

Looking Around the Corner using Ultrafast Transient Imaging

Ahmed Kirmani · Tyler Hutchison · James Davis ·
Ramesh Raskar

Received: 22 May 2010 / Accepted: 23 May 2011 / Published online: 23 June 2011
© Springer Science+Business Media, LLC 2011

Abstract We propose a novel framework called *transient imaging* for image formation and scene understanding through impulse illumination and time images. Using time-of-flight cameras and multi-path analysis of global light transport, we pioneer new algorithms and systems for scene understanding through time images. We demonstrate that our proposed transient imaging framework allows us to accomplish tasks that are well beyond the reach of existing imaging technology. For example, one can infer the geometry of not only the visible but also the hidden parts of a scene, enabling us to *look around corners*. Traditional cameras estimate intensity per pixel $I(x, y)$. Our transient imaging camera captures a 3D time-image $I(x, y, t)$ for each pixel and uses an ultra-short pulse laser for illumination. Emerging technologies are supporting cameras with a temporal-profile

per pixel at picosecond resolution, allowing us to capture an ultra-high speed time-image. This time-image contains the time profile of irradiance incident at a sensor pixel. We experimentally corroborated our theory with free space hardware experiments using a femtosecond laser and a picosecond accurate sensing device. The ability to infer the structure of hidden scene elements, unobservable by both the camera and illumination source, will create a range of new computer vision opportunities.

Keywords Light transport · Global illumination · Multi-path analysis · Inverse problems · Inverse rendering · Computational Imaging

The Marr Prize is awarded to the best paper(s) at the biannual flagship vision conference, the IEEE International Conference on Computer Vision (ICCV). This paper is an extended and re-reviewed journal version of the conference paper that received Honorable Mention in 2009.

Electronic supplementary material The online version of this article (doi:10.1007/s11263-011-0470-y) contains supplementary material, which is available to authorized users.

A. Kirmani (✉) · T. Hutchison · R. Raskar
MIT Media Lab, Cambridge, MA 02139, USA
e-mail: akirmani@mit.edu

T. Hutchison
e-mail: thutch@mit.edu

R. Raskar
e-mail: raskar@mit.edu

J. Davis
Department of Computer Science, UC Santa Cruz, Santa Cruz,
CA 95064, USA
e-mail: davis@cs.ucsc.edu

1 Introduction

How can we use time-of-flight cameras for performing multi-path analysis of global light transport? If so, then how can this allow the design of new algorithms and systems for scene understanding through time images? Can we use such a framework for estimating 3D geometry and appearance of not only visible but also hidden parts of a scene? We propose a novel theoretical framework for image formation and scene understanding using impulse illumination and time images. We call this framework *Transient imaging*. We also propose the design of an ultrafast imaging device based on this framework which we term as a *Transient imaging camera*. We used our inverse algorithms and measurements collected using the transient imaging camera to estimate the hidden scene geometry from a fixed single viewpoint.

The light field or the *Plenoptic function* describes the amount of light traveling in every direction through every point in space at any wavelength at any point in time. It is

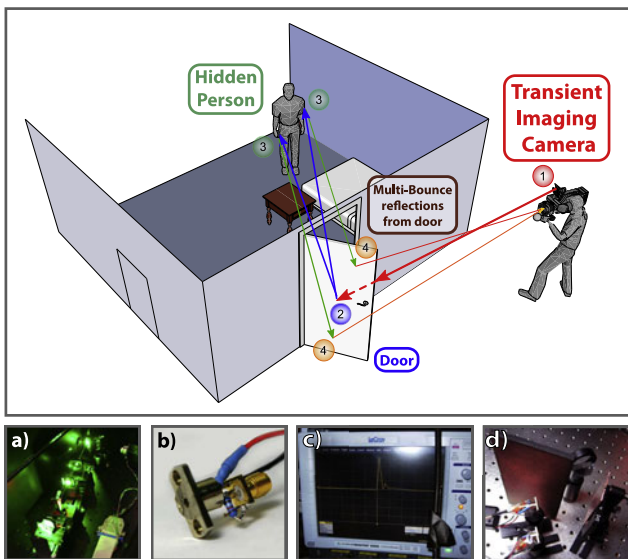


Fig. 1 Can you look around the corner into a room with no imaging device in the line of sight? In this example scenario, we conceptually illustrate that by emitting short pulses (labeled 1–2), and analyzing multi-bounce reflection from the door (4–1) using our proposed framework, it is possible to infer hidden geometry even if the intermediate bounces (3) are not visible. We experimentally validated our transient imaging framework for a simple scene shown in (d). Our transient imaging camera prototype comprises of (a) femtosecond laser illumination (b) picosecond-accurate detectors and (c) an Ultrafast sampling oscilloscope. We measured the Space Time Impulse Response (STIR) of a simple scene shown in (d) containing a hidden 1–0–1 barcode and reconstructed the hidden barcode

a function $L(x, y, z, \theta, \phi, \lambda, t)$ which is a function of three spatial dimensions, two angular dimensions, wavelength and time. The time parameter is usually ignored since traditionally all measurements are made in steady state, where the ray intensities do not change over time. In a traditional camera, a limited 2D projection of the complete light field of the scene is captured as an intensity value $I(x, y, c)$ per pixel by integrating along angular, temporal and wavelength dimensions during the exposure time.

$$I(x, y, c) = \int_{t=0}^{T_{exp}} \int_z \int_{\theta} \int_{\phi} L(x, y, z, \theta, \phi, \lambda, t) \delta(\lambda - c)$$

Many distinct scenes result in identical projections, and thus identical pixel values. Hence it is very difficult to use traditional imaging to estimate properties such as the depth of a mirror, the position of a sheet of glass, or the overall scale of a scene because these properties are not directly observable in the RGB values reported by a traditional camera. Another seemingly impossible problem is to estimate the structure and geometry of scene elements that are occluded from the camera and the light source. Since no light signal from the hidden scene parts will directly reach the sensor, the reflectance information of hidden scene elements gets blurred with the visible scene information in a non recoverable manner. In a room-sized environment, a microsecond

long 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.

It is also possible to sample geometric dimensions of the plenoptic function. Light field cameras (Veeraraghavan et al. 2007; Ng et al. 2005) sample both the incident angle of light and the spatial location. They have paved the way for powerful algorithms to perform many unconventional tasks such as extended depth of field, refocusing after a photo has been taken and depth estimation from fixed viewpoint. We propose that time sampling the incident light will provide even more powerful tools. This problem is inspired by the cumulative technological advances in ultra-fast imaging and photonics in the past decades. We exploit the fact that it is now possible to capture light at extremely short time scales. Light travels about 1 foot/nanosecond,¹ and existing commercial ultra-fast imaging hardware that is capable of sampling light at even sub-picosecond scales. Ultra-fast illumination sources such as femtosecond lasers have been available for decades. Note that it is well known that light scatters in a scene and takes a complex multipath from illumination source \rightarrow scene elements \rightarrow sensor (Nayar et al. 2006; Seitz et al. 2005; Sen et al. 2005). This process is termed as *light transport*. But because light travels so fast, we need ultra-fast light sampling to resolve these multipath.

Our methods exploit the dynamics of light transport. Critical to our work is the distinction between steady-state and transient light transport. Traditional cameras capture 2D images which are projections of light field generated by steady-state light transport. This corresponds to the familiar case of global light transport equilibrium, in which the speed of light is infinity and it is conveniently assumed that light takes no time to traverse all the different multi-paths in the scene to reach the state of steady state light transport. In transient imaging, we eschew this simplification. An impulse of light evolves into a complex pattern in time. The dynamics of transient light transport can be extremely complex, even for a simple scene. Recent advances in ultra-high speed imaging have made it possible to sample light as it travels 0.3 millimeter in 1 picosecond (1 picosecond = 10^{-12} s). At this scale, it is possible to reason about the individual paths light takes within a scene. This allows direct observation of properties such as distance, and the difference between primary and secondary reflections.

