

Looking Around the Corner using Transient Imaging

Ahmed Kirmani^{*1}, Tyler Hutchison¹, James Davis^{†2}, and Ramesh Raskar^{‡1}

¹MIT Media Laboratory

²UC Santa Cruz

Abstract

We show that multi-path analysis using images from a time-of-flight (ToF) camera provides a tantalizing opportunity to infer about 3D geometry of not only visible but hidden parts of a scene. We provide a novel framework for reconstructing scene geometry from a single viewpoint using a camera that captures a 3D time-image $I(x, y, t)$ for each pixel. We propose a framework that uses the time-image and transient reasoning to expose scene properties that may be beyond the reach of traditional computer vision. We corroborate our theory with free space hardware experiments using a femtosecond laser and an ultrafast photo detector array. The ability to compute the geometry of hidden elements, unobservable by both the camera and illumination source, will create a range of new computer vision opportunities.

1. Introduction

Camera-based 2D intensity images, $I(x, y)$, have long been used to observe and interpret scenes. New sensors and algorithms for scene understanding will clearly benefit many application areas such as robotics, industrial applications, user interfaces and surveillance. This paper introduces a novel framework called *Transient Light Transport* which allows new properties of scenes to be observed and interpreted. In a traditional camera, the light incident at a pixel is integrated along angular, temporal and wavelength dimensions during the exposure time to record a single intensity value. Distinct scenes may result in identical projections (images) and, hence, identical pixel values. Thus, it is challenging to estimate scene properties which are not directly observable. Steady-state light transport assumes an equilibrium in global illumination. In a room-sized environment, a microsecond exposure (integration) time is long enough for a light impulse to fully traverse all the possible multi-paths introduced due to inter-reflections between scene elements and reach steady state. Traditional video cameras sample light very slowly compared to the time scale at which the transient properties of light come into play.

^{*} {akirmani|thutch|raskar} at mit.edu

[†] davis at cs.ucsc.edu. Work supported in part by NSF #CCF-0746690

[‡] Supported by Nokia Research and Alfred P. Sloan research fellowship

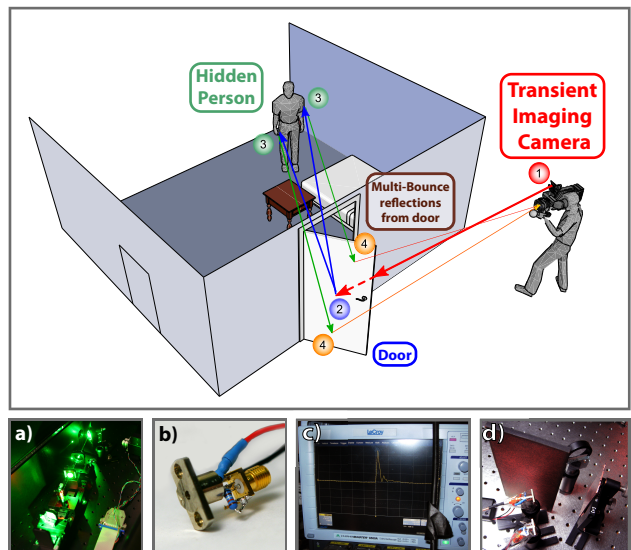


Figure 1. Can you look around the corner into a room with no imaging device in the line of sight? This paper shows that by emitting short pulses (1-2), and analyzing multi-bounce reflection from the door (4-1), we can infer hidden geometry even if the intermediate bounces (3) are not visible. The transient imaging camera prototype consists of (a) Femtosecond laser illumination (b) Picosecond-accurate detectors and (c) an Ultrafast sampling oscilloscope. We measure the Space Time Impulse Response (STIR) of the scene (d) containing a hidden 1-0-1 barcode and reconstruct the hidden surface. Please refer to supplementary video 8 for more results.

In our transient light transport framework, light takes a finite amount of time to travel from one scene point to the other. Recent advances in ultra-high speed imaging have made it possible to sample light as it travels 0.3 millimeter in 1 picosecond. The dynamics of transient light transport in response to a single ray impulse illumination can be extremely complex, even for a simple scene. Unlike a traditional 2D pixel, which measures the total number of photons, transient light transport measures photon arrival rate as a function of time. In Section 3, we propose a transient imaging camera model which samples incident light continuously. In Section 4, we propose the theoretical framework called *inverse* transient light transport for estimating the geometry of scenes that may contain elements occluded from both the camera and illumination. Section 5 describes our hardware prototype comprising a femtosec-

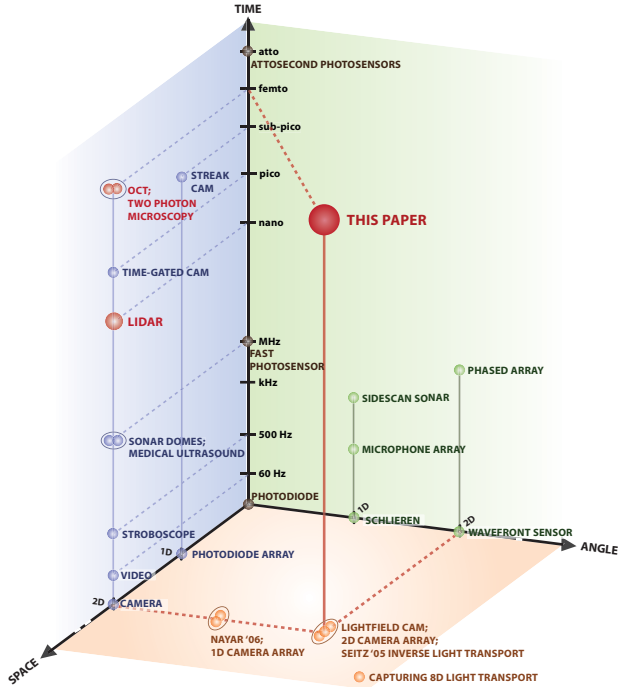


Figure 2. Popular imaging methods plotted in the Space-Angle-Time axes. With higher dimensional light capture, we expand the horizons of scene understanding. Our paper uses LIDAR-like imaging hardware, but, in contrast, we exploit the multi-path information which is rejected in both LIDAR and OCT.

ond laser and a directionally sensitive, picosecond accurate photo sensor array. Using our prototype, we demonstrate all the key functionalities required in a transient imaging camera: geometry, photometry, multi-bounce observations and free-space functioning. Finally, in Section 6, we experimentally demonstrate how our proposed imaging model enables novel scene understanding which allows us to look around the corner without any device in the line of sight (Figure 1).

