hamamatsu C5680 streak camera that captures one spatial dimension, i.e., a line segment in the scene, with an effective time resolution of 15 ps and a quantum efficiency of about 10%. The position and viewing direction of the camera are fixed. The diffuser wall is covered with Edmund Optics NT83 diffuse white paint.

Reconstruction Technique We use a set of Matlab routines to implement the back-projection based reconstruction. Geometry information about the visible part of the scene, i.e., diffuser wall could be collected using our time of flight system. Reconstructing the three dimensional geometry of a visible scene using time of flight data is well known [2]. We omit this step and concentrate on the reconstruction of the hidden geometry. We use a FARO Gauge digitizer arm to measure the geometry of the visible scene and also to gather data about a sparse set of points on hidden objects for comparative verification. The digitizer arm data is used as ground truth for later independent verification of the position and shape of hidden objects as shown via transparent planes

in Figure 4e. After calibration, we treat the camera and laser as a rigid pair with known intrinsic and extrinsic parameters [23].

We estimate the oriented bounding box around the hidden object using a lower resolution reconstruction. We reduce the spatial resolution to 8 mm/voxel and downsample the input data by factor of 40. We can scan a 40 cm x 40 cm x 40 cm volume spanning the space in front of the wall in 2-4 s to determine the bounding box of a region of interest. The finer voxel grid resolution is 1.7 mm in each dimension. We can use the coarse reconstruction obtained to set up a finer grid within this bounding box. Alternatively we can set an optimized bounding box from the collected ground truth. To minimize reconstruction time, we used this second method in most of the published reconstructions. We confirmed that apart from the reconstruction time and digitization artifacts, both methods produce the same results. We compute the principal axis of this low resolution approximation and orient the fine voxel grid with these axes.

In the post-processing step, we use a common approach to improve the surface visualization. We estimate the local contrast and apply a soft threshold. The confidence value for a voxel is $V' = \tanh(20(V - V_0))V/m_{\mathrm{loc}}$, where V is the original voxel value in filtered heatmap and m_{loc} is a local maximum computed in a 20 x 20 x 20 voxel sliding window around the voxel under consideration. Division by m_{loc} normalizes for local contrast. The value V_0 is a global threshold and set to 0.3 times the global maximum of the filtered heatmap. The $\tanh$ function achieves a soft threshold.

System SNR The laser emits a pulse every 13.3 ns (75 MHz) and consequently the reflected signal repeats at the same rate. We average 7.5 million such 13.3 ns windows in a 100 ms exposure time on our streak tube readout camera. We add 50 to 200 such images to minimize noise from the readout camera. The light returned from a single hidden patch is attenuated in the second, third and fourth path segments. In our setups this attenuation factor is 10^{-8}. Attenuation in the fourth path segment can be partially counteracted by increasing the camera aperture.

Choice of laser positions on the wall Recall that we direct the laser to various positions on the diffuser wall and capture one streak image for each position. The position of a hidden point s (Figure 1) is determined with highest confidence along the normal N to an ellipsoid through s with foci at the laser spot, L, and the wall point, w. Large angular diversity through a wide range of angles for N for all such pairs to create baselines is important. Note also that N is the angle bisector of $\angle Lsw$.

The location and spacing of the laser positions on the wall can have a big impact on reconstruction performance. To obtain good results one should choose the laser positions so as to provide good angular diversity. We use 60 laser positions in 3-5 lines perpendicular to the line on the wall observed by our 1D streak camera. This configuration yielded significantly better results than putting the laser positions on few lines parallel to the camera line.

Scaling the system Scaling up the distances in the scene is challenging because higher resolution and larger distances lead to disproportionately less light being transferred through the scene. A less challenging task may be to scale the entire experiment including the hidden object, the pulse length, the diffuser wall and the camera aperture. The reduction in resolution to be expected in this scaling should be equal to the increase in size of the hidden object.

To understand this consider a hidden square patch in the scene. To resolve it we require discernible light to be reflected back from that patch after reflections or bounces off other patches. Excluding the collimated laser beam, there are a three paths as described earlier. For each path, light is attenuated by approximately $d^2/(2\pi r^2)$, where r is the distance between the source and the destination patch and d is the side length of the destination patch. For the fourth segment, the destination patch is the camera aperture and d denotes the size of this aperture. If r and d are scaled together for any path, the contributed energy from the source patch to the destination patch does not change. This may allow us to scale the overall system to larger scenes without a prohibitively drastic change in performance. However, increasing the aperture size is only possible to a certain extend.

Non-Lambertian surfaces Our reconstruction method is well-suited for Lambertian reflectance of surfaces. Our methods is also robust for near-Lambertian surfaces, e.g., surfaces with a large diffuse component, and they are implicitly handled in our current reconstruction algorithm. The surface reflectance profile only affects the relative weight of the backprojection ellipses and not their shapes. The shape is dictated by the time of flight which is independent of the reflectance distribution.

Surfaces that are highly specular, retroreflective or have a low reflectance make the hidden shape reconstruction challenging. Highly specular, mirror-like and retroreflective surfaces limit the regions illuminated by the subsequent bounces and may not reflect enough energy back to the camera. They also could cause dynamic range problems. Subsurface scattering or additional inter-reflections extend the fall time in reflected time profile of a pulse. But the onset due to reflection from the first surface is maintained in the time profile and hence the time delayed reflections appear as background noise in our reconstruction. Absorbing low-reflectance black materials reduce the SNR but the effect is minor compared to the squared attenuation over distances.

Although near-Lambertian surfaces are very common in the proposed application areas, reconstruction in the presence of varying reflectance materials is an interesting future research topic.