We propose a conceptual framework for scene understanding through the modeling and analysis of global light

¹ 1 ns = 10^{-9} s. Precisely light travels 0.3 mm in 1 picosecond = 10^{-12} s.

transport using time images. We explore new opportunities in multi-path analysis of light transport using time-of-flight sensors. Our approach to scene understanding is four-fold:

1. Measure the scene's transient photometric response function using the directional time-of-flight camera and active impulse illumination.
2. Estimate the structure and geometry of the scene using the observed STIR.
3. Use the estimated structure and geometry along with *a priori* models of surface light scattering properties to infer the scene reflectance.
4. Higher order inference engines can be constructed that use the estimated scene properties for higher level scene abstraction and understanding.

Our solution to the problem of single viewpoint looking around corners using transient imaging can be summarized in the pseudo-equation: 2-dimensional single viewpoint projection + 1-dimensional time sampling = 4-dimensional multiple viewpoint illumination and sensing.

1.1 Contributions

In this paper, we propose a first, initial approach for solving the hard inverse problem of recovering structure of completely occluded scenes based on transient imaging methods as opposed to traditional steady state imaging. Our method involves computational processing of time-samples of the light scattered in response to active, impulsive illumination of visible parts of the scene. We made the following contributions through this paper:

1. Developed a theoretical framework called *transient light transport* that uses time images and impulse illumination for scene understanding.
2. We proposed a transient imaging camera model which time samples incident light continuously and uses spatio-temporal impulse illumination as the light source.
3. We built a proof-of-concept hardware prototype comprising a femtosecond laser and a directionally sensitive, picosecond accurate photo sensor array, intensified CCD cameras and picosecond streak cameras to demonstrate the practical use of transient imaging theory. Using our prototype, we demonstrated all the key functionalities required in an imaging device: geometry, photometry, multi-bounce (multi-path) light separation and free-space functioning.
4. We also demonstrated 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. In particular, we reconstruct the structure and geometry of hidden planar scenes using our transient imaging theory and camera prototype. We have verified this through both simulation and experiments.

2 Related Work

Transient imaging is a new imaging domain. To our best knowledge, neither the transient light transport theory nor the presented imaging experiments have ever been conceived in literature. The basic differences between transient imaging and related imaging methods are summarized next. Also see Fig. 2 for a quick comparison and discussion of transient imaging with other popular imaging methods.

2.1 Theoretical Prior Art

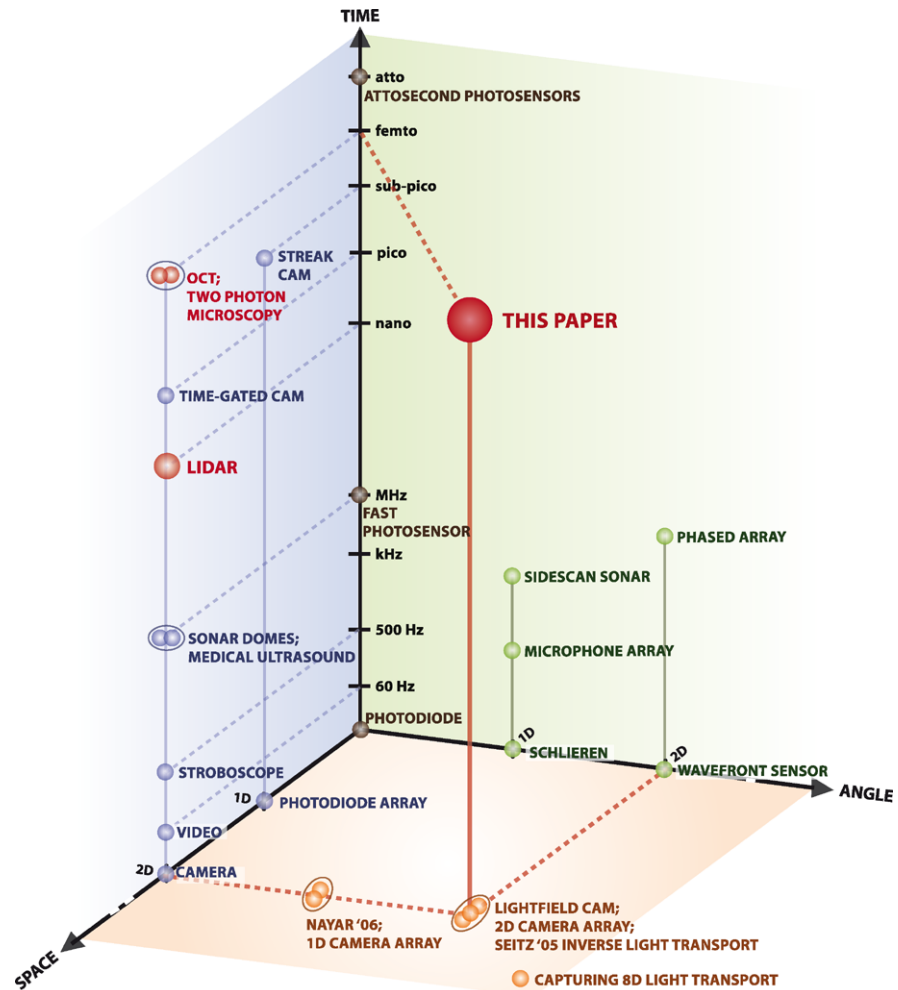
Global Light Transport Light often follows a complex path between the emitter and sensor. A description of steady-state light transport in a scene is referred to as the *rendering equation* (Kajiya 1986). Extensions have been described to include time in light transport (Arvo 1993). In Raskar and Davis (2007) proposed inverse analysis using a 5D time-light transport matrix to recover geometric and photometric scene parameters. In addition, Smith et al. (2008) proposed a modification of the rendering equation via a transient rendering framework. Accurate measurement of physical scene properties is called inverse-rendering (Patow and Pueyo 2003). Complex models have been developed for reconstructing specular (Kutulakos and Steger 2007), transparent (Morris and Kutulakos 2007), Lambertian (Nayar et al. 1990) scenes and joint lighting and reflectance (Ramamoorthi and Hanrahan 2001).

Capturing and Analysis of Light Transport Recent work in image-based modeling and computational photography has also shown several methods for capturing steady-state light transport (Sen et al. 2005). The incident illumination is represented as a 4D illumination field and the resultant radiance is represented as a 4D view field. Taken together, the 8D reflectance field represents all time-invariant interaction of a scene. Our work is 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 (Nayar et al. 2006). Seitz et al. (2005) have decomposed images into multi-bounce components under the assumption that the scene is Lambertian. Although the dual photography approach (Sen et al. 2005) 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.

