Decomposing Global Light Transport Using Time of Flight Imaging

Di Wu, Andreas Velten, Matthew O'Toole, Belen Masia, Amit Agrawal, Qionghai Dai & Ramesh Raskar

International Journal of Computer Vision

ISSN 0920-5691

Int J Comput Vis
DOI 10.1007/s11263-013-0668-2

Springer

Int J Comput Vis
DOI 10.1007/s11263-013-0668-2

Decomposing Global Light Transport Using Time of Flight Imaging

Di Wu · Andreas Velten · Matthew O'Toole · Belen Masia · Amit Agrawal · Qionghai Dai · Ramesh Raskar

Received: 1 February 2013 / Accepted: 7 October 2013

Abstract Global light transport is composed of direct and indirect components. In this paper, we take the first steps toward analyzing light transport using the high temporal resolution information of time of flight (ToF) images. With pulsed scene illumination, the time profile at each pixel of these images separates different illumination components by their finite travel time and encodes complex interactions between the incident light and the scene geometry with spatially-varying material properties. We exploit the time profile to decompose light transport into its constituent direct, subsurface scattering, and interreflection components. We show that the time profile is well modelled using a Gaussian function for the direct and interreflection components, and a decaying exponential function for the subsurface scattering component. We use our direct, subsurface scattering, and interreflection separation algorithm for five computer vision applications: recovering projective depth maps, identifying subsurface scattering objects, measuring parameters of analytical subsurface scattering models, performing edge detection using ToF images and rendering novel images of the captured scene with adjusted amounts of subsurface scattering.

Keywords Light transport analysis · Direct/global separation · Time of flight imaging · Transient imaging · Femto-photography

Electronic supplementary material The online version of this article (doi:10.1007/s11263-013-0668-2) contains supplementary material, which is available to authorized users.

D. Wu · M. O'Toole · B. Masia · R. Raskar
MIT Media Lab, Room E14-474G, 75 Amherst St, Cambridge, MA02139, USA
e-mail: raskar@media.mit.edu

D. Wu
Tsinghua University, Beijing, China

A. Velten (✉)
Medical Engineering, Morgridge Institute for Research, 330 North Orchard Street, Madison, WI 53715, USA
e-mail: velten@gmail.com

A. Velten
Laboratory for Optical and Computational Instrumentation, University of Wisconsin at Madison, 1675 Observatory Drive, Madison, WI 53706, USA

M. O'Toole
University of Toronto, Toronto, ON, Canada

B. Masia
Universidad de Zaragoza, Saragossa, Spain

A. Agrawal
Mitsubishi Electric Research Labs, Cambridge, MA, USA

Q. Dai
Department of Automation, Tsinghua University, Room 410, Central Main Building, Beijing, China

1 Introduction

Light transport analysis is a lasting and challenging area in computer vision and graphics. Existing methods for light transport analysis usually rely on structured light techniques that use active spatially and/or angularly varying illumination patterns. They enable a broad range of applications, including material acquisition (Holroyd et al. 2010; Mukaigawa et al. 2010), light transport separation (Nayar et al. 2006; Seitz et al. 2005), and robust geometry acquisition (Chen et al. 2007; Gupta et al. 2011, 2009; Zhang and Nayar 2006). On the other hand, time of flight (ToF) imaging is traditionally limited to geometry acquisition. Recently, Velten et al. introduced *femto-photography*, a novel imaging technique with an effective resolution of about one trillion frames per second (Velten et al. 2012, 2013). The proposed novel combination of

Published online: 31 October 2013

re-purposed hardware and computational photography techniques allows to visualize the propagation of light through table-top scenes, including phenomena such as scattering, caustics, diffraction, or second and third order indirect lighting. This femto-second photography technique has already spawned new applications in computer vision and graphics, such as single-view BRDF capture (Naik et al. 2011), or reconstructing hidden geometry (Velten et al. 2012).

In this paper, we propose a separation method based on ToF imaging. Our approach is new in two aspects. First, we utilize ToF images to measure 3D light transport; second, we use the time profile of each pixel to decompose different light transport effects. The key insight of this research is to exploit pulsed scene illumination along with ultra-high time resolution in light transport by differentiating light transport effects with different time profiles. The time models of different light effects are analyzed and verified in this paper.