5 Author Contributions

The method was conceived by A. Velten and R. R., the experiments were designed by A. Velten, M. G. B. and R. R. and performed by A. Velten. The algorithm was conceived by A. Velten, T. W., O. G., and R. R. and implemented and optimized by

A. Velten, T. W., O. G., A. Veeraraghavan, and R. R.. All authors took part in writing the paper.

6 Acknowledgements

This work was funded by the Media Lab Consortium Members, DARPA through the DARPA YFA grant, and the Institute for Soldier Nanotechnologies and U.S. Army Research Office under contract W911NF-07-D-0004. A. Veeraraghavan was supported by the NSF IIS Award #1116718.

We would like to thank Amy Fritz for her help with acquiring and testing data and Chinmaya Joshi for his help with the preparation of the figures.

Competing financial interests: The authors declare no competing financial interests.

References

[1] Kamerman, G. W. Laser Radar [M]. Chapter 1 of *Active Electro-Optical System, Vol.* **6***, The Infrared and Electro-Optical System Handbook.*, 1993.

[2] Busck, J. & Heiselberg, H. Gated viewing and high-accuracy three-dimensional laser radar. *Appl. Optics*, **43**(24):4705–4710, 2004.

[3] 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**(5021):769–771, Aug 1991.

[4] Xia H. & Zhang C. Ultrafast ranging lidar based on real-time Fourier transformation. *Opt. Lett.*, **34**:2108–2110, 2009.

[5] Sen, P., Chen, B., Garg, G., Marschner, S. R., Horowitz, M., Levoy, M. & Lensch, H. P. A. Dual photography. In *Proc. of ACM SIGGRAPH*, **24**(2):745–755, 2005.

[6] Repasi, E., Lutzmann P., Steinvall, O., Elmqvist, M., Ghler, B. & Anstett, G. Advanced short-wavelength infrared range-gated imaging for ground applications in monostatic and bistatic configurations. *Appl. Optics*, **48**:5956–5969, 2009.

[7] Sume, A., Gustafsson, M., Jnis, A., Nilsson, S., Rahm, J. & rbom, A. Radar detection of moving objects around corners. In *Proc. of SPIE*, **7308**:73080V–73080V, 2009.

[8] Chakraborty, B., Li, Y., Zhang, J.J., Trueblood, T., Papandreou-Suppappola, A. & Morrell, D. Multipath exploitationwith adaptivewaveform design for tracking in urban terrain. In *IEEE ICASSP Conference*, 3894 – 3897, March 2010.

[9] Raskar, R. & Davis, J. 5d time-light transport matrix: What can we reason about scene properties? *MIT Technical Report*, 2008. URL: http://hdl.handle.net/1721.1/67888

[10] Kirmani, D., Hutchison, T., Davis, J. & Raskar, R. Looking around the corner using transient imaging. In *Proc. of IEEE ICCV*, 159–166, 2009.

[11] Seitz, S. M., Matsushita, Y., & Kutulakos, K. N. A theory of inverse light transport. In *Proc. of IEEE ICCV*, **2**:1440–1447, 2005.

[12] Kak, A.C., & Slaney, M. *Principles of Computerized Tomographic Imaging*, Chapter 3. IEEE Service Center,Piscataway, NJ, 2001.

[13] Blackmon, T. T., et. al. Virtual reality mapping system for chernobyl accident site assessment. In *Proceedings of the SPIE*, **3644**:338–345, 1999.

[14] Burke, J. L., Murphy R. R., Coovert, M. D. & Riddle, D. L. Moonlight in miami: a field study of humanrobot interaction in the context of an urban search and rescue disaster response training exercise. *Hum.-Comput. Interact.*, **19**:85–116, 2004.

[15] Ng, T. C., Ibanez-Guzman, J., Shen, J., Gong, Z., Wang ,H. & Cheng, C. Vehicle following with obstacle avoidance capabilities in natural environments. In *Robotics and Automation, 2004. Proceedings. ICRA '04. 2004 IEEE International Conference on*, **5**:4283 – 4288.

[16] Feld, M. S., Yaqoob, Z., Psaltis, D. & Yang, C. Optical phase conjugation for turbidity suppression in biological samples. *Nature Phot.*, **2**:110–115, 2008.

[17] Dylov, D. V. & Fleischer, J. W. Nonlinear self-filtering of noisy images via dynamical stochastic resonance. *Nature Phot.*, **4**:323–328, 2010.

[18] Popoff, S., Lerosey, G., Fink, M., Boccara, A. C. & Gigan, S. Image transmission through an opaque material. *Nature Comm.*, 2010.

[19] Vellekoop, I. M., Lagendijk, A. & Mosk, A. P.. Exploiting disorder for perfect focusing. *Nature Phot.*, **4**:320–322, 2010.

[20] Choi, Y., et. al. Overcoming the diffraction limit using multiple light scattering in a highly disordered medium. *Phys. Rev. Lett.*, **107**(2):023902, Jul 2011.

[21] Katz, O., Small, E., Bromberg, Y. & Silberberg, Y. Focusing and compression of ultrashort pulses through scattering media. *Nature Phot.*, **5**(6):372–377, Jul 2011.

[22] Nayar, S. K., Krishnan, G., Grossberg, M. D. & Raskar, R. Fast separation of direct and global components of a scene using high frequency illumination. *ACM Trans. Graph.*, **25**(3):935–944, 2006.