2.2 Hardware and Experimental Prior Art

SONAR (SOUND Navigation And Ranging), is a technique that uses sound propagation in a medium such as air or water to detect and locate remote objects. The speed of sound

Fig. 2 Popular imaging methods plotted in the *Space-Angle-Time* axes. With higher dimensional light capture, we expand the horizons of scene understanding. A LIDAR generates 2D image, but unlike a time gated camera, it is essentially a 1D (time/range) sensor, and image is generated by scanning. Our concept transient imaging camera shown in Fig. 6 uses LIDAR-like imaging hardware, but, in contrast, we exploit the multi-path information which is rejected in both LIDAR (Light Detection And Ranging) and OCT (Optical coherence tomography). A phased array (antenna) an angle related device—but unlike the wave-front sensor that is used for measuring angles, a phased array is usually combined, for example with a pulse radar. Note that the phase here is a function of both the beam steering during transmission and the angular sensitivity during reception. Note that this figure mixes time and frequency in one axis. Additionally, many phase arrays also work at gigahertz range



is six orders of magnitude slower than the speed of light, and therefore easier to detect. Nevertheless work in SONAR has produced intricate models of the effects of many surfaces with complicated scattering properties. In addition, detection *over the corner* as well as exploiting multi-path in the context of Over the Horizon RADAR, and Synthetic Aperture RADAR have been demonstrated before (Garren et al. 2004, 2005; Sarunic et al. 2001).

LIDAR (Light Detection And Ranging) systems modulate light, typically on the order of nanoseconds, and measure the phase of the reflected signal to determine depth (Kamerman 1993). Flash LIDAR systems use a 2D imager to provide fast measurement of full depth maps (Iddan and Yahav 2001; Lange and Seitz 2001). Importantly, a number of companies (Canesta <http://canesta.com/>; MESA <http://www.mesa-imaging.ch/>; 3DV <http://www.3dvsystems.com/>; PMD <http://www.pmdtec.com/>) 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 (Gonzalez-Banos and Davis 2004). 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 (Vandapel et al. 2004).

LIDAR systems have been successful in some circumstances, but they work well only for certain types of surface reflectance, and do little to help estimate other global properties such as relationship between scene patches. In addition they are used in restrictive configurations by carefully placing emitters near the receivers. We need a generalized sensor which fundamentally records a greater portion of the light transport in a scene. This sensor could then be used to design new algorithms and specialized sensing methods.

Time-gated imaging captures a gated image, $I(x, y, t_\delta)$, 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. It is possible to construct the transient photometric response function using gated imagers, e.g. Busck et. al show a TPRF measured to 100 picosecond accuracy.

Streak cameras Streak cameras are ultrafast photonic recorders which deposit photons across a spatial dimension, rather than integrating them in a single pixel. Using a 2D array, $I(x, y_\delta, t)$ can be measured. Sweeping the fixed direction, y_δ , allows $I(x, y, t)$ to be captured. Picosecond streak cameras have been available for decades (Campillo and Shapiro 1983). Modern research systems can function in the attosecond range (Itatani et al. 2002).

Femtosecond Imaging Optical coherence tomography (OCT) (Schmitt 1999), an interferometric technique, and two-photon microscopy (Denk et al. 1990), using fluorescence, allow high-quality, micrometer resolution 3D imaging of biological tissue. Both these methods are based on pulsed femtosecond illumination. Femtosecond windows also allow ballistic photons to be separated from scattered photons while imaging in biological tissue. 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. All the existing methods based on time sampling of light make no use of global light transport reasoning to infer scene characteristics. They 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 Imaging Framework

In our transient light transport framework, we assume that the speed of light is some finite value and light takes a finite amount of time to travel from one scene point to the other. As light scatters around a scene, it takes different paths, and longer paths take a longer time to traverse. Even a single pulse of light can evolve into a complicated pattern in time. The dynamics of transient light transport can be extremely complex, even for a simple scene. The theory of light transport describes the interaction of light rays with a scene. Incident illumination provides first set of light rays that travel towards other elements in the scene and the camera. The direct bounce is followed by a complex pattern of inter-reflections whose dynamics is governed by the scene geometry and material properties of the scene elements. This process continues until an equilibrium light flow is attained.

We consider a scene \mathcal{S} (Fig. 3) composed of M small planar facets (patches with unit area) p_1, \dots, p_M with geometry $G = \{Z, D, N, V\}$ comprising of the patch positions

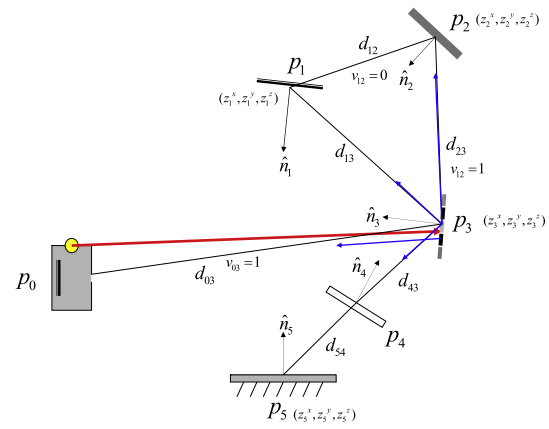


Fig. 3 A scene consisting of $M = 5$ patches and the illumination-camera patch p_0 . The patches have different spatial coordinates (z_i^x, z_i^y, z_i^z) , orientations \hat{n}_i and relative visibility between patches v_{ij} . The patches also have different material properties, for instance p_4 is diffused, p_4 is translucent and p_5 is a mirror

$Z = [z_1, \dots, z_M]$ where each $z_i \in \mathbb{R}^3$; the distance matrix $D = [d_{ij}]$ where $d_{ij} = d_{ji}$ is the Euclidean distance between patches p_i and p_j ; the relative orientation matrix $N = [\hat{n}_1, \dots, \hat{n}_M]$ consists of unit surface normal vectors $n_i \in \mathbb{R}^3$ at patch p_i with respect to a fixed coordinate system; and the visibility matrix $V = [v_{ij}]$ where $v_{ij} = v_{ji} = 0$ or 1 depending on whether or not patch p_i is occluded from p_j . For analytical convenience, we consider the camera (observer) and illumination (source) as a single patch denoted by p_0 . All the analysis that follows can be generalized to include multiple sources and the observer at an arbitrary position in the scene.

We introduce a variation of the rendering equation (Kajiya 1986; Immel et al. 1986) in which we represent the time light takes to traverse distances within the scene by a finite delay. Let t denote a time instant and $\{L_{ij} : i, j = 0, \dots, M\}$ be the set of radiances for rays that travel between scene patches. Transient light transport is governed by the following dynamical equation which we term as *transient light transport equation*:

$$L_{ij}[t] = E_{ij}[t] + \sum_{k=0}^M f_{kij} L_{ki}[t - \delta_{ki}] \quad (1)$$

Equation (1) states that the scalar ray radiance $L_{ij}[t]$ leaving patch p_i towards p_j at time t is the sum of emissive radiance $E_{ij}[t]$ and the form factor weighted delay sum of the radiances from other patches. For simplicity, let the speed of light $c = 1$. Then the propagation delay δ_{ij} is equal to the distance d_{ij} (see Fig. 5). We assume that all delays δ_{ij} are integer multiples of a unit delay. The scalar weights f_{kij} or *form factors* denote the proportion of light incident from

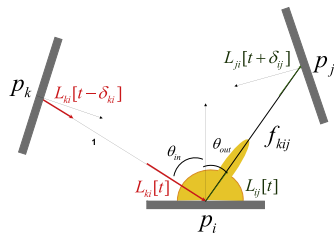


Fig. 4 Form factors f_{kij} are the proportion of light incident from patch p_k on to p_i that will be directed towards p_j

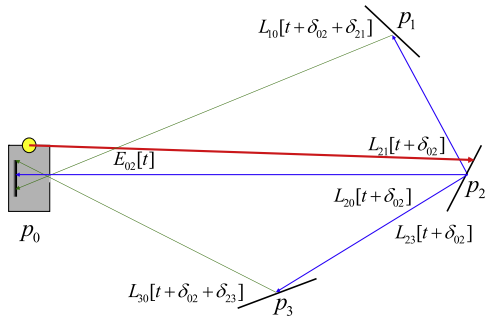


Fig. 5 A ray impulse $E_{02}[t]$ directed towards patch p_2 at time t . This ray illuminates p_2 at time instant $t + \delta_{02}$ and generates the directional radiance vector $[L_{20}[t + \delta], L_{21}[t], L_{23}[t]]$. These light rays travel towards the camera p_0 and scene patches p_1 and p_3 resulting in global illumination

patch p_k on to p_i that will be directed towards p_j :

$$f_{kij} = \rho_{kij} \left(\frac{\cos(\theta_{in}) \cos(\theta_{out})}{\|z_i - z_j\|^2} v_{ki} v_{ij} \right)$$

where ρ_{kij} is the directional reflectance which depends on the material property and obeys Helmholtz reciprocity ($\rho_{kij} = \rho_{jik}$), θ_{in} is the incident angle and θ_{out} is the viewing angle (see Fig. 4). Additionally, if the patch does not interact with itself and then $f_{kij} = 0$ for $k = i$ or $i = j$. We assume that the scene is static and material properties are constant over the imaging interval. The source and observer patch p_0 does not participate in inter-reflections.