Contributions We show how to decompose ToF videos into direct, subsurface scattering, and interreflection components by the analysis of time profiles and apply it to the following problems:

(i) recovering projective depth from the direct component in the presence of global scattering;
(ii) identifying and labelling different types of global illumination effects;
(iii) measuring parameters of subsurface scattering materials from a single point of view;
(iv) performing edge detection using spatial-temporal ToF information;
(v) adjusting subsurface scattering in rendered novel images of the scene.

The paper is organized as follows: While Sect. 2 discusses related work, Sect. 3 introduces ToF images, the image formation model, the theory of light transport involved and the algorithms for light component separation. Section 4 presents a set of techniques by which a number of computer vision applications benefit from the prior light component separation. In Sect. 5 the datasets and scenes used to test the proposed algorithms are described. Finally, Sect. 6 shows results for the algorithms presented, both for light component separation and for its applications, and Sect. 7 summarizes and concludes the paper.

2 Related Work

The process of light transport forming an image of a scene is modelled by the rendering equation (Kajiya 1986). In rendering applications, ToF is generally ignored, because conventional slow sensors are only able to detect the time independent, steady state light transport. The addition of time to the theoretical model has been proposed by Arvo (1993). A frequency domain analysis of time dependent light transport has been provided by Wu et al. (2012). The process of obtaining information about light transport by analysis of a captured scene image is known as inverse rendering.

Global illumination effects in light transport, including interreflections, caustics, and subsurface scattering, are important visual features of real-world scenes. Structured light methods utilize global light transport when decomposing multi-bounce light transport into individual bounces (Bai et al. 2010; Seitz et al. 2005), extracting geometric information from second-bounce light transport (Liu et al. 2010), and compensating for global illumination (Ng et al. 2012; Wetzstein and Bimber 2007). A method for separating high-frequency direct transport from low-frequency global transport requires as little as a single photo under structured lighting (Nayar et al. 2006). Spatially varying illumination can also be used to separate light transport into a long-range (diffuse interreflection) and near-range (subsurface scattering) component (Reddy et al. 2012). O'Toole et al. (2012) introduced primal-dual coding to sample specific light paths from a scene using coded illumination and exposure.

These direct/global separation techniques find uses in range acquisition applications for robustly capturing scene depth information in the presence of global illumination (Chen et al. 2007; Gupta et al. 2009; Zhang and Nayar 2006). Specifically designed structured illumination patterns also allow for robust geometry acquisition that accounts for both interreflections and subsurface scattering light transport effects within a scene (Gupta et al. 2011). A related application is the ability to see a hidden scene around a corner using structured illumination of the hidden scene to provide light transport information (Sen et al. 2005).

ToF imaging methods provide another means of analyzing global transport effects. ToF methods used in incoherent LiDAR and gated viewing systems measure the flight time of the direct component after pulse illumination of the scene to obtain a depth measurement (Gleckler and Gelbart 2000; Brooker 2009; Busck and Heiselberg 2004). LiDAR systems are commonly employed in large outdoor scenes and operate on timescales of nanoseconds. Indirect light components are suppressed or discarded as noise by these systems. Other methods exploit phase coherence between detected light and scene illumination to detect ToF. Among these methods are coherent LiDAR (Xia and Zhang 2009), Light in Fight Holography (Abramson 1978), Optical Coherence Tomography (Huang et al. 1991), and White Light Interferometry (Wyant 2002). Unfortunately, coherence between illumination and detection light is preserved only for the direct light component. Light loses coherence when undergoing diffuse reflections and therefore global illumination

components do not preserve a constant phase relation with the illumination over sufficiently large portions of the light field. This limits coherence based detection schemes to direct illumination.