2. Related work

Global light transport: Light often follows a complex path between the emitter and sensor. A description of steady-state light transport in a scene is referred to as the *rendering equation* [1]. Extensions have been described to include time in light transport [2]. In [3], Raskar and Davis proposed inverse analysis using a 5D time-light transport matrix to recover geometric and photometric scene parameters. In addition, Smith et. al. [4] proposed a modification of the rendering equation via a transient rendering framework. Accurate measurement of physical scene properties is called inverse-rendering [5]. Complex models have been developed for reconstructing specular [6], transparent [7], Lambertian [8] scenes and joint lighting and reflectance [9]. Recent work in image-based modeling and computational photography has also shown several methods for capturing steady-state light transport [10]. Our work

is highly influenced by the following pioneering efforts in steady-state global light transport analysis. Nayar et. al. decomposed an image into its direct and indirect components under the assumption that the scene has no high-frequency components [11]. Seitz et. al. [12] have decomposed images into multi-bounce components under the assumption that the scene is Lambertian. Although the dual photography approach [10] can see an object hidden from a camera, it requires a projector in the object’s line of sight. Our method exploits transient, rather than steady-state transport, to estimate more challenging scene properties.

LIDAR and Time-gated imaging: LIght Detection And Ranging systems modulate light, typically on the order of nanoseconds, and measure the phase of the reflected signal to determine depth [13]. Flash LIDAR systems use a 2D imager to provide fast measurement of full depth maps [14, 15]. Importantly, a number of companies (Canesta, MESA, 3DV, PMD) are pushing this technology towards consumer price points. The quality of phase estimation can be improved by simulating the expected shape of the reflected signal or estimating the effect of ambient light [16]. Separately detecting multiple peaks in the sensor response can allow two surfaces, such as a forest canopy and a ground plane, to be detected, and waveform analysis can detect surface discontinuities [17]. Time-gated imaging captures $I(x, y, t_\delta)$ by integrating the reflected pulse of light over extremely short windows. Multiple captures at incremental time windows, t_δ , allow the time image $I(x, y, t)$ to be captured at up to 100 picosecond accuracy. Nanosecond windows are used for imaging tanks at the range of kilometers and picosecond gating allows imaging in turbid water.

Femtosecond Imaging: Optical coherence tomography (OCT) [18], an interferometric technique, and two-photon microscopy [19], using fluorescence, allow high-quality, micrometer-resolution 3D imaging of biological tissue. Both these methods are based on pulsed femtosecond illumination. Experiments in this paper are the first attempt at free-space use of femto-laser illumination in contrast to their established use in optical fibers or millimeter-scale biological samples. Streak cameras are ultrafast photonic recorders which deposit photons across a spatial dimension, rather than integrating them in a single pixel. Picosecond streak cameras have been available for decades [20]. Modern research systems can function in the attosecond range [21]. All the existing methods based on time sampling of light make no use of global light transport reasoning to infer scene characteristics. They instead image in a single direction time-gated window to improve SNR and reject multi-path scattering. This paper shows that complex global reasoning about scene content is possible given a measured multi-path time profile.

3. Transient Light Transport

The theory of light transport describes the interaction of light with a scene. Incident illumination causes the first bounce. This direct bounce is followed by a complex pattern of inter-reflections whose dynamics are governed by the scene geometry and material properties of the scene elements. At any time instant we only observe a projection of the complete set of scene light rays. This projection only comprises the rays that are directed towards the camera.

We consider a scene \mathcal{S} composed of M small planar facets p_1, \dots, p_M with 3D positions $z_i \in \mathbb{R}^3$. Let $\mathbf{Z} = [z_1, \dots, z_M]$. Define relative visibility between patches, $v_{ij} = v_{ji} = 0$ or 1 depending on whether or not patch p_i is occluded from p_j . Let $\mathbf{D} = [d_{ij}]$ be the Euclidean distance matrix containing pairwise distances. For analytical convenience, we consider the camera (observer) and illumination (source) as a single patch denoted by p_0 . It is straightforward to extend the following analysis to include multiple sources and the camera at an arbitrary position in the scene. We assume that the scene is static and material properties are constant over sub-nanosecond imaging intervals.

Transient Imaging Camera: Our model comprises a generalized sensor and a pulsed illumination source. Each sensor pixel observes a unique patch in the scene. It also continuously time samples the incoming irradiance, creating a 3D *time image*, $I(x_i, y_i, t)$. The pixel at sensor position (x_i, y_i) observes the patch p_i over time. The pulsed illumination source generates arbitrarily short duration and directional impulse rays. The direction of an impulse ray aimed at patch p_i is specified by (θ_i, ϕ_i) . The sensor and illumination are synchronized for precise measurement of Time Difference Of Arrival (TDOA).

Space Time Impulse Response of the scene \mathcal{S} denoted by $\text{STIR}(\mathcal{S})$ is a collection of time images, each captured with an impulse ray illuminating a single scene patch p_j . This is a 5D function: $\text{STIR}(x_i, y_i, \theta_j, \phi_j, t)$. We measure STIR using the following (Figure 3):

- 1 For each patch $p_j : j = 1, \dots, M$.
 - 1a Illuminate p_j with an impulse ray (θ_j, ϕ_j) .
 - 1b Record time image $\{I(x_i, y_i, t) : i = 1 \dots M; t = 0 \dots T\} = \text{STIR}(x_i, y_i, \theta_j, \phi_j, t)$.