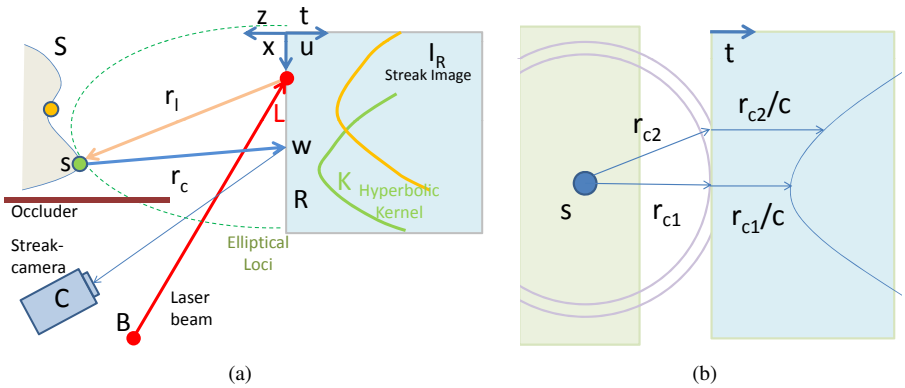
[23] Forsyth, D. & Ponce, J. *Computer Vision, A Modern Approach*. Prentice Hall, 2002.

Recovering Three-dimensional Shape Around a Corner using Ultrafast Time-of-Flight Imaging

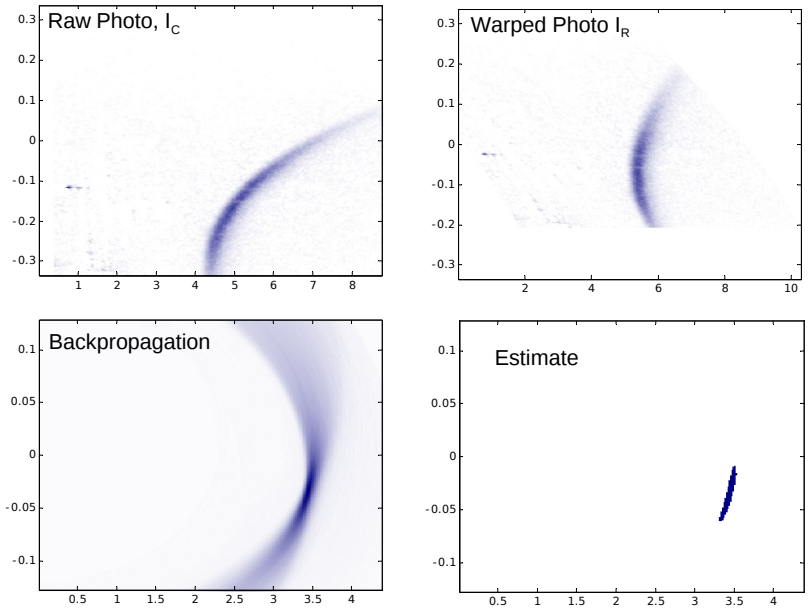
Supplementary Information

Andreas Velten, Thomas Willwacher, Otkrist Gupta,
Ashok Veeraraghavan, Moungi Bawendi, Ramesh Raskar

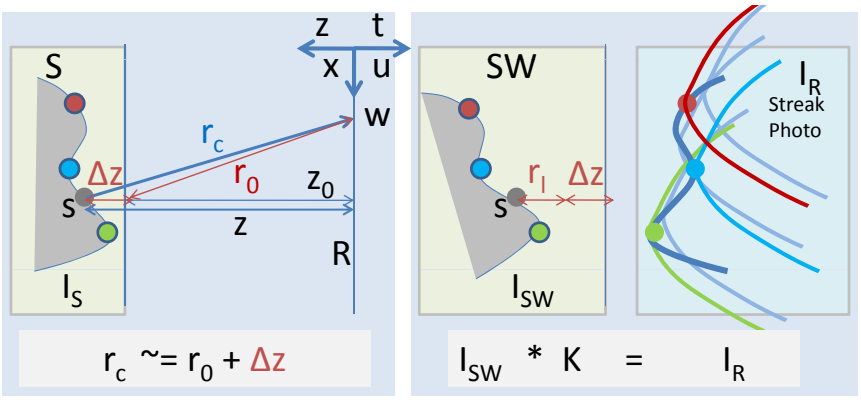
Supplementary Figures



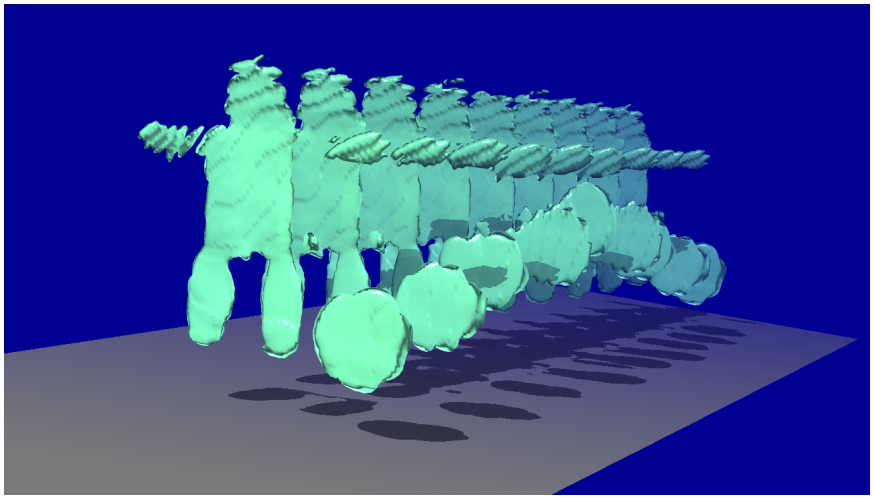
Supplementary Figure S1: Forward Model. (a) The laser illuminates the surface S and each point $s \in S$ generates a wavefront. The spherical wavefront contributes to a hyperbola in the space-time streak image, I_R. (b) Spherical wavefronts propagating from a point create a hyperbolic space-time curve in the streak image.