The recent femto-photography technique captures data at picosecond resolution, including direct and global illumination components (Velten et al. 2012, 2013). In their work, the authors reconstruct videos of the propagation of femtosecond light pulses through different macroscopic scenes. Additionally, they introduce the concept of *time-unwarping* to account for the effects of the finite speed of light in the sensor- and the world-reference systems. Femto-photography techniques have already been used to reconstruct the geometry of occluded objects *around the corner* through the analysis of the arrival time of multiple-bounce light transport (Pandharkar et al. 2011; Velten et al. 2012; Gupta et al. 2012), to allow for in-the-wild BRDF acquisition without any encircling equipment (Naik et al. 2011), or to synthesize faithful representations of light transport including relativistic effects (Jarabo et al. 2013). Follow-up work has introduced the use of photonic mixer devices (PMDs), although the achieved temporal resolution and sensitivity is not as high (Heide et al. 2013). In this work we show how to analyze the captured data, decomposing global illumination into its different components of direct, indirect and subsurface scattering.

3 Light Transport Analysis Using ToF Imaging

Light traveling through a real-world scene may undergo a complex array of interactions with objects that have different radiometric properties. For example, partially translucent objects (including tomatoes, wax, and skin) have subsurface scattering properties, causing light to penetrate their surface, scatter beneath the surface, and exit the surface at a different position. Geometric concavities produce interreflections where light might bounce multiple times across different reflecting surfaces within the concavity. In general, there are many classifications for the optical phenomena that occur in real-world scenes, including specular reflections for mirror-like objects, volumetric scattering due to participating media, and caustics from light focused by reflective or refractive objects.

Our objective is to separate an input image P into its constituent light transport elements. Specifically, our aim is to reduce an image to a sum of multiple images, each image containing the radiance from the scene due to a specific subset of light paths. In this work, we represent every image as the sum of three components:

$$P = D + S + I \tag{1}$$

where the direct component D consists of all one-bounce light paths, the subsurface scattering component S contains short-range scattering through the surface of translucent materials, and the interreflection component I represents longer-range scattering through the scene.

The key observation is that each light transport components can be uniquely identified using ToF imaging. In this work, we represent each ToF image P as a volume, as illustrated in Fig. 1a. The pixel value $P(x, y, t)$ corresponds to the radiance received by pixel (x, y) at a time t, where the temporal resolution is on the order of picoseconds (10^{-12} of a second). As explained in detail in Velten et al. (2013), the ToF image can be visualized as a sequence of frames as shown by Fig. 1c, d, each frame representing an image of the scene observed at a different instance in time.

Decomposing light transport using ToF imaging is made possible because each light transport component behaves differently as a function of time. We define a *time profile* $P(t)$ as a non-negative vector representing the radiance observed at a fixed pixel (x, y) as a function of time t. The radiance measured by the time profile is in response to a light pulse turned on at a time $t = 0$, and received by the camera pixel at a later time. As a result of Eq. 1, this time profile can be decomposed into a sum of individual time profile components, each representing the radiance due to a different set of light paths.

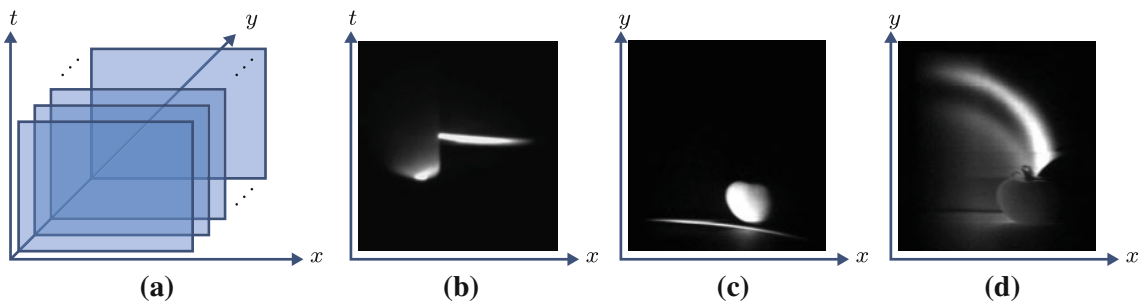


Fig. 1 (**a**) The ToF image format, as described in Hamamatsu (2012). (**b**) Example of one $x - t$ image, where $y = 121$. (**c**) $x - y$ image of the tomato screen scene (vi) (see Sect. 5 for details) when the direct component appears on the tomato at time $t_1 = 196$. (**d**) $x - y$ image when the direct component appears on the back wall at time $t_2 = 301$

3.1 Characteristics of Time Profile Component

The unique characteristics of each time profile component is a result of the distances covered by the corresponding light paths through a scene, and the fixed speed of light that relates distances to time.