We model illumination using the emitter patch p_0 . All other patches in the scene are non-emissive, $E_{ij}[t] = 0 : i = 1, \dots, M; j = 0, \dots, M; t = 0, \dots, \infty$. Illumination is the set of radiances $\{E_{0j}[t] : \forall j = 1, \dots, M; t = 0, \dots, \infty\}$ representing the light emitted towards all scene patches at different time instants. The outgoing light at patch p_i is the vector of directional radiances,

$$\mathbf{L}[i, t] = [L_{i0}[t], \dots, L_{iM}[t]]$$

and for the entire scene we have the transient light transport vector $\mathbf{L}[t] = [\mathbf{L}[1, t], \dots, \mathbf{L}[M, t]]$ which contains $(M(M - 1) + M)$ scalar irradiances. We can only observe the projection of $\mathbf{L}[t]$ that is directed towards the camera p_0 . At each time t we record a vector of M intensity values $\mathbf{L}^c[t] = [L_{10}[t - \delta_{10}], \dots, L_{M0}[t - \delta_{M0}]]^T$.

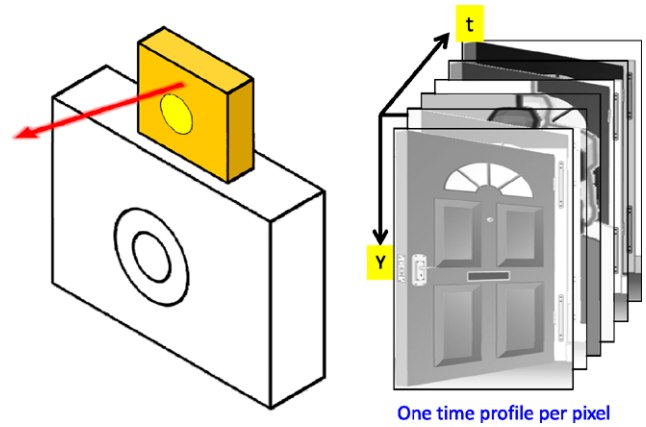


Fig. 6 Transient Imaging Camera captures a 3D time image in response to a ray impulse illumination

3.1 Transient Imaging Camera

The transient imaging camera 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*, $I(x_i, y_i, t)$ (see Fig. 6). 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).

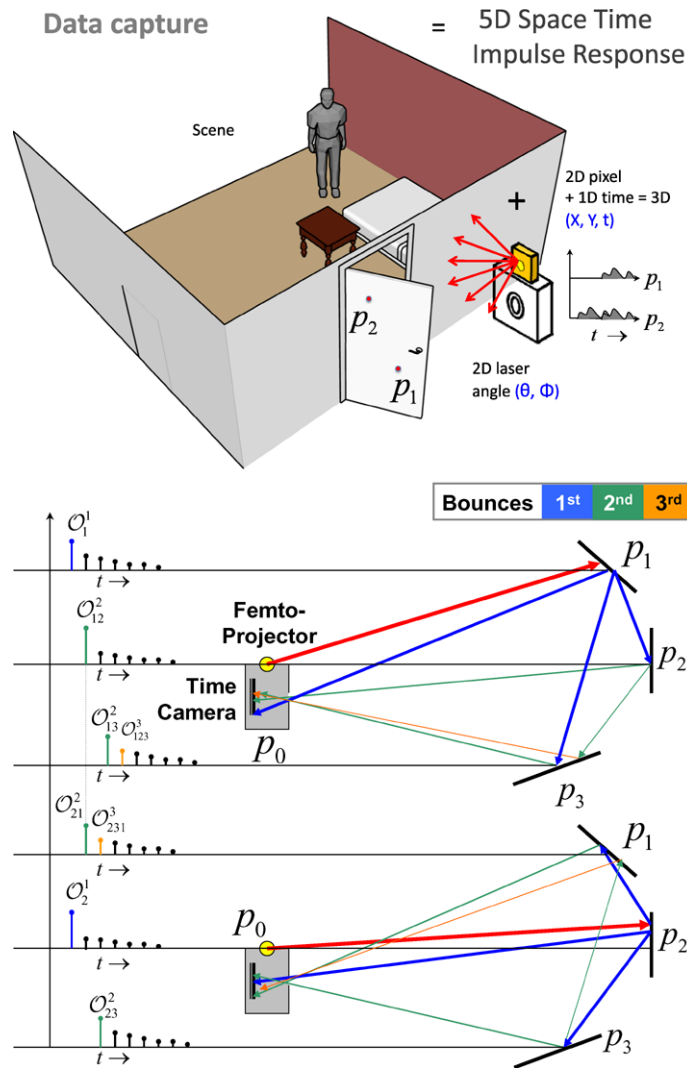
3.2 Space Time Impulse Response

The Space Time Impulse Response or the STIR of a scene \mathcal{S} denoted by $\mathbf{STIR}(\mathcal{S})$ is a collection of time images, each captured with an impulse ray aimed the direction (θ_j, ϕ_j) , illuminating a single scene patch p_j . This is a 5D function: $\mathbf{STIR}(x_i, y_i, \theta_j, \phi_j, t)$. We measure the STIR as follows (see Fig. 7):

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

In still scenes with static material properties, transient light transport is a Linear and Time Invariant (LTI) process. By linear, we imply that if we illuminate the scene as just described and record the STIR but with a different scalar multiple of the illumination intensity, then the received intensities that comprise the STIR will also be scaled with the same scalar multiple. By linearity, we also imply that the time-image that is received in response to simultaneous impulse illumination of two patches, say, p_i and p_j , is equal to the sum of the individual time-images, $I(x_i, y_i, t)$ and

Fig. 7 Measuring the STIR of a scene using the transient imaging camera. We successively illuminate a single patch at a time and record a 3D time image. Collection of such time images constitutes the 5D STIR



$I(x_j, y_j, t)$. By time-invariance, we imply that if we time-shift the illumination pulse by a certain delay or advance, the STIR time-profiles will also be time-delayed or advanced by exactly the same amount as the outgoing illumination pulse. We can express any general, time-varying illumination of the scene as a linear, time-shifted combination of STIR time-images, $I(x_i, y_i, t)$. Hence, the STIR of the scene characterizes its appearance under general illumination from a given transient imaging camera view-point. Note that this is similar to the characterization of a LTI system using its time impulse response except that the notion of an impulse in our case extends to both space and time. We note that the transient light transport equation (1) describes a Multiple-input Multiple-output linear time invariant (LTI) system. One of the important properties of an LTI system is that scaled and time shifted inputs will result in a corresponding scale and time shift in the outputs. Hence the STIR of the scene can be used to completely characterize its behavior under any general illumination.