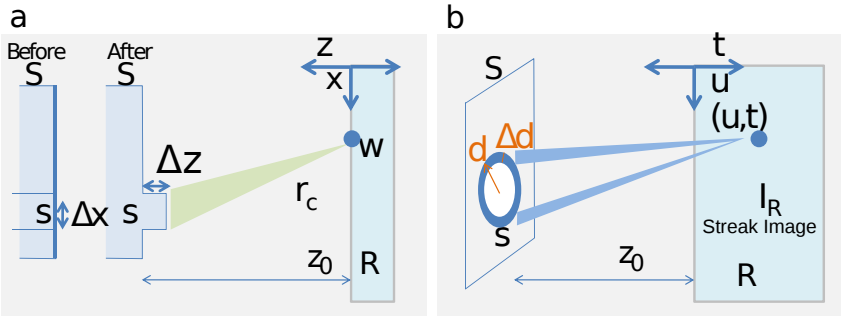
Supplementary Figure S2: Backprojection. A space time transform on a raw streak image allows us to convert a 4 segment problem into a sequence of 2 segment problems. The toy scene is a small 1cm$\times$1cm patch creating a prominent (blurred) hyperbola in the warped image. Backpropagation creates low frequency residual but simple thresholding recovers the patch geometry.



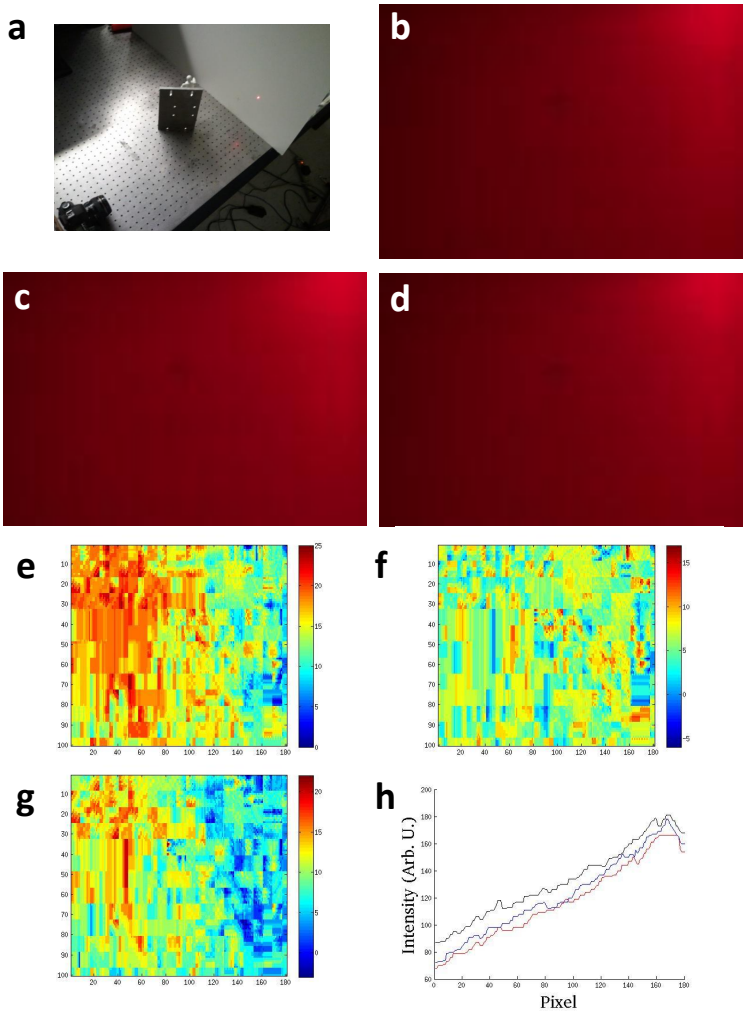
Supplementary Figure S3: Fresnel Approximation for convolution operation. With a near-constant depth assumption of $\Delta z \ll z_0$, the streak image I_R is approximated as a convolution of the *warped shape* countour image I_{SW} with the hyperbolically shaped kernel K. The warped shape image in turn is the true shape (S), deformed along the z direction according to laser distance. We assume an opaque object and hence the contributions are only from the points on the curve (surface) and not from the area behind the curve.



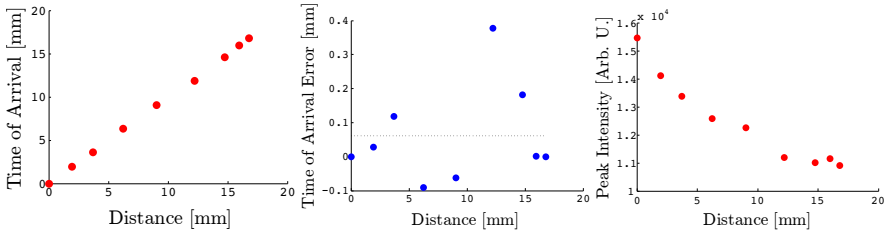
Supplementary Figure S4: Stop motion reconstruction. Results of a multi-pose stop motion animation dataset after filtered backprojection and soft-thresholding. A hidden model of a *man with a ball* is captured in various poses. The rendering shows the sequence of reconstructions created by our filtered backprojection algorithm and demonstrates the ability to remove low-frequency artifacts of backprojection. The mislabeled voxels remain consistent across different poses indicating stability of our capture and inversion process. Shadows are introduced to aid visualization.