Among all light paths that arrive at a particular camera pixel (x, y) at different times t, the path consisting of a single bounce travels the shortest distance amongst all possible light paths. This produces two distinct traits within the corresponding time profile component $D(t)$. First, the time profile consists of a single non-zero value $D(t_0)$ representing the direct component, where t_0 corresponds to the distance travelled by the light path (assuming the speed of light is fixed, i.e., there is a single medium). Second, because the direct light path travels the shortest distance amongst all possible light paths, any other radiance observed from the scene must arrive after t_0.

The subsurface scattering time profile $S(t)$ has an initial intensity value α at the time t_0 when light enters the surface, and decreases in value as light escapes from the surface. Subsurface scattering is physically described by the radiative transfer equation (Wang and Wu 2007), which in the case of anisotropic scattering, and low absorption can be approximated as the diffusion equation:

$$\frac{\partial \Phi(\mathbf{r}, t)}{\upsilon \partial t} + \mu_a \Phi(\mathbf{r}, t) - \mathcal{D}\nabla^2 \Phi(\mathbf{r}, t) = \boldsymbol{\mathcal{S}}(\mathbf{r}, t) \tag{2}$$

where Φ is the flux density, $\boldsymbol{\mathcal{S}}$ is the light source, the absorption coefficient μ_a and the diffusion constant $\mathcal{D}$ describes the scattering properties of the material. The speed of light is denoted by υ to avoid confusion with other uses of the letter c in different sections. We assume $\Phi(\mathbf{r}, t)$ can be expressed as the product of a time dependent and a space dependent component $R(\mathbf{r})S(t)$. After the initial illumination pulse has passed ($\boldsymbol{\mathcal{S}}(\mathbf{r}, t) = 0$), Eq. 2 can be written as

$$\frac{1}{\upsilon}\frac{\left(\frac{\partial S(t)}{\partial t}\right)}{S(t)} + \mu_a = \mathcal{D}\frac{\nabla^2 R(\mathbf{r})}{R(\mathbf{r})}. \tag{3}$$

This type of equation is referred to as a Sturm-Liouville problem in the literature. Its left hand side is clearly independent of $\mathbf{r}$, the right hand side is independent of time t. Since both sides are equal for all $\mathbf{r}$ and t the terms on both sides can depend on neither $\mathbf{r}$ nor t and have to be constant. We introduce two constants k_r and k_t to separate both equations.

$$\mathcal{D}\nabla^2 R(\mathbf{r}) = k_r R(\mathbf{r}) \quad \text{and} \quad \frac{\partial S(t)}{\upsilon \partial t} + \mu_a S(t) = k_t S(t) \tag{4}$$

The time dependent part has the following solution:

$$S(t) = \alpha e^{-\upsilon(k_t - \mu_a)(t - t_0)} \tag{5}$$

After the illumination pulse has passed, the light distribution in the sample will therefore decay approximately exponentially.

In general, subsurface scattering is very complex and depends on several factors such as scene geometry, lighting conditions, and material properties. The assumption of low absorption underlying our derivation specifically means that, in its path through the material, a light particle is much more likely to be scattered than to be absorbed. It will thus likely undergo many scattering events before absorption occurs. This is the case for many materials that provide enough subsurface scattering to become relevant in a computer vision or computer graphics context, i. e. in macroscopic scenes with usually moderate time resolutions. For example we can assume low absorption for the tomato in our scenes, and for marble or skin.

Interreflection is the result of the reflected light from one object illuminating another object. For example, a scene consisting of as little as two points may produce interreflected light, where light may bounce between both points multiple times before being observed. The corresponding time profile component $I(t)$ for such a scene is an impulse train. Note that the intensity of a k bounce light path decreases exponentially as a function of k. Note also, that subsurface scattering can be considered a set of interreflections between very small surfaces that are very close together (i.e. the scattering centers inside a medium). It is thus to be expected, that for decreasing path length between surfaces (for example in corners), time responses of interreflections become similar to subsurface scattering. Similarly, the overall light content of a complex scene is expected to decay approximately exponentially when viewed with a low time resolution.