We propose a conceptual framework for scene understanding through the modeling and analysis of global light transport. The transient imaging camera maps a scene to its space time impulse response (STIR), but how can we infer the scene structure given its space time impulse response? In Sect. 4 we propose an algorithmic framework for looking around corners, i.e., for reconstructing hidden scene structure for planar layered scenes.

4 Hidden Geometry Estimation

Unlike traditional time-of-flight imaging, our goal is to compute the direct distances, d_{0i} , using the first bounce, and the pair wise distances, d_{ij} . Instead of using intrinsic camera calibration, we exploit second and higher order bounces to estimate scene geometry. First, we use the onset information contained in the STIR to estimate pair wise distances. Then we compute a robust isometric embedding to determine patch coordinates.

4.1 Inverse Geometry Problem

We develop our formulation for a scene with the following strict assumptions:

1. Each patch is visible from all the other patches ($v_{ij} = 1, \forall i, j$). We also assume that there are no limiting cases in which the incoming illumination direction is perpendicular to the surface normal.
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 particular, when any patch is illuminated from any direction it scatters a non-zero amount of light towards all other patches.

In Sect. 4.1, we discuss the extension of our transient imaging framework to scenes consisting of patches hidden from the camera and illumination.

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 (Fig. 7). 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 (Fig. 7). 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 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 Fig. 7. The linear system for this scene is constructed as:

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

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.

Structure from Pairwise Distances The problem of estimating scene structure, \mathbf{Z} , from pair wise 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, Dattorro 2006). 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 Fig. 8. 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 (Dattorro 2006).

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 $\hat{\mathbf{Z}} = \mathbf{R}Z_e + \mathbf{T}$
-

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:

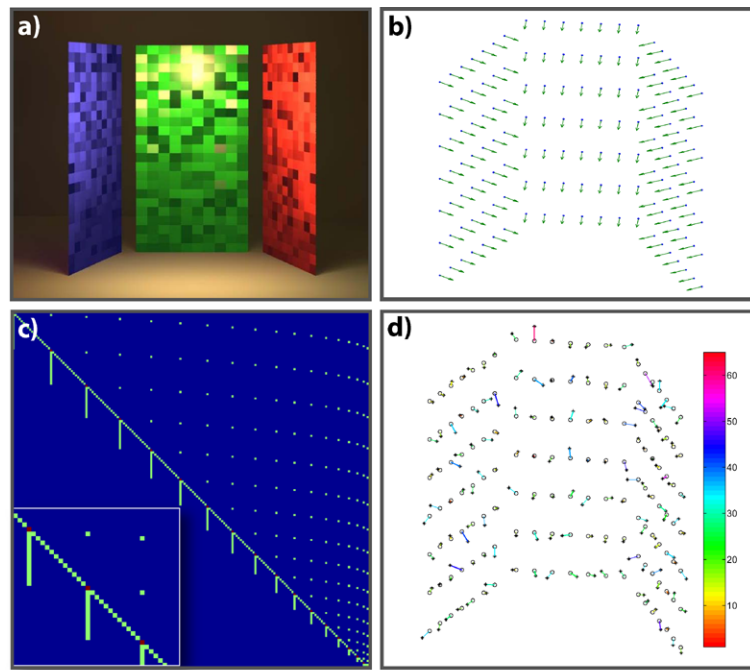
1. Labeling third onsets.
2. Inferring distances to hidden patches from integrated path lengths.

To estimate the structures of the hidden patches, we make the following three strong assumptions:

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*, $\mathcal{O}^3 = \{O_{ijk}^3 : i, j, k =$

Fig. 8 (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 $\sim \mathcal{N}(\mu, \sigma^2)$, $\mu =$ device resolution = 250 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



$1, \dots, M; i \neq j; j \neq k$ }, that corresponds to third bounces. Note that there are $O(M)$ first onsets, $O(M^2)$ second onsets and $O(M^3)$ third onsets. Also, Euclidean geometry imposes that $O_{ijk}^3 = O_{kji}^3$. Labeling the onsets contained in \mathcal{O}^3 is non-trivial. An 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 $\mathcal{O}^1, \mathcal{O}^2, \dots, \mathcal{O}^{N+3}$.

As a simple example, consider the scene in Fig. 9. 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 Sect. 4.1. The distances (d_{21}, d_{24}) and (d_{31}, d_{34}) are not directly observable. Once these distances are estimated, 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 $\mathbf{STIR}(\mathcal{S})$. 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 Fig. 9). This labeling procedure can be generalized for multiple hidden patches:

1. Estimate the distances to all the visible scene patches (Sect. 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 Sect. 4.1. An example of reconstructing hidden 3D geometry is shown in Fig. 10.

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

5 Hardware Prototype and Experiments

5.1 Transient Imaging Camera Prototype

The framework developed in Sect. 4.1 was corroborated with experiments conducted using a prototype transient

Fig. 9 (Color online) 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

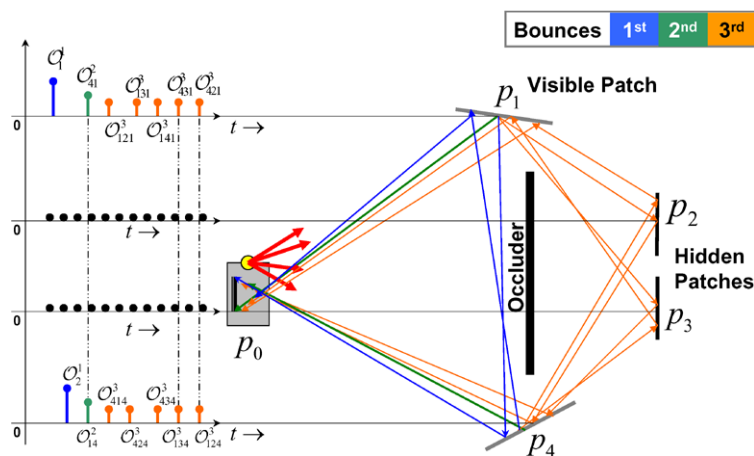
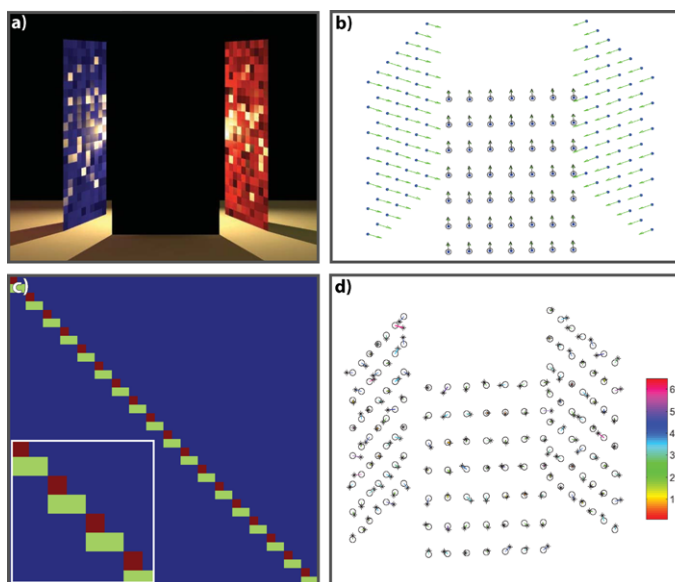