In still scenes with static material properties, light transport is a Linear and Time Invariant (LTI) process. Hence, the STIR of the scene characterizes its appearance under any general illumination from a given camera view. In this paper, we use the STIR for developing novel algorithms and models for scene understanding, particularly for *seeing* hidden scene patches.

4. Inverse Geometry

Unlike traditional time-of-flight imaging, our goal is to compute the direct distances, d_{0i} , using the first bounce, and the pairwise distances, d_{ij} . Instead of using intrinsic camera

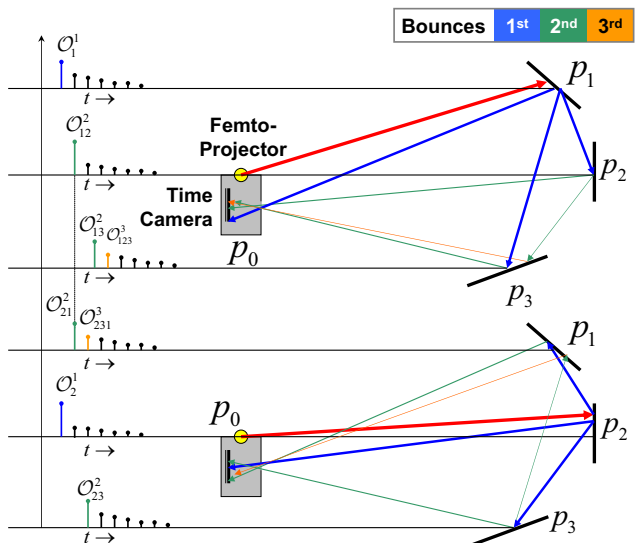


Figure 3. Measuring the STIR of a scene with 3 patches using the transient imaging camera. We successively illuminate a single patch and record a 3D time image. Collection of such time images creates a 5D STIR.

calibration, we exploit second and higher order bounces to estimate scene geometry. First, we use the onset information contained in the STIR to estimate pairwise distances. Then we compute a robust isometric embedding to determine patch coordinates. We develop our formulation for a scene with the following strict assumptions:

- 1 Each patch is visible from all the other patches ($v_{ij} = 1, \forall i, j$). If not, then we image locally with a set of patches for which this is true.
- 2 The reflectance of each patch p_i has a non-zero diffuse component. This assumption ensures that we are able to estimate direct distances d_{0i} .

In Section 4.3, we discuss the extension of our transient imaging framework to scenes consisting of patches hidden from the camera and illumination.

4.1. Distances from STIR

Define $\mathcal{O}^1 = \{O_i^1 | i = 1, \dots, M\}$ as the set of *first onsets*: the collection of all time instants, O_i^1 , when the pixel observing patch p_i receives the first non-zero response while the source illuminates the same patch p_i (Figure 3). O_i^1 is the time taken by the impulse ray originating at p_0 directed towards p_i to arrive back at p_0 after the first bounce; this corresponds to the direct path $p_0 \rightarrow p_i \rightarrow p_0$. Similarly, we define $\mathcal{O}^2 = \{O_{ij}^2 | i, j = 1, \dots, M; j \neq i\}$ as the set of *second onsets*: the collection of times when the transient imaging camera receives the first non-zero response from a patch p_i while illuminating a different patch p_j (Figure 3). This corresponds to the multi-path $p_0 \rightarrow p_j \rightarrow p_i \rightarrow p_0$. $O_{ij}^2 = O_{ji}^2$. It is straightforward to label the onsets in \mathcal{O}^1 and \mathcal{O}^2 because they correspond to the first non-zero responses in STIR time images.

In order to compute \mathbf{D} using \mathcal{O}^1 and \mathcal{O}^2 , we construct the forward distance transform, \mathbf{T}_2 , of size

$(M(M + 1)/2 \times M(M + 1)/2)$ which models the sum of appropriate combinations of path lengths contained in the distance vector $\mathbf{d} = \text{vec}(\mathbf{D})$ and relates it to the vector of observed onsets \mathbf{O} . Then we solve the linear system $\mathbf{T}_2 \mathbf{d} = \mathbf{O}$ to obtain distance estimates $\hat{\mathbf{d}}$. As an example, consider a scene with 3 patches ($M = 3$) as shown in Figure 3. The linear system for this scene is constructed as:

$$\begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ \hline 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} d_{01} \\ d_{12} \\ d_{13} \\ d_{02} \\ d_{23} \\ d_{03} \end{bmatrix} = c \begin{bmatrix} O_1^1 \\ O_{12}^2 \\ O_{13}^2 \\ O_1^1 \\ O_{23}^2 \\ O_1^3 \end{bmatrix}$$

For any M , matrix \mathbf{T}_2 is full rank and well-conditioned. Due to synchronization errors, device delays and response times the observed onsets have measurement uncertainties which introduce errors in distance estimates. We use the redundancy in second onset values ($O_{ij}^2 = O_{ji}^2$) to obtain multiple estimates, $\hat{\mathbf{d}}$, and reduce error by averaging them. A detailed noise analysis can be found in the supporting technical report 8.

4.2. Structure from Pairwise Distances

The problem of estimating scene structure, \mathbf{Z} , from pairwise distance estimates, \mathbf{D} , is equivalent to finding an *isometric embedding* $\hat{\mathbf{Z}} \subset \mathbb{R}^{M \times 3} \rightarrow \mathbb{R}^3$ (Algorithm 1, [22]). For computational convenience we take p_0 to be the origin ($z_0 = (0, 0, 0)$). A computer simulation that recovers the scene structure from noisy distance estimates using the isometric embedding algorithm is shown in Figure 4. We used the estimated coordinates, $\hat{\mathbf{Z}}$, iteratively to recompute robust distance estimates. The use of convex optimization to compute optimal embeddings in the presence of distance uncertainties is explained in [22].