3.2 Image Formation Model

The image formation model of ToF images presents unique challenges when attempting to decompose real time profiles into their constituent parts. An illustration of the image formation model in Fig. 2 highlights the sources of two such challenges: temporal resolution and dynamic range.

The measured time profile is a convolution of the original time profile with the camera's impulse response function. The impulse response function represents the camera's finite time resolution and is the result of several physical attributes of the system. First, the time profile characteristics discussed in the previous section assume that the incident light pulse has negligible width. Second, the light pulse is assumed to fire at a time 0, though this depends on the stability of the synchronization between camera and light pulse. Third, the impulse response may also depend on the aperture of the system. From measured data of the direct component within a scene, we observe that our camera impulse response function is well-approximated by a Gaussian, as illustrated in Fig. 3.

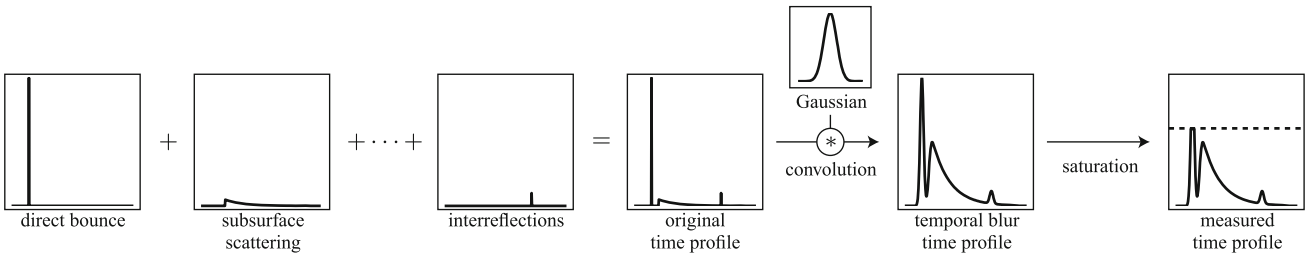


Fig. 2 Flowchart illustrating the image formation model for our ToF camera. Light interacts with a scene in any number of ways (i.e. direct reflection, subsurface scattering, interreflections) to produce a time profile. The temporal resolution power of the camera depends on the blur kernel for the camera. The final measured signal is quantized. As a result of the very high dynamic range of ToF images, most of the measured time profiles from our datasets are oversaturated by the light from the direct bounce. The objective of this work is to recover the individual components of light transport from this measured time profile

The dynamic range of ToF images is high relative to images captured using conventional photography. The intensity of the direct light paths is concentrated to a specific time within the time profile, whereas the intensity associated with global light paths is distributed over time. As a result of this large difference between bright and dark areas of the time profile, ToF images often contain both over-exposed and under-exposed regions. In general, to obtain a good signal in low intensity parts of the ToF image, we allow some saturation in the images (e.g. on specular reflections). If, instead, the gain of the camera is reduced to avoid saturation in those spots, the rest of the image becomes darker, and much more image information may be lost due to low signal-to-noise ratio.

There is an inherent trade-off between temporal resolution and dynamic range of a ToF imaging system. Increasing the temporal resolution of the system also increases the dynamic range of the ToF images, causing issues such as over-saturation. A camera impulse response function that blurs the time profile, effectively reducing the temporal resolution, also decreases the dynamic range of the time profile.

3.3 Direct and Global Separation

We have analyzed and modelled attributes of several constituent light transport components. The first intuition to separate them is to fit them to the model proposed above. In this section, we first propose a model fitting decomposition algorithm, and show its limitations in the presence of image distortions such as saturation. Further, we propose an interpolation based decomposition algorithm that locates the direct component and interpolates the global component from the coupled time profile by incorporating physically valid constraints.