Fig. 10 (a) Estimating distances in scenes with hidden patches. Unknown to the estimation algorithm, the hidden patches are on a plane (shown in black). (b) Original patch geometry. We use 1st, 2nd and 3rd bounce onsets, our labeling algorithm and the T_3 operator (c) to estimate hidden geometry. (d) The isometric embedding error plot verifies negligible reconstruction error and near co-planarity of patches. Onset noise and color bar schemes are same as Fig. 8



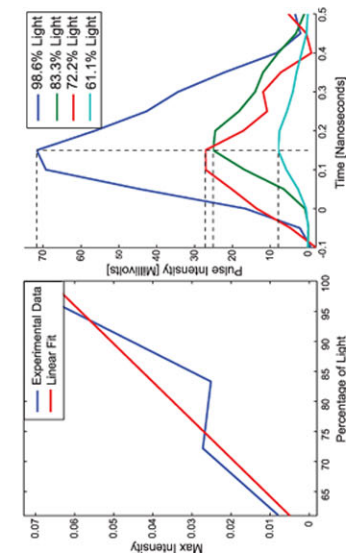
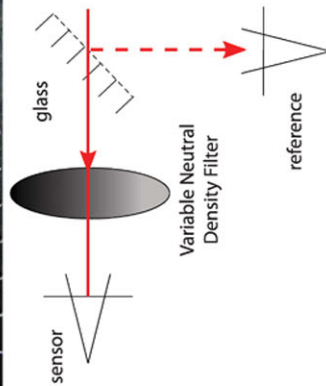
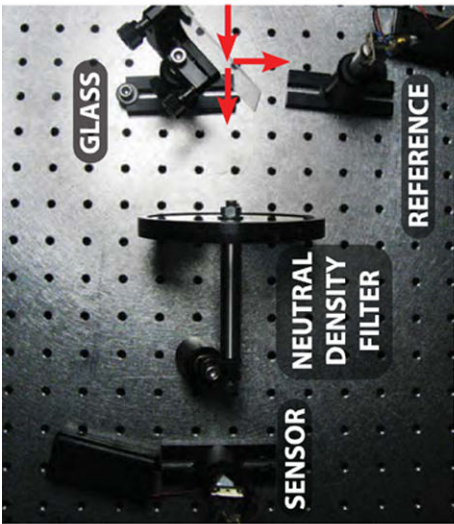
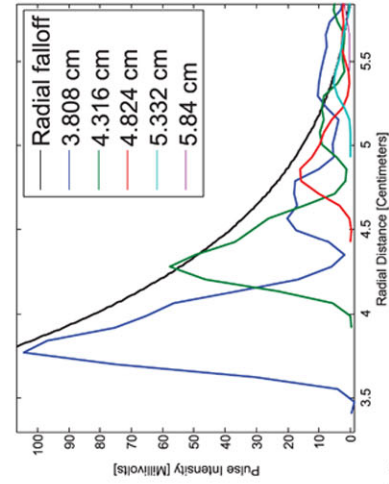
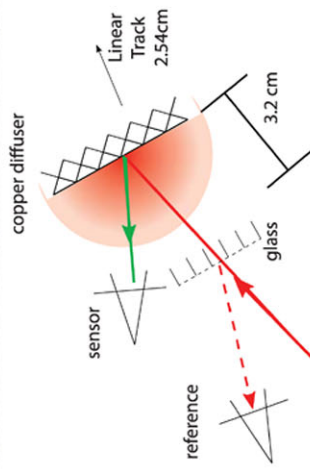
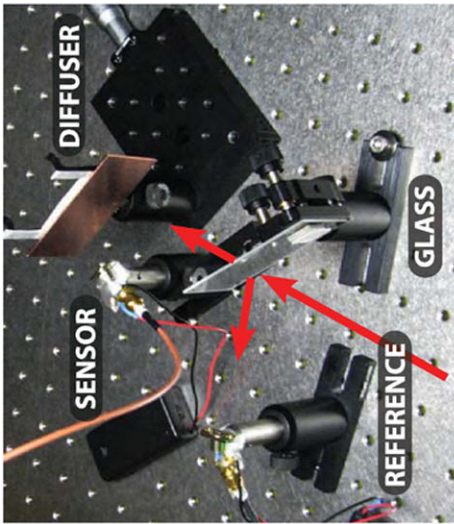
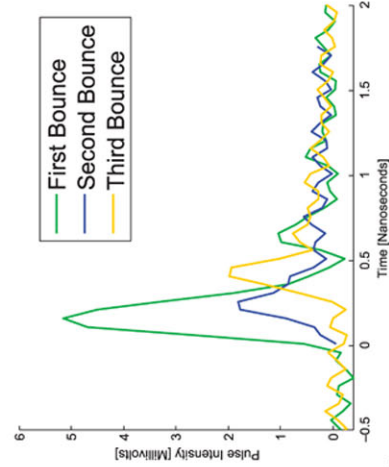
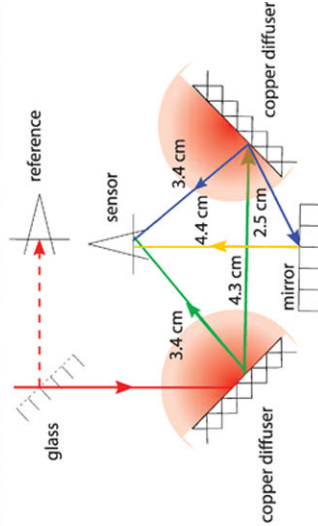
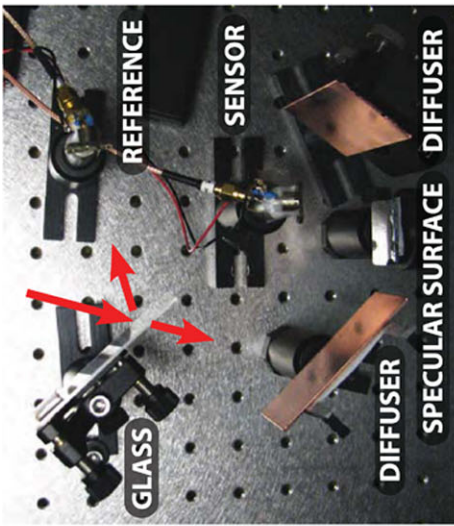
imaging camera. We demonstrated the feasibility with proof of concept experiments but not a full-fledged imaging apparatus. In particular, this prototype (Fig. 1) was developed with the intent to show that it is feasible to reason about multi-bounce global transport using the STIR.

A commercially-available reverse-biased silicon photo sensor (Thorlabs FDS02, \$72) was used as the ultrafast sensor. This sensor has an active area of 250 microns in diameter and a condensing lens to gather more light. A 5 GHz ultrafast oscilloscope digitized the photo-currents. The least count was 50 ps (1.5 cm light travel). The ray impulse source was a mode locked 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. The average laser power was 420 milliwatts, corresponding to a peak power of greater than 85 kW.

We were required to sample the incident light with picosecond resolution and be highly sensitive to a low pho-

ton arrival rate. Our depth resolution is limited by the response time of the detector and digitizer (250 ps, 7.5 cm 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.

Four proof-of-concept experiments (Fig. 11) were conducted 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. Synchronization was achieved 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.



(c)

(b)

(a)

Fig. 11 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

5.2 Hidden Geometry Reconstruction Experiments

The transient imaging prototype was used along with the algorithmic framework developed in Sect. 4.1 to estimate geometry for objects that do not reflect any light to camera due to specularity or occlusion.

Missing direct reflection Consider the example shown in Fig. 12 (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, estimated 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 , was estimated via stereo triangulation (using the known angle of the laser beam) or ToF (Sect. 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 (Sen et al. 2005), 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.