Supplementary Figure S5: Improving depth and lateral resolution. (a) In a still camera, the ability to discern displacement of a patch with area $(\Delta x)^2$ by a distance Δz is limited by camera sensitivity. (b) Using time resolution, the ability to discern the same patch is improved and possible within practical camera sensitivity. The pixel (u, t) receives energy only from inside the ring. For simplicity, the diagrams in this document show the scene in flat-land and the surfaces are drawn as 2D curves.

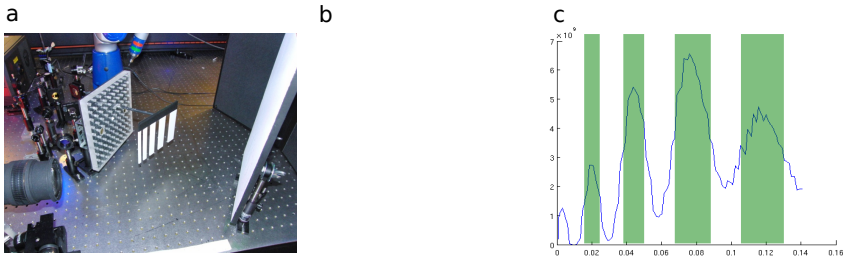


Supplementary Figure S6: Reconstruction attempt with a slow camera. We performed an experiment to demonstrate the challenges in imaging around the corner with a conventional, low temporal resolution laser and camera. (a) A setup with hidden mannequin but using a red continuous laser and a Canon 5D camera. (b) An image of the wall recorded with the Canon 5D camera with the room lights turned off and no hidden object present. (The recorded light is due to the reflections from walls behind the laser and camera.) (c) An image recorded with the hidden mannequin present. The increased light level on the wall is marginal, is low spatial frequency and shows no noticeable high frequency structure. (d) An image of the wall with the hidden mannequin moved away from the wall by 10 cm. The reduction in light level on the wall has no visible structure. (e) The difference between image in (b) and (c) using a false color map. (f) The difference between (b) and (d). (g) The difference between (c) and (d). (h) The plot of intensities along the centered horizontal scanline of each of the images (b=red, c=black, d=blue).



(a) (b) (c)

Supplementary Figure S7: Resolution in depth. (a) Distance estimation. Time here is measured in mm of traveled distance at the speed of light 1 mm≈0.3 ps. (b) Error is less than 1 mm. (c) Plot of intensity as a small patch is moved perpendicular to the first surface.



Supplementary Figure S8: Resolution in lateral dimension. (a) Setup with chirp pattern (occluder removed in this photo) (b) Raw streak photo from streak camera (c) The blue curve shows reconstruction of the geometry and indicates that we can recover features with 0.5 cm in lateral dimensions in the given scenario.

Supplementary Methods

Modelling Light Pulse Propagation

In this section, we analyze the relationship between the hidden scene geometry and the observed space time light transport in order to design methods to estimate the shape of the hidden objects. Consider a scene shown in Supplementary Figure S1. The scene contains a hidden object (whose surface we are interested in estimating) and a diffuser wall. A laser beam(B) emits a short light pulse and is pointed towards the diffuser wall to form a laser spot L. The light reflected by the diffuser wall reaches the hidden surface, is reflected and returns back to the diffuser wall. The streak camera is also pointed towards the wall.

For each location of the laser spot L, a $3D$ image (2 spatial and 1 temporal dimension) is recorded. The laser spot is moved to multiple locations on the wall ($2D$). The two dimensions for laser direction and the three dimensions for recording lead to a $5D$ light transport data. The pulse return time at each location on the wall depends upon several known parameters such as the location of the laser spot and unknown parameters such as the hidden surface profile. The idea is to exploit the observed $5D$ light transport data to infer the hidden surface profile.

For an intuitive understanding, consider the hidden scene to be a single point, as shown in Supplementary Figure S1. The reflected spherical wavefront propagating from that hidden scene point reaches the different points on the wall at different times creating a hyperbolic curve in the space-time streak image (Supplementary Figure S2). When the hidden scene contains a surface instead of individual and isolated scene points, the space-time hyperbolas corresponding to the different surface points are added together to produce the captured streak images and so we need to demultiplex or deconvolve these signals. In general, we could use a captured $5D$ light transport data but in our experiments, we are restricted to a streak camera that has a one spatial dimension. Thus, our system captures only a four dimensional light transport.

Bounce Reduction

In our setup, the optical path for light travel consists of 4 segments (Supplementary Figure S1): (1) from the laser B to the spot on the wall L, (2) from L to the scene point s, (3) from s again to a point on the wall w, and (4) finally from w to the camera C where it is recorded. However, the first and the fourth segment are directed segments and do not involve diffuse scattering. This allows us to precalibrate for these segments and effectively reduce the tertiary scattering problem to a primary (single) scattering problem. More concretely, suppose the camera records the streak image $I_C(p,t)$, where p is the pixel coordinate and t is time. In I_C, $t = 0$ corresponds to the instant the laser pulse is emitted from B. Then I_C is related to the intensity $I_R(w,t)$ of light incident on the receiver plane by the transformation

$$I_R(w,t) = I_C(H(w), t - ||L - B|| - ||C - w||). \tag{S1}$$

Here H is the projective transformation (homography) mapping coordinates on R to camera coordinates. The time shift by the distance from camera to screen, $||C - w||$, varies hyperbolically with the pixel coordinate w. Since the geometry of wall, R, is known, H, $||L-B||$ and $||C-w||$ can be computed in advance. Note there is no $\cos(\theta)$ factor or $1/r^2$ fall off in the above formula as the camera integrates over more pixels for oblique and distant patches. For this to hold, it is also important that R is Lambertian, as we assume. To summarize, the processing step (S1) reduces the problem to a single scattering problem, with an unfocused point source at L emitting a pulse at $t = 0$ and an unfocused virtual array of receivers on R recording the intensity of the reflected wavefront, $I_R(w, t)$.