Algorithm 1 ISOEMBED [$\hat{\mathbf{D}}$]

1. Compute $h_{ij} = \frac{1}{2} (d_{0i}^2 + d_{0j}^2 - d_{ij}^2)$. Construct Gram matrix $\mathbf{H}_{M \times M} = [h_{ij}]$
 2. Compute the SVD of $\mathbf{H} = U \Sigma V^T$
 3. Pick 3 largest eigenvalue-vectors $\Sigma_3^{3 \times 3}, U_3^{M \times 3}, V_3^{3 \times M}$
 4. Compute embedding $Z_e = (\Sigma_3)^{1/2} V_3$
 5. Rotate and translate to align $\hat{\mathbf{Z}} = \mathbf{R} Z_e + \mathbf{T}$
-

4.3. Scenes with Occluders

We now consider a scene that contains a set of patches (say H) hidden from both the camera and the source. Hidden surface estimation is viewed as two sub-problems: (1) Labeling third onsets and (2) Inferring distances to hidden patches from integrated path lengths. To estimate the structures of the hidden patches, we make the following three strong assumptions:

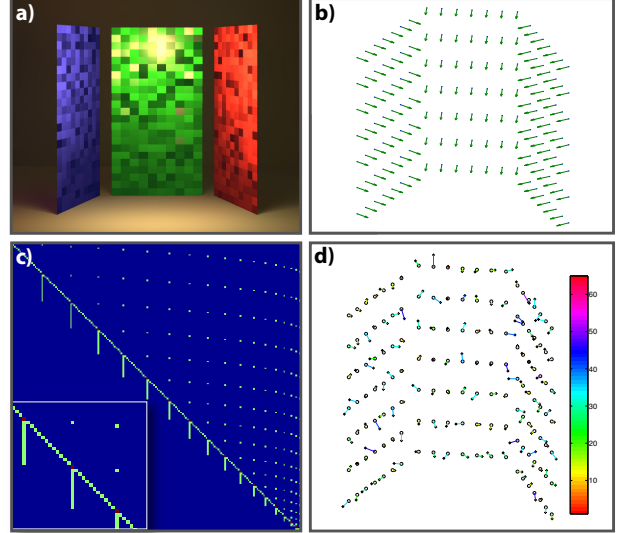


Figure 4. (a) Estimating distances in an all-visible scene comprising of 3 rectangles which are discretized as 49 patches. Note that reflectance is not relevant. (b) Original geometry shows the surface normals in green. (c) We used noisy 1st and 2nd time onsets (Gaussian noise $\sim \mathcal{N}(\mu, \sigma^2)$, $\mu = \text{device resolution} = 250\text{ps}$ and $\sigma = 0.1$) to estimate the distances using the \mathbf{T}_2 operator (inset shows enlarged view). (d) This is followed by isometric embedding and surface fitting. The reconstruction errors are plotted. Color bar shows %-error in reconstructed coordinate values.

- 1 The number of hidden patches is known or assumed.
- 2 All third bounces arrive before fourth and higher order bounces.
- 3 No two or more distinct third bounces arrive at the same time in the same time profile $\text{STIR}(x_i, y_i, \theta_j, \phi_j, t = 0 \dots T)$.

The second assumption is true for scenes that have no inter-reflection amongst hidden patches. The third assumption is generally valid because we measure the STIR one patch at a time. If a patch, p_i , is hidden from p_0 , then the first and second onsets involving p_i cannot be observed, i.e the vector of distances $\mathbf{d}_H = [d_{ij}] : p_i \in H, j = 0, \dots, M$ cannot be estimated using just \mathcal{O}^1 and \mathcal{O}^2 . Hence, we need to consider the set of *third onsets*, $\mathcal{O}^3 = \{O_{ijk}^3 : i, j, k = 1, \dots, M; i \neq j; j \neq k\}$, that corresponds to third bounces. Note that there are $O(M)$ first onsets, $O(M^2)$ second onsets and $O(M^3)$ third onsets¹. Also, Euclidean geometry imposes that $O_{ijk}^3 = O_{kji}^3$. Labeling the onsets contained in \mathcal{O}^3 is non-trivial. As a simple example, consider the scene in Figure 5. Assume that the patches p_2 and p_3 are hidden. We first compute the distances involving visible patches, d_{01}, d_{04}, d_{14} as explained in Section 4.1. The distances (d_{21}, d_{24}) and (d_{31}, d_{34}) are not directly observable. Once these distances are estimated,

¹An important generalization of the hidden patches scenario is to estimate distances in the case of multiple interactions between hidden patches. If a hidden patch has at most N inter-reflections with the other hidden patches, then we need to utilize onsets that correspond to up to $(N + 3)$ bounces i.e. the sets $\mathcal{O}^1, \mathcal{O}^2, \dots, \mathcal{O}^{N+3}$

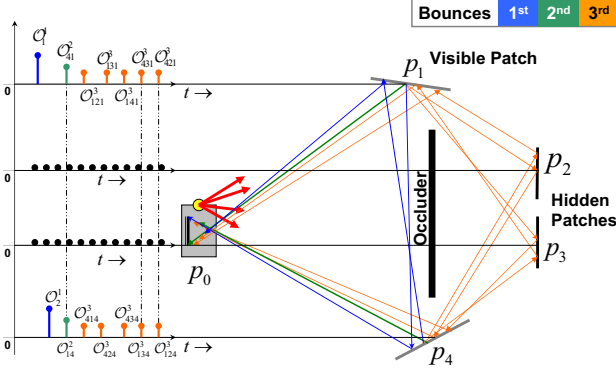


Figure 5. A scene with $M = 4$ patches. Patches p_2 and p_3 are hidden. The blue (first) and green (second) onsets are a result of directly observing visible patches p_1 and p_4 . The pattern of arrival of third onsets depends on the relative distance of the hidden patches p_2 and p_3 from the visible patches. The onsets that correspond to light traversing the same Euclidean distance are readily identified. Once the onsets are labeled, they are used to obtain distances that involve hidden patches.