Looking Around the Corner We demonstrated an example of multi-path analysis in a scene that contains patches which were not visible to either the camera or the illumination source. Consider the ray diagram shown in Fig. 12 (see bottom row). 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 were estimated using first bounce onsets. We then raster scanned across the diffuser and measured 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 Fig. 12 (bottom). We incurred a maximum error of 0.9574 cm in coordinate reconstruction.

6 Limitations

Transient imaging is in a preliminary stage of theoretical modeling and experimentation. Our work, as it stands, has several limitations which make it challenging to generalize the transient imaging method to more complex, general scenes.

6.1 Theoretical Limitations

Fundamentally, transient imaging is an inverse problem. As with all inverse problems, there are inherent issues related to robustness and success rate. The inverse light transport problems posed in Sect. 4.1 have degenerate cases in which multiple solutions (scenes) exist for the same observable STIR. In the looking around corner scenario, we make two strong assumptions on the hidden scene geometry which are limiting: we assume that the number of hidden patches is known, and that the order of times-of-arrival of the onsets is preserved. Also, there is no proof for the uniqueness of our proposed solution or for the feasibility and scalability of our approach to more complex scenes, such as ones with occlusions and nonlinear geometry which causes hidden scene inter-reflections.

Some of these issues can be alleviated with the use of suitable prior models for scene geometry. The resulting parameter estimation problems may be solved using optimization based regularization schemes and robustness constraints. But it may still not be possible to guarantee 100% accuracy while using transient methods. Importantly, although our additional time-domain data is noisy, it still restricts the class of solutions to a greater extent than using the more limited data of traditional cameras.

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. This problem is circumvented 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

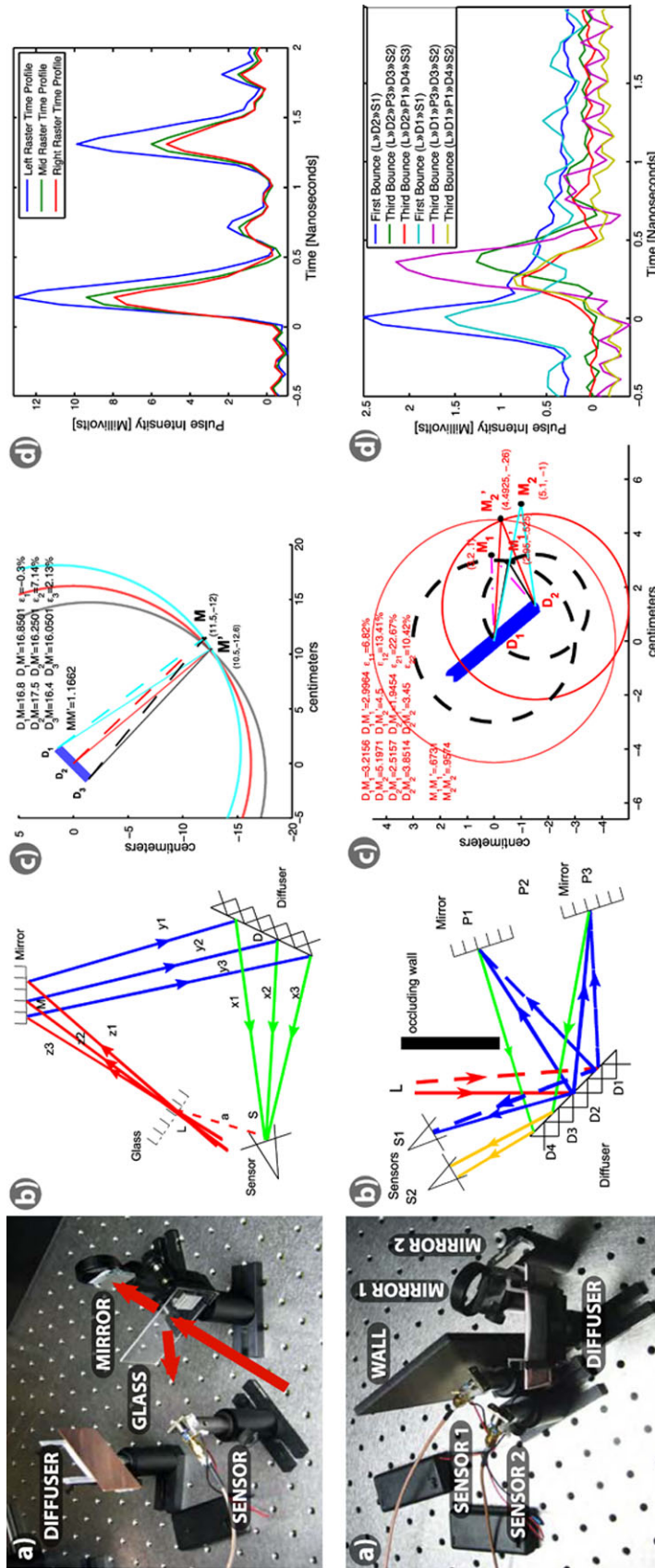


Fig. 12 *Missing direct reflection (top): (a)* A photo of the setup. *(b)* Ray diagram describing the light pulse path in 3 raster scans. *(c)* Plot showing multilateration using the 3 raster scans data: original and reconstructed scene geometries. *(d)* Oscilloscope data plot showing the TDOA between 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

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.

6.2 Experimental Limitations

Our approach shares practical limitations with existing active illumination systems in terms of power and use scenarios: the sources must overcome ambient illumination, and only scenes within finite distance from the camera can be imaged. The most significant practical limitation is very low signal-to-noise ratio (SNR) and dynamic range. This is because of the exponential decay of the signal with increasing number of bounces.

In addition, we require precise high-frequency pulsed opto-electronics at a speed and quality that are not currently available at consumer prices. An important challenge we face is to collect strong multi-path signals. Some of the other immediate technological challenges we face in implementing transient imaging for real scenarios include focusing light at extremely high sampling speeds, single photon sensitive detectors etc. Hardware solutions to some of our challenges may not be available immediately but recent advances trends in photonics and ultrafast sensing have indicated progress in that direction. Transient imaging may pose as an impetus to rapid advancements in device physics and photonics.

Also, our pulsed light source, a femtosecond laser has a pulse repetition frequency (PRF) of 93.68 MHz. This translates into an un-ambiguous range of 1.6 meters which includes higher order reflections. This range was sufficient for our experiment but is short for practical use. By decreasing the PRF or increasing the time-separation between subsequent light pulses, we can image large scale scenes, provided we can collect sufficient amount of light.

Direct reflections from scene elements are significantly stronger than light which has traveled a more complex path, possibly reflecting from diffuse surfaces. Every bounce off a diffuse surface creates considerable light loss and, thus, impacts the SNR. Thus a major 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 (200 ps or lower). The illumination source must be powerful enough to overcome ambient light. We expect that solid state lasers will continue to increase in power and frequency, doing much to alleviate these concerns but in the meanwhile we may make heavy use of existing LIDAR and OCT hardware in order to demonstrate preliminary applications of

this technique. 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.

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. 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.

7 Future Work

The approach presented in this paper is in the nascent stage. Our technique as it stands poses several open theoretical questions, such as, under what circumstances is there a unique solution to the scene structure? Since the problem is inherently ill-posed, can different viewpoints be combined to resolve the ambiguities? Can our problem benefit from other inverse methods, such as computerized tomography? Also, we assume that the form factor function (BRDF) of hidden scene patches is unknown. It is possible that assuming Lambertian reflectance may simplify the problem significantly. There are also a number of practical challenges, such as, how to deal with the low photon count and huge contrast since the intensity falls off exponentially with the number of reflections.

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. 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.