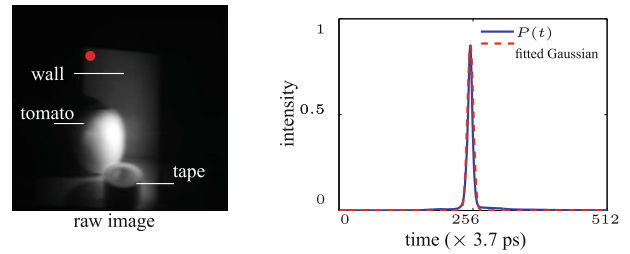
To validate our algorithm, we modify the graphics renderer Physically Based Ray Tracer (PBRT) to generate time-profiles for multiple bounces (please refer to Sect. 5 for more details about the simulated dataset).

3.3.1 Model Fitting

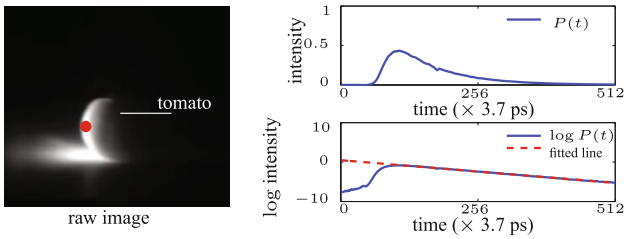
As discussed above, direct light transport is a Gaussian impulse, subsurface scattering can be modelled by an exponential decaying function, and interreflections from reflective surfaces can also be represented by a Gaussian impulse. Multiple closely spaced interreflections from diffuse surfaces form a decaying time profile which can be approximated by an exponential function as well. Given these models of light transport effects, the intuitive method is to fit the time profile $P(t)$ to the mixture of models: $e^{\frac{(t-t_1)^2}{2\sigma^2}}$ for the direct component and $e^{-\alpha(t-t_2)}$ for the global component, where t_1 and t_2 encode the temporal location of the signal. By minimizing Eq. 6, we obtain the model parameters $(\lambda_1, \lambda_2, \sigma, \alpha, t_1, t_2)$. In the cases when the system's temporal impulse response function width σ is known, we only need to optimize for the remaining 5 parameters.

$$\min \| P(t) - \lambda_1 e^{\frac{t^2}{2\sigma^2}} * \delta(t - t_1) - \lambda_2 (e^{\frac{t^2}{2\sigma^2}} * e^{-\alpha(t-t_2)}) u(t - t_2) \|_2 \quad (6)$$

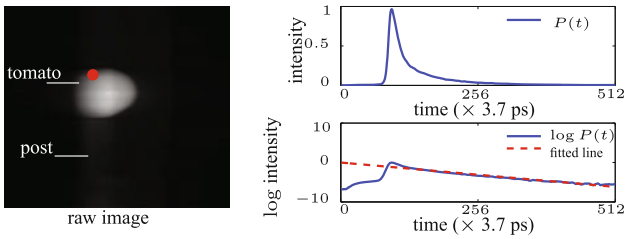
In Fig. 4a, we show the original time profile of a point in the simulation data (please refer to Sect. 5 for more details) in cyan. In the rendering, the temporal impulse response function is a perfect delta function. We simulate the real captured time profile by convolving the rendered data with the impulse response function of the capture system, which then gives an approximation of the captured $P(t)$, $D(t)$ and $G(t)$ as shown in black, red and blue respectively. The global component of the rendered time profile is decaying as an exponential, which demonstrates our model for diffuse interreflections. The model fitting decomposition result is shown in Fig. 4b. The parameter σ is known in this case. Although the direct and global components are deeply coupled due to short range effects, the fitting well approximates the ground truth. We optimize Eq. 6 using `lsqcurvefit` in MATLAB. Although convergence to a global minimum is not guaran-