Scattering of the light pulse

Generating Streak Images After the homography correction, we can consider a simplified scenarios of just two surfaces, the wall R and the hidden surface S. The surface S is illuminated by a light source at L. The surface R (receivers) can be assumed to host a virtual array of ultrafast photodetectors. The virtual photodetectors create an image $I_R(w, t)$ as intensity pattern of the incoming light as a function of time, t, and the position w. Hence the image, $I_R(w, t)$, is the intensity observed at $w \in R$ at time t. Experimentally, the virtual photodetectors are realized by using a Lambertian object R observed by a streak camera with ps time resolution (Supplementary Figure S1). Ignoring occlusions, the intensity pattern at R takes the following approximate form

$$I_R(w,t) = \int_S \int_\tau \frac{1}{\pi r_c^2} \delta(r_c - t + \tau) I_S(s,\tau) d\tau d^2 s \tag{S2}$$

where $w \in R$, $s \in S$, $t, \tau \in \mathbb{R}$ and $r_c = ||w - s||$. Furthermore, $I_S(s, \tau)$ encodes the hidden 3D shape S as the intensity of the light emitted by the transmitter at $s \in S$ at time τ. Note that we use units in which the speed of light $c = 1$. In other words, we measure time in units of distance. Note also that we make an approximation in neglecting the dependence on the normals to surfaces R and S. In the situation of interest to us, the object S is a diffuse (Lambertian) object illuminated by a single point source at position $L \in \mathbb{R}^3$. Concretely, this point source is the surface patch the laser is directed to. Hence, neglecting the normal dependence, $I_S(s, \tau) = I\delta(\tau - r_l)/(\pi r_l^2)$ with $r_l = ||L - s||$. Equation (S2) becomes

$$I_R(w,t) = \int_S I \frac{1}{\pi r_c^2} \frac{1}{\pi r_l^2} \delta(t - r_c - r_l) d^2 s \tag{S3}$$

The propagation of laser to wall and wall to camera is ignored in I_R. Laser to wall propagation is corrected using an offset value for time. The wall to camera sensor propagation is inverted by using a homography. In summary, the recorded streak image, I_C, which involves three or more bounces is converted to image, I_R, which involves only one bounce. For simplicity, we will ignore I_C and consider I_R as the streak image for rest of the discussion.

Hyperbolic Contribution For a fixed laser position, L, and sensor location, w, at a time t, the allowed values of s all lie on an ellipsoid with focal points L and w, given by the equation $t = r_c + r_l$. (More specifically, the locus of s lies on a *prolate spheroid*, i.e., an ellipsoid with two equal equatorial radii, smaller than the third equatorial radius.)

If we fix L and s this equation describes a two sheeted hyperboloid in (w, t)-space:

$$t - r_l = r_c = \sqrt{(x-u)^2 + (y-v)^2 + z(x,y)^2} \tag{S4}$$

where (u, v) are the two coordinates of w in the plane of the receiver wall. In particular, each point on the hidden surface S will contribute a hyperboloid to the image $I_R(u, v, t)$. The hyperboloids will have different shapes, depending on the depth $z(x, y)$, and will be shifted along the t-axis. Smaller depth $z(x, y)$ increases eccentricity and leads to higher curvature at the vertex of the hyperboloid .

Modified Fresnel Approximation Suppose that the hidden surface S has a small depth variation. We can write $z(x, y) = z_0 + \Delta z(x, y)$, with approximate mean depth z_0 and minor variations $\Delta z(x, y)$. Hence, $\Delta z(x, y) \ll z_0$. In this case, we apply an additional approximation, which is the analog of the Fresnel approximation in Fourier optics. Note that we are dealing with incoherent and pulsed light, so we call it the *modified Fresnel approximation*. Concretely, we expand the square root in (S4) and assume that $z_0 \gg (x - u)$ or $(y - v)$. The right hand side of (S4) becomes $r_c = \sqrt{(x-u)^2 + (y-v)^2 + z_0^2} + \Delta z(x, y)$, or $r_c = r_0 + \Delta z(x, y)$. Using this approximation in the argument of the delta function in (S3), and neglecting Δz in the denominator, we can express I_R as a convolution.

$$\begin{aligned} I_R(u, v, t) &\approx \int_{x,y} \frac{\delta(t - r_l - \Delta z - r_0)}{\pi^2 2 r_c^2 r_l^2} dx dy && \text{(S5)} \\ &= \int_{x,y,\tau} \frac{\delta(t - \tau - r_0)\delta(\tau - r_l - \Delta z)}{\pi^2 2 r_c^2 r_l^2} dx dy d\tau \\ &= (K * I_{SW})(u, v, t) && \text{(S6)} \end{aligned}$$

The hidden shape S is expressed using a delta function $I_S = \Delta z(x, y)$. Supplementary Figure S3 shows that, after a transform due to laser position, L, we have a new *warped shape approximation* $I_{SW}(x, y, \tau) = \delta(\tau - r_l - \Delta z(x, y))/(\pi r_l^2)$. We split the delta function inside the intergral above and re-write the equation as a convolution (in 3-dimensional (u, v, t)-space) of the warped shape approximation I_{SW}. This warped image I_{SW} "cramps up" information about the shape S in the time domain, warped by the additional "deformation" r_l, given by the distance to the laser. Finally the convolution kernel $K(x, y, t) = \delta(t - r_k)/(\pi r_k^2)$, with $r_k = \sqrt{x^2 + y^2 + z_0^2}$, is a hyperboloid, whose eccentricity (or curvature at the vertex) depends on z_0.