d_{02}, d_{03} and d_{23} can be computed using multilateration. Now, we apply our labeling algorithm to identify third onsets. The onsets, O_{141}^3 and O_{414}^3 , are readily labeled using TDOA, since we know the distances to patch p_1 and p_4 . The onsets $O_{121}^3, O_{131}^3, O_{424}^3, O_{434}^3, O_{124}^3, O_{134}^3, O_{421}^3, O_{431}^3$ are disambiguated using the facts that $O_{421}^3 = O_{124}^3, O_{431}^3 = O_{134}^3$ and the onsets arrive in different time profiles of the $\text{STIR}(\mathcal{S})$. We sort the remaining onsets based on their arrival times and label them based on the *a priori* assumption of the proximity of hidden patches to visible patches. In this example, w.l.o.g we assume that p_2 is closer to p_1 than p_3 . Hence, the onset O_{121}^3 arrives earlier than O_{131}^3 (see onset arrival profile in Figure 5). This labeling procedure can be generalized for multiple hidden patches:

- 1 Estimate the distances to all the visible scene patches (Section 4.1) and use the arrival times to label all third bounce onsets corresponding to visible geometry.
- 2 Fix an arbitrary ordering of hidden patches based on their proximity to some visible patch.
- 3 Use arrival times to identify the third onset pairs corresponding to same path length ($O_{ijk}^3 = O_{kji}^3$). Label them with the ordering of step 2.
- 4 Sort the remaining onsets according to their arrival times and use step 2 ordering to label them.

We construct the distance operator, \mathbf{T}_3 , that relates third bounces arrival times involving hidden patches, \mathbf{O}_H , and the distances to the hidden patches, \mathbf{d}_H . We solve the resulting linear system $\mathbf{T}_3 \mathbf{d}_H = \mathbf{O}_H$ and obtain the complete distance set, \mathbf{D} . We then estimate the structure, \mathbf{Z} , as discussed in Section 4.2. An example of reconstructing hidden 3D geometry is shown in Figure 6.

$$\begin{bmatrix} 2 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} d_{21} \\ d_{24} \\ d_{31} \\ d_{34} \end{bmatrix} = c \begin{bmatrix} O_{121}^3 - O_1^1 \\ O_{124}^3 - (O_1^1 + O_4^1)/2 \\ O_{131}^3 - O_3^1 \\ O_{134}^3 - (O_1^1 + O_4^1)/2 \end{bmatrix}$$

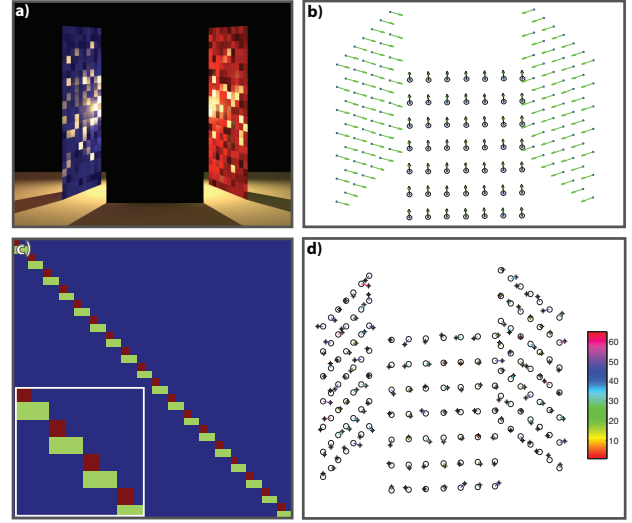


Figure 6. (a) Estimating distances in scenes with hidden patches. Unknown to the estimation algorithm, the hidden patches are on a plane (shown in black). (b) Original patch geometry. We use 1st, 2nd and 3rd bounce onsets, our labeling algorithm and the \mathbf{T}_3 operator (c) to estimate hidden geometry. (d) The isometric embedding error plot verifies negligible reconstruction error and near co-planarity of patches. Onset noise and color bar schemes are same as Figure 4.

5. Hardware Prototype

We corroborate the framework developed in Section 4 with experiments conducted using a prototype transient imaging camera. Our experiments demonstrate feasibility but not a full-fledged imaging apparatus. In particular, we intend this prototype (Figure 1) to show that it is feasible to reason about multi-bounce global transport using the STIR.

We used a commercially-available reverse-biased silicon photo sensor (Thorlabs FDS02, \$72). This sensor has an active area of 250 microns in diameter and a condensing lens to gather more light. Photo-currents were digitized by a 5 GHz oscilloscope. Our least count was 50 ps (1.5cm light travel). Our ray impulse source was a modelocked Ti-Sapphire laser with a center wavelength of 810 nm, that emitted 50 femtosecond long pulses at a repetition rate of 93.68 MHz. The spatial bandwidth of these pulses greatly exceeds the response bandwidth of the sensor. Average laser power was 420 milliwatts, corresponding to a peak power of greater than 85 kW.

We need to sample the incident light with picosecond resolution and be highly sensitive to a low photon arrival rate. Our depth resolution is limited by the response time of the detector and digitizer (250 ps, 7.5cm light travel). The high peak power of our laser was critical for registering SNR above the dark current of our photo sensor. Also, our STIR acquisition times are in nanoseconds, which allows us to take a large number of *exposures* and time average them to reduce Gaussian noise. In absence of a 2D photo sensor array, we emulated directionality by raster scanning the scene with a steerable laser and sensor.