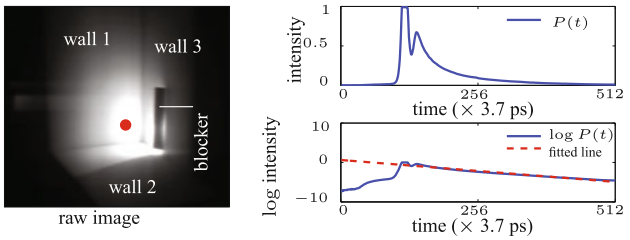
(a) tomato tape scene



(b) back-lit tomato scene



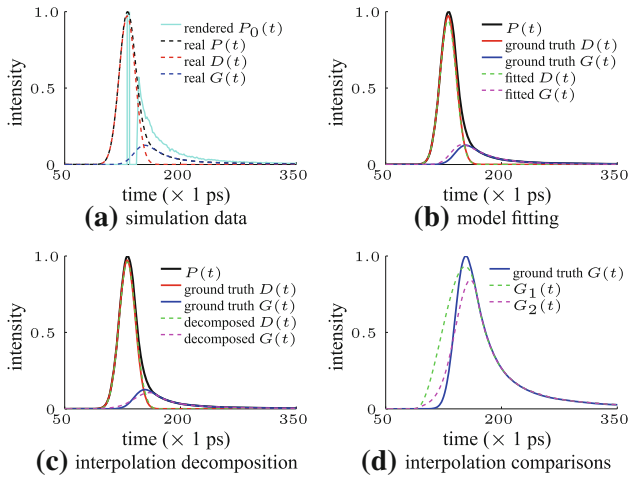
(c) mounted tomato scene



(d) corner scene

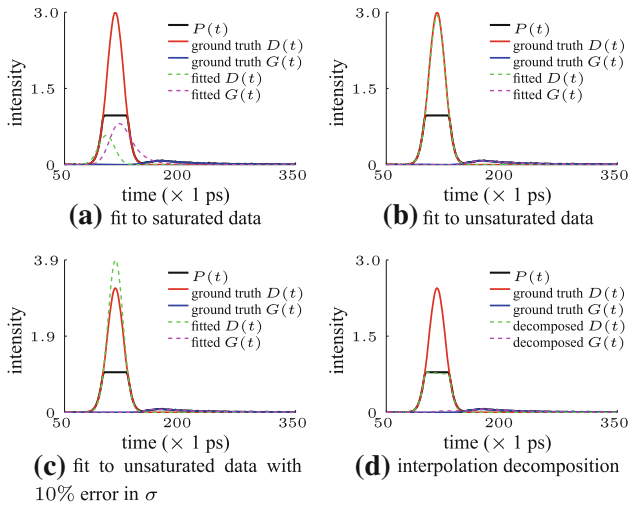
Fig. 3 (**a**) *Tomato tape scene:* A Gaussian closely approximates the time profile of a point containing only the direct component. (**b**) *Backlit tomato scene*: The time profile of a point lit by subsurface scattered light decays exponentially over time. A line closely approximates the log of the time profile. (**c**) *Mounted tomato scene*: A time profile containing both direct and subsurface scattering light transport is a sum of a Gaussian and an exponentially decaying function. (**d**) *Corner scene*: A time profile contains both direct and complex diffuse interreflection light transport due to light trapped in the corner

teed, the fitting does work well in practice provided a good initial estimate is given. We use the time instant of the maximum of the profile as an initial guess for t_1, as well as for σ; for the rest, $(t_2, \alpha, \lambda_1, \lambda_2) = (5, 0.01, 1, 0.1)$ worked well as initial estimates for all cases tested in the absence of jitter noise and saturation (see Figs. 4b, 5b).



(a) simulation data **(b)** model fitting **(c)** interpolation decomposition **(d)** interpolation comparisons

Fig. 4 Time profile of point 1 in simulation data (Fig. 13) and its decompositions. Please note that the intensity of direct impulse in the *cyan curve* in **a** is 8.9, we show the intensity range of [0,1] for the visibility of $P(t)$, $D(t)$ and $G(t)$



(a) fit to saturated data **(b)** fit to unsaturated data **(c)** fit to unsaturated data with 10% error in σ **(d)** interpolation decomposition

Fig. 5 Simulation data with noise and saturation and their decompositions. The time profile corresponds to point 2 in simulated corner scene (Fig. 13). (**a**–**c**) use the separation method described in Sect. 3.3.1, (**d**) uses the method described in Sect. 3.3.2

Limitations The model fitting works well in an experimental best case scenario (Fig. 4b), when the parameter σ is known and the data is not distorted by effects other than noise and a finite impulse response. However, in real data, the captured ToF information might suffer from jitter or saturation problems, as discussed in Fig. 2. Jitter could cause σ to change intermittently. Saturation is a more fundamental problem: Because ToF illumination is localized in space, as well as in time, the dynamic range inherent to the data is larger than in a steady state image, resulting in saturation, as discussed in Sect. 3.2.