We propose to build a hardware prototype that functions in general scenes with better SNR and high spatio-temporal resolution. We need to develop a robust algorithmic inference framework to process the captured data-sets and estimate scene geometry. Through collaboration with applied optics and photonics researchers, we hope to use emerging solid state lasers and optics to build a single-shot time-image camera and port algorithms to modern processors. In the longer term, multiple problems need to be addressed to scale the solution beyond a room-sized environment. Ultimately,

a hybrid detector made up of a transient imager and a long wavelength imager will support better scene understanding in estimating hidden objects.

8 Summary of Approach

The goal of this paper is to explore the opportunities in multi-path analysis of light transport. 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 was the exploration of a new area of algorithms for solving hard problems in computer vision based on time-image analysis.

If the STIR of the scene is available then it can be directly used to obtain the geometry G of the scene by solving the inverse geometry problem discussed in Sect. 4.1. This procedure is summarized as:

1. Estimate the distance matrix D using time onsets.
2. Compute the coordinate set Z using isometric embedding.
3. Compute the surface normals N using smoothness assumption.

The goal of transient imaging is to explore the opportunities in multi-path analysis of light transport. We developed the theoretical basis for analysis and demonstrated potential methods for recovering scene properties for a range of practical, real world scenarios including guidance and control, and, rescue and planning.

9 Applications

Transient imaging will be useful in a range of applications where the inverse imaging problem is intractable with today's sensors such as back-scatter reconstruction in medical imaging through scattering medium (including in-contact medical imaging), tracking beyond line of sight in surveillance and robot path planning, segmentation by estimating material index of each surface in a single photo, micro machining and optical lithography.

With extended observable structure we will enable better layout understanding for fire and rescue personnel, and car collision avoidance at blind corners. Other applications of non line-of-sight imaging include real time shape estimation of potential threats around a corner, computing motion parameters of hidden objects, rescue and planning in high risk environments, blind assistance, inspection of industrial objects (with hidden surfaces).

Some of the theoretical methods and results we have developed may be used to augment existing LIDAR imaging for improved 3D reconstruction and better scene understanding. The theoretical and experimental tools and techniques developed for transient imaging could be extended to other domains such as acoustics, ultrasound and underwater imaging. It may be hard to identify other such application areas immediately but as transient imaging becomes a widely used tool, researchers and engineers will find new uses for its methods.

10 Conclusion

Transient imaging opens up a completely new problem domain and it requires its own set of novel solutions. Although we have demonstrated initial feasibility by proof-of-concept experiments, our research poses more questions than it answers. There are several unexplored interesting possibilities and future directions. Before transient imaging becomes a field ready tool, we require significant theoretical and experimental research solving the real world inverse geometry and reflectance problems. In order to tap the full potential of transient imaging, we need to build dedicated hardware systems including, long range, semiconductor femtosecond lasers, sub-picosecond accurate, single photon sensitive detector arrays and powerful on board analog/digital signal processing. This process involves bringing together advanced algebraic, differential geometrical models of real world scene formation, cutting edge and experimental research in applied physics and semiconductors together with state of art on board computational signal processing.

Acknowledgements 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, Biyeun Buczyk assisted in carrying out initial experiments. We finally thank Gavin Miller, Adam Smith and James Skorupski for several initial discussions.

Davis was supported in part by NSF #CCF-0746690 and Raskar was supported in part by Nokia Research and Sloan research fellowship.

References

- Arvo, J. (1993). Transfer equations in global illumination. In *Global Illumination, SIGGRAPH '93 Course Notes*.
- Campillo, A., & Shapiro, S. (1983). Picosecond streak camera fluorometry—a review. *IEEE Journal of Quantum Electronics*.
- Dattoro, J. (2006). *Convex optimization & Euclidean distance geometry*. Morrisville: Lulu.com.
- Denk, W., Strickler, J.H., & Webb, W.W. (1990). Two-photon laser scanning fluorescence microscopy. *Science*, 248, 73–76.

- Garren, D., Goldstein, J., Obuchon, D., Greene, R., & North, J. (2004). SAR image formation algorithm with multipath reflectivity estimation. In *Proceedings of the IEEE radar conference, 2004* (pp. 323–328).
- Garren, D., Sullivan, D., North, J., & Goldstein, J. (2005). Image preconditioning for a SAR image reconstruction algorithm for multipath scattering. In *Proc. of IEEE int. radar conference*
- Gonzalez-Banos, H., & Davis, J. (2004). Computing depth under ambient illumination using multi-shuttered light. *Computer Vision and Pattern Recognition*.
- Iddan, G. J., & Yahav, G. (2001). 3D imaging in the studio (and elsewhere...). In *SPIE*.
- Immell, D. S., Cohen, M. F., & Greenberg, D. P. (1986). A radiosity method for non-diffuse environments. In *ACM SIGGRAPH*.
- Itatani, J., Quéré, F., Yudin, G., Ivanov, M., Krausz, F., & Corkum, P. (2002). Attosecond streak camera. *Physical Review Letters*.
- Kajiya, J. T. (1986). The rendering equation. In *ACM SIGGRAPH*.
- Kamerman, G. (1993). Active electro-optical system. In *The infrared and electro-optical system handbook: Vol. 6. Laser radar [M]. Chapter 1*.
- Kutulakos, K. N., & Steger, E. (2007). A theory of refractive and specular 3d shape by light-path triangulation. *International Journal of Computer Vision*.
- Lange, R., & Seitz, P. (2001). Solid-state time-of-flight range camera. *IEEE Journal of Quantum Electronics*.
- Morris, N. J. W., & Kutulakos, K. N. (2007). Reconstructing the surface of inhomogeneous transparent scenes by scatter trace photography. *International Conference on Computer Vision*.
- Nayar, S. K., Ikeuchi, K., & Kanade, T. (1990). Shape from interreflections. *International Conference on Computer Vision*.
- Nayar, S. K., Krishnan, G., Grossberg, M. D., & Raskar, R. (2006). Fast separation of direct and global components of a scene using high frequency illumination. In *ACM SIGGRAPH*.
- Ng, R., Marc, L., Mathieu, B., Gene, D., Mark, H., & Pat, H. (2005). *Light field photography with a hand-held plenoptic camera*. Stanford University Computer Science Tech Report.
- Patow, G., & Pueyo, X. (2003). A survey of inverse rendering problems. *Computer Graphics Forum*.
- Ramamoorthi, R., & Hanrahan, P. (2001). A signal-processing framework for inverse rendering. *Computer Graphics and Interactive Techniques*.
- Sarunic, P., White, K., & Rutten, M. (2001). Over-the-horizon radar multipath and multisensor track fusion algorithm development.
- Schmitt, J. M. (1999). Optical coherence tomography (oct): a review. *IEEE Quantum Electronics*.
- Seitz, S. M., Matsushita, Y., & Kutulakos, K. N. (2005). A theory of inverse light transport. *IEEE International Conference on Computer Vision*.
- Sen, P., Chen, B., Garg, G., Marschner, S. R., Horowitz, M., Levoy, M., & Lensch, H. P. A. (2005). Dual photography. In *ACM SIGGRAPH*.
- Vandapel, N., Amidi, O., & Miller, J. (2004). Toward laser pulse waveform analysis for scene interpretation. *IEEE International Conference on Robotics and Automation*.
- Veeraraghavan, A., Raskar, R., Agrawal, A., Mohan, A., & Tumblin, J. (2007). Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocussing. In *ACM SIGGRAPH*.
- Raskar, R., & Davis, J. 5d time-light transport matrix: What can we reason about scene properties, Int. Memo 2007.
- Smith, A., Skorupski, J., & Davis, J. Transient rendering, UC Santa Cruz TR UCSC-SOE-08-26, Feb 2008. 2.