Note that equation (S6) is highly nonlinear in the unknown depths Δz, but linear in the warped shape I_{SW}, from which these depths can be determined. In conclusion, using

the modified Fresnel approximation, for every depth, we can express the forward propagation as a convolution with a hyperboloid. But for each depth, z_0, the curvature and position of the hyperboloid in space-time streak image, I_R, is progressively different.

Algorithms for surface reconstruction

Problem statement as a system of linear equations Let us express the results of the last section using linear algebra. Let us discretize the bounding box around the hidden shape and the corresponding 3D Cartesian space into voxels and arrange the voxels into a vector $f_S \in \mathbb{R}^N$. Here N is the number of voxels in the 3D volume of interest. The value of $f_S(i)$ is set to zero for all the voxels that do not contain a surface point $s \in S$. The value of $f_S(i)$ for voxels that lie on the surface of the objects is the albedo (or reflectance) of the corresponding surface point. Voxels that are interior to the object are occluded by voxels on the surface of the object and do not return any signal energy, so they are also set to zero. Consider now the streak image I_R recorded with the laser at position L_1. Vectorize the streak image pixels into a single vector $g_{R,1} \in \mathbb{R}^M$, where M is the total number of spatio-temporal pixels present. The pixel values will depend linearly on the albedos in f_S and hence satisfy a linear equation of the form

$$g_{R,1} = A_1 f_S, \tag{S7}$$

for some $M \times N$ matrix A_1. Concretely, the entries of A_1 can be read off from equations (S2) and (S3). If multiple streak images $1, \ldots, n$ are recorded corresponding to different locations of the laser, then those different streak images are stacked on top of each other in a vector y, which satisfies the linear equation

$$g_R = \begin{bmatrix} g_{R,1} \\ g_{R,2} \\ \vdots \\ g_{R,n} \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{bmatrix} f_S = A f_S$$

Our goal is to analyze and solve the above linear system. The Fresnel approximation of the last section gives an intuition. The Fresnel approximation allows us to rewrite the linear system as $g_R = A_F f_{SW}$, where A_F is a block circulant matrix that represents the convolution with the hyperbolic kernel K.

Backprojection for surface reconstruction

Each voxel in the 3D world contributes signal energy to only a very small subset of the spatio-temporal bins (streak camera pixels) that are imaged. The specific spatio-temporal bins or pixels that contain signal energy are related to the hidden 3D surface. Further, if a particular pixel in the image contains no signal energy, this means that every voxel in the 3D world that would have contributed energy to it was empty. Both these pieces of information can be used to contruct an algorithm for reconstruction. The

basic intuition behind this algorithm (a backprojection algorithm, similar to algorithms used tomographic reconstruction) is very simple. Each observed pixel contributes energy back to all the source voxels that could have contributed energy to this pixel and the contribution made is proportional to the observed signal energy. Voxels in the world that were occupied receive contributions from all the pixels in the streak image that they contributed to and therefore have a large response. This response energy with appropriate filtering and thresholding can be used to recover the 3D surface. If the working volume is shallow (i.e., can be represented with a plane at depth z_0 plus minor depth variations), we can use a single kernel of Frensel approximation leading to a block circulant matrix A. Below, we first explore reconstruction with Frensel approximation. Then we strive for a more accurate result. We use a different kernel for each depth, leading to a non-block circulant matrix.

Depth Independent Fresnel Approximation In order to implement the backpropagation algorithm, in practice, it is necessary to model the forward propagation of the spherical wavefronts from each of the hidden surface voxels. Although approximate, we first use the Fresnel approximation based forward propagation model described in the Section "Scattering of the light pulse" for a better understanding and ability to easily analyze the invertibility. Under the Fresnel approximation the captured streak images can be written as a convolution of the unknown surface with the hyperbolic blur kernel as given by

$$I_R = K * I_{SW}$$

with $K(x, y, t) = \delta(t - \sqrt{x^2 + y^2 + z_0^2})/(\pi(x^2 + y^2 + z_0^2))$. The backprojection kernel, on the other hand, is $\tilde{K}(x, t) = \delta(t + \sqrt{x^2 + y^2 + z_0^2})$. Hence backprojection is, up to the distance attenuation prefactor, the adjoint of propagation. If the Fresnel approximation is valid, the effect of backprojection on the captured streak images can be described as a convolution with the point spread function $psf = \tilde{K} * K$. The function psf can be computed analytically. This function has a large peak at the center, surrounded by a butterfly shaped low frequency component. This peak implies that when one performs backprojection peaks will be observed at all the locations where there is a 3D scene point. The limitations of backprojection are also evident from the function psf. Since the peak is surrounded by low frequency components, this approach without any post-processing (filtering) will lead to overly smoothened results. Rephrasing these observations in the matrix notation introduced at the beginning of this section, one can say that the backprojection operation is described by the matrix $\tilde{A}_F$, which is the same as A_F^T,up to the distance attenuation factors. The composition of propagation and backprojection $\tilde{A}_F A_F$ is close to the identity matrix.

Depth Dependent backprojection While it is very useful to use the Fresnel approximation to analyse the effect of backprojection, the approximations made lead to inaccuracies when (a) the scene has significant depth variation or (b) there are occlusions. In those cases we need to use the more precise formulas (S2), (S3). Propagation can then no longer be written as a convolution, since the integral kernel, i.e., the hyperbola, changes shape with varying depth.

Limitations of Backprojection Backprojection suffers from several limitations. The results of backprojecton are smoother than the original surface and there are still a few false positive surface points. We also observe that surface slopes beyond 45 degrees are extremely hard to reconstruct. This can be explained theoretically, at least in the Fresnel approximation as follows. The Fourier transform of the hyperbolic convolution kernel falls off gently (by a power law) in the direction orthogonal to the receiver plane, but falls off exponentially in the parallel direction. Hence features having high parallel spatial frequencies, such as high slope regions, are very hard to recover. In order to tackle these limitations it would be necessary to use additional prior information about the 3D scene being reconstructed.