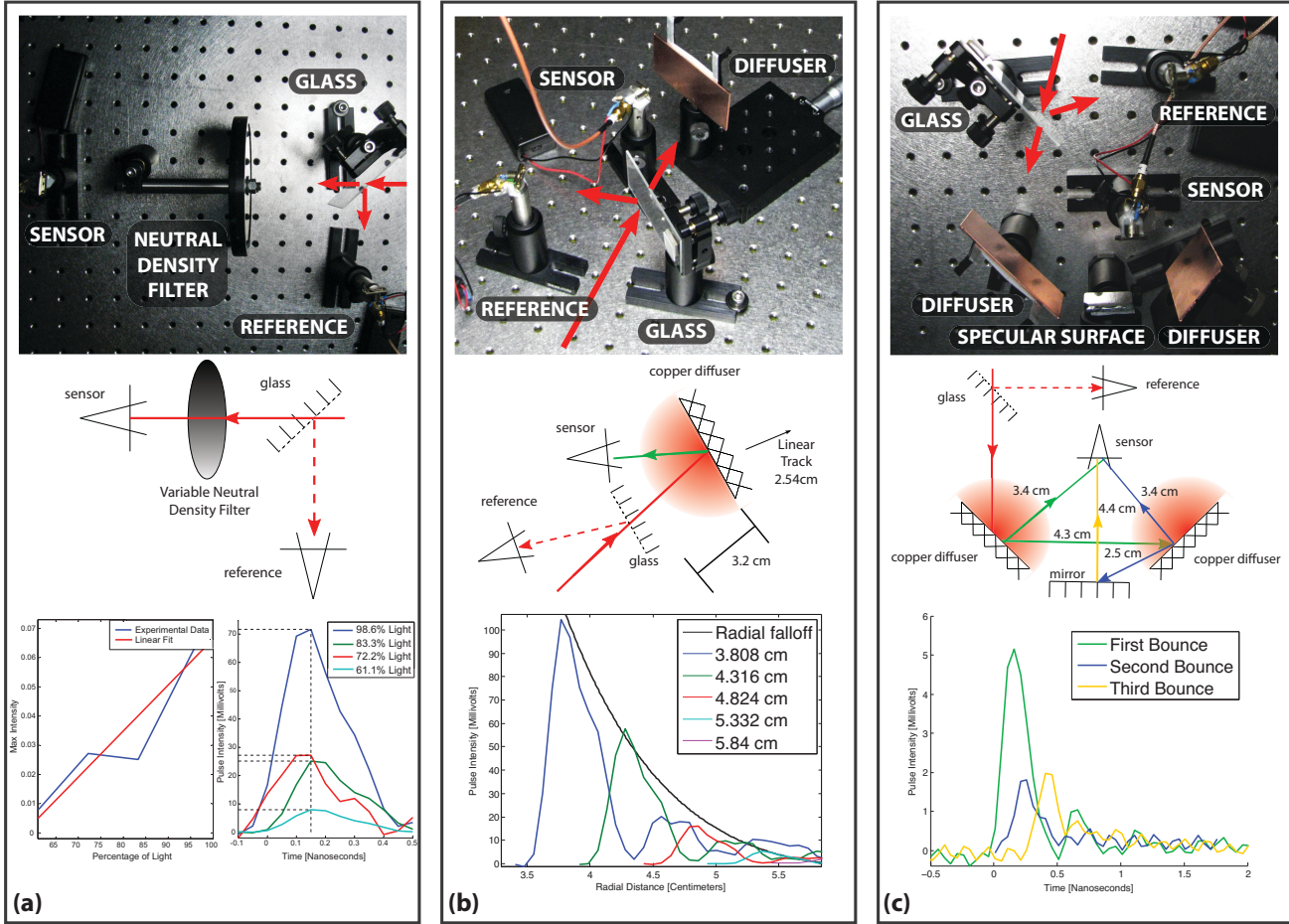


Figure 7. **Design and Verification of a Transient Imaging Camera.** (a) The ray impulses are recorded after being attenuated by a varying neutral density filter. The peak pulse intensity decreases linearly with the attenuation. (b) The intensity of the first bounce from a diffuser obeys the inverse square fall-off pattern. (c) We are able to record pulse intensities that are discernible from the noise floor even after the ray impulse has been reflected by three (2 diffuse and 1 specular) patches. The time shifts are linearly proportional to the multi-path length.

We conducted four proof-of-concept experiments (Figure 7) in flatland (2D) to demonstrate the following key properties of a transient imaging camera: free space functioning, linearity, multi-path light collection, inverse square intensity falloff and time invariance. We achieved synchronization by triggering our pulses based on a reference photo sensor. A small part of the laser pulse was deflected into a reference photo sensor using a semi-reflective glass patch and all pulse arrivals (onsets) were measured as TDOA with respect to the reference pulse.

6. Applications and Experiments

We use the transient imaging prototype and the algorithmic framework developed in Section 4 to estimate geometry for objects that do not reflect any light to camera due to specularly or occlusion.

Missing direct reflection: Consider the example shown in Figure 8(top) comprising a mirror and a diffuser. In traditional cameras it is difficult to estimate the distance to

a specular surface because there is no direct reflection received at the camera. Using transient imaging analysis, we can estimate the distances to specular surfaces by observing indirect bounces. If we aim the laser, L , towards a mirror (in a known direction) it will strike an unknown point on M . The reflected light will then illuminate points on the diffuser. Separately, the position and the depth of the diffuser, x , is estimated via stereo triangulation (using the known angle of the laser beam) or ToF (Section 4.1). When the laser illuminates M , the total path length sensed at a pixel observing D is $(z + y + x)$. Since x is known, the point M is obtained using conic multilateration. Note that, in dual photography [10], we create the dual image, i.e. the projector view, but that does not allow 3D estimation. We conducted 3 raster scans and assumed $z_1 = z_2 = z_3 = z$. The path lengths $z_i + x_i + y_i, i = 1, 2, 3$ were estimated using TDOA. In this experiment, we incurred a position error of 1.1662 cm and a maximum distance error of 7.14% in coordinate reconstruction by multilateration.