We use the traditional approach that removes the low frequency artifacts around sharp features. The method is known as filtered backprojection that involves using a carefully chosen high pass filter. In our case, a good choice is second derivative in the z direction of the voxel grid. Supplementary Figure S4 shows the results after such filtering and applying a soft threshold described in the main paper.

Note that backprojection is a voxel-based technique and does not take into account the surface-based properties like orientation or reflectance profile. Hence our technique is expected to work best for nearly Lambertian surfaces, for which the recorded images do not depend strongly on the surface normals.

Necessity of ultrafast imaging

We consider the achievable resolution and space-time dimension tradeoffs in hidden shape recovery.

Limits of Traditional Photography

Even with a still camera, one can in principle detect the displacement of a small hidden area of a scene as shown in Supplementary Figure S5(a), but the problem is ill-conditioned. To see this, let us consider for simplicity a near planar scene at depth z_0, illuminated homogeneously by a far away source. The intensity of light incident at a point $r \in R$ that was emitted by a surface patch above r of size Δx by Δx is proportional to $I(\Delta x)^2/(z_0^2)$, where I is the total intensity received. Moving the patch by $\Delta z \ll z_0$ in depth, the contributed intensity will change by $\Delta I \propto I(\Delta x)^2 \Delta z / z_0^3$. Hence we conclude that $\Delta I/I \propto (\Delta x)^2 \Delta z / z_0^3$. As in typical scenario, the spatial resolutions ($\Delta x, \Delta z \approx 5$mm, $z_0 \approx 20$cm, we require intensity resolution, $\Delta I/I \sim 3 \times 10^{-5}$. This means one has to distinguish intensities of 1 from 1.00003. This is not possible in practice. Note that the intensities received after tertiary scattering are already very small, so it is hard to obtain a good signal to noise ratio. We show the limits of traditional low temporal resolution photography via an example in Supplementary Figure S6.

Benefits of Time Resolution

For an ordinary camera, two conditions make the problem ill-conditioned: The relative intensity contributed by an emitter changes only slightly ($\propto \Delta z/z_0$) and this small change is overwhelmed by the contribution of the background with area A, yielding the factor $(\Delta x/z_0)^2$. Using a ps accurate high-speed light source and sensors these problems can be circumvented.

A change in patch position s means it contributes to a different pixel in the streak photo, provided $\Delta z/c > \Delta t$, where $c =$ speed of light and Δt is time resolution.

Unlike an ordinary sensor, not all patches on S contribute to a pixel (time bin) in a streak photo making the mixing easier to invert. The locus of points contributing to a fixed sensor and time-bin position, (u, t), lie on a ring with radius $d = \sqrt{(ct)^2 - z_0^2}$ (Supplementary Figure S5(b)). If the time bin has a width of $\Delta t \ll t$, the width of the ring is approximately $\Delta d = c^2 t \Delta t/d$. Hence the total area of the ring is $2\pi d\Delta d = 2\pi c^2 t \Delta t$. We want to detect changes in the intensity emitted by a patch of size $\Delta A = (\Delta x)^2$. Hence the change in total intensity is approximately $\Delta I/I = \Delta A/(2\pi d\Delta d) = (\Delta x)^2/(2\pi c^2 t \Delta t)$. In our scenario typically $\Delta x \approx 3c\Delta t$. Furthermore $ct \approx z_0$. Hence $\Delta I/I \approx 3\Delta x/(2\pi z_0)$. Thus the required intensity increment is linearly proportional to $\Delta x/z_0$, and not quadratically as before. In our case, this ratio is a reasonable $\sim 10^{-2}$. This gives the guidance on time-resolution. In addition, the time resolution of the light source should not be worse than that of the sensor.

Performance Validation

We performed a series of tests to estimate the spatial resolution perpendicular and parallel to the visible surface, i.e., the wall. We use the FARO Gauge measurement arm to collect independently verifyable geometric position data (ground truth) and compared with positions recovered after multiple scattering using our algorithm. In our system, translation along the direction perpendicular to the diffuser wall can be resolved with a resolution of 400 μm better than the full width half maximum (FWHM) time resolution of the imaging system (Supplementary Figure S7, a and b). Lateral resolution in a plane parallel to the wall is lower and is limited to 0.5–1 cm depending on proximity to the wall (Supplementary Figure S8).

Choice of resolutions

There are four important parameters for our shape estimation setup: The spatial resolution, i.e., the spacing between sensor pixels, the temporal resolution of sensors and the laser, the intensity resolution and signal to noise ratio of our sensors (and power of the light source) and the angular diversity determined by the geometry of the setup. We saw in Section "Benefits of Time Resolution" that time resolution is critical and gives us an approximate lower bound on the resolution of our 3D reconstruction This is the same as in the case of a direct view traditional time of flight camera. However,

our situation differs from a direct view time of flight camera. Neither our sensors nor our light source have any directional resolution into the hidden space after a diffuse reflection.

Spatial Camera Resolution If we could determine the correspondences, i.e., which part of the received signal at a sensor was emitted by which transmitter (surface patch), spatial resolution of a streak camera would actually be unnecessary. Time resolution will directly determine reconstructed 3D resolution of hidden objects. Despite these two challenges, finite time resolution of the streak camera and the loss of correspondence, the sufficiently high spatial resolution allows us to exploit the local structure in streak photo to recover shape without explicitly solving the correspondence problem.