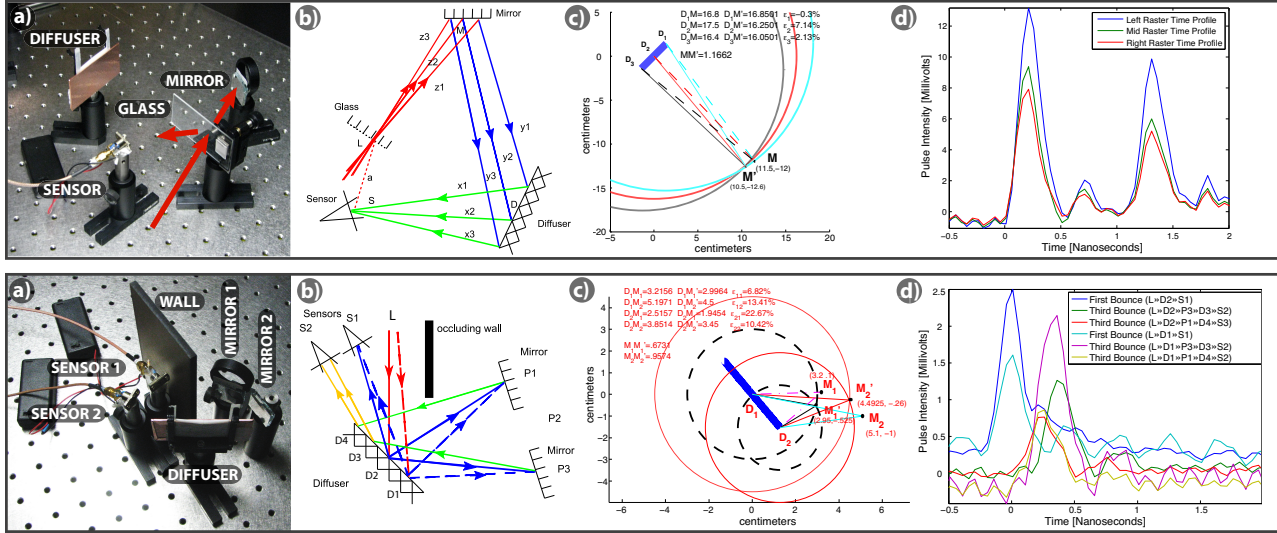


Figure 8. **Missing direct reflection (Top):** (a) A photo of the setup. (b) Ray diagram describing the light pulse path in 3 raster scans. (c) Plot showing multilateration using the 3 raster scans data: original and reconstructed scene geometries. (d) Oscilloscope data plot showing the TDOA between the 2nd bounce and the reference signal. **Looking around the corner (Bottom):** (a) A photo of the setup showing hidden 1-0-1 barcode. The sensors and the laser are completely shielded from the barcode. (b) Ray diagram tracing the paths of 1st and 3rd bounces in the 2 raster scans. (c) Plot showing the scene geometry reconstructed using multilateration. (d) Oscilloscope data plot showing the 1st bounce and the two separately recorded 3rd bounces, for both raster scans. Note the very small delay (≤ 200 ps) between two 3rd bounce arrivals. Please zoom in the PDF version for details.

Looking Around the Corner: We show an example of multi-path analysis in a scene that contains patches which are not visible to either the camera or the illumination source. Consider the ray diagram shown in Figure 8(bottom). Only light rays that have *first* bounced off the diffuser reach the hidden patches P_1, P_2, P_3 . Light that is reflected from the hidden patches (second bounce) can only reach the camera once it is reflected off the diffuser again (third bounce). The position and depth of the points on the diffuser are estimated using first bounce onsets. We then raster scan across the diffuser and measure the time difference of arrival (TDOA) between the first and third bounce onsets. We imaged a hidden 1 – 0 – 1 barcode using the first and third bounces off of a single diffuser. We used sensors, S_1 and S_2 , and a femtosecond laser source, L , neither of which had the barcode in their line of sight. The patches P_1 and P_3 were ground mirrors and P_2 was free space. The mirrors were aligned to maximize the SNR required for registering a third bounce. The maximum separation between P_1 and P_3 was limited to 5 cm because of SNR considerations. The first bounce, LD_1S_1 , was recorded by S_1 , and the two third bounces from the hidden patches, $LD_1P_1D_4S_2$ and $LD_1P_3D_3S_2$, arrived at S_2 within 200 ps of each other. Our current sensor was not fast enough and could only record the sum of the two third bounces. The two bounces can be recorded more accurately with a faster picosecond sensor or separated using deconvolution using S_2 's impulse response. As a proof of concept, we computed a high quality estimate by blocking P_1 and P_3 , one at a time. The reconstruction results are shown in Figure 8(bottom). We incurred a maximum error of 0.9574 cm in coordinate reconstruction.

7. Future Work

Our work, as it stands, has several limitations which make it challenging to generalize the transient imaging method to more complex, general scenes. However, with sophisticated modeling and advanced instrumentation, it is possible to alleviate these restrictions.

Experimental: Isolating onsets in practice is noisy, as onsets do not arrive at discrete instances; rather, they arrive as a continuous time profile. Though we have assumed a discrete patch model, future research should include continuous surface models and utilize tools in differential geometry to model the transport in general scenes. Additionally, the ray impulses are low pass filtered by the sensor response. All these reasons cause a broad temporal blur, rather than a sharp distinct onset. A noise model for the transient imaging camera can be found in supporting material 8. Defocus blur causes some scene patches to correspond to the same camera pixel. We alleviate this by working within the camera's depth-of-field. Every bounce off a diffuse surface creates considerable light loss and, thus, impacts the SNR. Another challenge is to collect strong multi-path signals (requiring single photon sensitivity) with ultra-fast time sampling. Commercial solutions, such as Intensified CCD cameras, allow image acquisition at very low light levels and at relatively high gating speeds (200ps or lower). The illumination source must be powerful enough to overcome ambient light. Also, our current method will not work for scenes which have arbitrarily placed highly specular objects, though reasoning may be improved with the use of appropriate priors.

Theoretical: Novel noise models for the transient imaging camera are required to account for uncertainties due to light-matter interplay. If two or more scene patches are occluded from each other ($v_{ij} = 0, i, j \neq 0$), our theoretical model fails. We circumvent this problem by using our transient imaging framework locally, with a subset of scene patches that satisfy our assumptions. The number of STIR measurements grow polynomially with number of patches, but the onset labeling complexity is exponential in the number of bounce orders used for inversion. Our framework will benefit from optimization-based onset labeling to account for time arrival uncertainties. We made a set of strong *a priori* assumptions for hidden surface estimation. Statistical regularization schemes, along with scene geometry priors, will allow us to extend transient reasoning to complex scenes where hidden surfaces may involve local scattering.

8. Conclusion

The goal of this paper is to explore the opportunities in multi-path analysis of light transport. To our knowledge, neither the transient light transport theory nor the presented imaging experiments have ever been conceived in literature. We developed the theoretical basis for analysis and demonstrated potential methods for recovering scene properties for a range of simple scenarios. Emerging trends in femtosecond accurate emitters, detectors and nonlinear optics may support single-shot time-image cameras. Upcoming low-cost solid state lasers will also support ultra-short operation. The key contribution here is exploration of a new area of algorithms for solving hard problems in computer vision based on time-image analysis. This, in turn, has tremendous applications in other areas such as scatter-free reconstruction in medical imaging, better layout understanding for fire and rescue personnel (Figure 1), tracking beyond line of sight in surveillance and car collision avoidance at blind corners, and robot path planning with extended observable structure. New research will adapt current work in structure from motion, segmentation, recognition and tracking to a novel time-image analysis that resolves shapes in challenging conditions.

Supplementary Material

<http://cameraculture.media.mit.edu/femtotransientimaging>

Acknowledgments

We thank the anonymous reviewers for critical suggestions. We are very thankful to Prof. Neil Gershenfeld (CBA), Prof. Joseph Paradiso, Dr. Franco Wong, Dr. Manu Prakash (MIT) for making available their opto-electronic apparatus that made our experiments possible. We would also like to thank MIT EECS Professors George Verghese, Franz Kaertner, Rajeev Ram for invaluable discussions. MIT undergraduate students: George Hansel, Kesavan Yogeswaran, assisted in carrying out initial experiments. We

thank Biyeun Buczyk (MIT) for greatly improving the figures and Paula Aguilera (MIT) for creating the video presentation. We finally thank Gavin Miller, Adam Smith and James Skorupski for several initial discussions.

References

- [1] J Kajiya, “The rendering equation”, in *SIGGRAPH '86*. 2
- [2] J Arvo, “Transfer equations in global illumination”, in *SIGGRAPH '93 Course Notes*. 2
- [3] R Raskar and J Davis, “5d time-light transport matrix: What can we reason about scene properties”, *Int. Memo '07*. 2
- [4] A Smith, J Skorupski, and J Davis, “Transient rendering”, *UC Santa Cruz TR UCSC-SOE-08-26, Feb '08*. 2
- [5] G Patow and X Pueyo, “A survey of inverse rendering problems”, *Computer Graphics Forum '03*. 2
- [6] N Kiriakos and E Steger, “A theory of refractive and specular 3d shape by light-path triangulation”, *IJCV '07*. 2
- [7] N Morris and K Kutulakos, “Reconstructing the surface of inhomogeneous transparent scenes by scatter trace photography”, in *ICCV '07*. 2
- [8] S K Nayar, K Ikeuchi, and T Kanade, “Shape from inter-reflections”, in *ICCV '90*. 2
- [9] R Ramamoorthi and P Hanrahan, “A signal-processing framework for inverse rendering”, in *SIGGRAPH '06*. 2
- [10] P Sen, B Chen, G Garg, S Marschner, M Horowitz, M Levoy, and H Lensch, “Dual photography”, in *ACM SIG. '05*. 2, 6
- [11] S K Nayar, G Krishnan, M Grossberg, and R Raskar, “Fast separation of direct and global components of a scene using high frequency illumination”, in *SIGGRAPH '06*. 2
- [12] S M Seitz, Y Matsushita, and K N Kutulakos, “A theory of inverse light transport”, in *ICCV '05*. 2
- [13] G W Kamerman, “Laser radar”, *Chap 1 of The Infrared and Electro-Optical System Handbook*. 2
- [14] G J Iddan and G Yahav, “3d imaging in the studio (and elsewhere...)”, in *Proc. SPIE, '01*. 2
- [15] R Lange and P Seitz, “Solid-state time-of-flight range camera”, *Quan. Elec., IEEE Jour. of, '01*. 2
- [16] H G Banos and J Davis, “Computing depth under ambient illumination using multi-shuttered light”, in *CVPR '04*. 2
- [17] N Vandapel, O Amidi, and J R Miller, “Toward laser pulse waveform analysis for scene interpretation”, in *ICRA '04*. 2
- [18] J M Schmitt, “Optical coherence tomography (oct): A review”, *IEEE Quan. Elec. '99*. 2
- [19] Denk W, Strickler J, and Webb W, “Two-photon laser scanning fluorescence microscopy”, *Science '90*. 2
- [20] A Campillo and S Shapiro, “Picosecond streak camera fluorometry -a review”, *Quan. Elec., IEEE Jour. of, '83*. 2
- [21] J Itatani, F Quéré, G L Yudin, M Y Ivanov, F Krausz, and P B Corkum, “Attosecond streak camera”, *Physical Review Letters '02*. 2
- [22] J Dattorro, “Convex optimization & euclidean distance geometry”, *Meboo Publishing USA*. 4