The performance of our model fitting method decreases in the presence of variations in σ and saturation. Figure 5 pro-

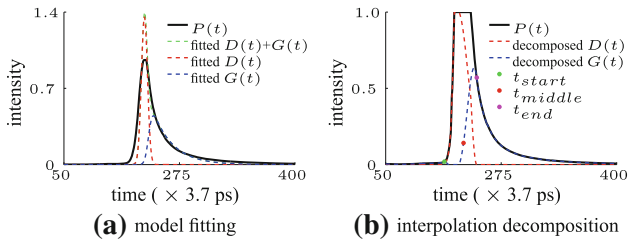


Fig. 6 Comparison of direct and global decomposition on real data, the time profile corresponds to point 3 in the tomato tape scene (Fig. 13). (**a**) The model fitting result underestimates the tail of the global component. (**b**) The interpolation-based algorithm preserves the details

vides the model fitting results on simulated data with noise and saturation. If we saturate our simulated data, our algorithm performs as shown in Fig. 5a. Figure 5c shows algorithm performance when the assumed σ deviates from the one used when generating the data.

Further, the global light transport in a real world scene is more complicated than pure subsurface scattering or interreflections. Figure 6a shows the model fitting result of a point in the tomato tape scene (v). This point receives interreflections from the floor and the tape, which further cause subsurface scattering inside the tomato, and this process goes on till the intensity has diminished to the detection limit. The fitting of the global part fails to represent the tail of the time profile with a single Gaussian convolved exponential decaying function. The fitting algorithm works well for simulated data with the exception of cross-combinations of light transport categories, such as subsurface scattering created by multibounce light and multibounce components created by subsurface scattering illumination.

3.3.2 Interpolation Based Decomposition

Saturation is common in real data and the lowering of camera gain to prevent saturation can destroy information in the ToF image. It is therefore desirable to develop a method that works well in the presence of saturation. Unfortunately, saturation imposes a very non-physical and highly non-linear artifact on the data that is hard to grasp in an elegant model. We can however find points of attack by investigating the structure of our data, and by noting that while a saturated direct component is unphysical, the underlying global component obeys physical constraints regarding rise and smoothness of the signal.

The time profiles of basic types of light transport effects offer important understanding of the constituent signals and how they are coupled together in the time dimension. To decompose the direct from the global component, we need to decouple the two in their overlapped time region.

We propose an interpolation based decomposition method. The algorithm separates direct and global time profiles by localizing the direct component within the time profile $P(t)$, and extracting the direct component $D(t)$ by imposing smoothness constraints on the global component $G(t)$. Note that some pixels within a ToF image may receive no direct illumination. Multibounce and subsurface scattering components are usually much weaker in intensity than the direct component, even when accounting for different material reflectances. We thus reject time profiles without a direct component by finding the first intensity peak in the time profile of each pixel for simple datasets. However, to get more accurate and robust direct light illuminated regions, we exploit a simple algorithm as described in Sect. 6.

The key assumptions of the algorithm are: (1) the first impulse observed in the time profile $P(t)$ is the direct component, (2) the global component is smooth during the time range of the direct component, (3) the decomposition obeys physically-based properties such as energy conservation.

Algorithm 1 Direct/global separation

Require: the time profile $P(t)$

1. Compute the derivative of the time profile $P'(t)$
2. Detect position of the first peak t_{peak} in the time profile by finding the local maximum; if the data is saturated, we set t_{peak} to the center of the saturated region, and this generates t_{middle}
3. The last point in time that satisfies $P(t) < \gamma P(t_{middle})$ (the typical value for γ is 0.01) determines t_{start}
4. Compute the end of the direct component $t_{end} = t_{start} + 2(t_{middle} - t_{start})$
5. Smoothly interpolate time profile values between t_{start} and t_{end} using C^1 and C^2 continuity to generate the global component $G(t)$
6. Extract direct component $D(t) = P(t) - G(t)$

In Algorithm 1, steps 1 and 2 localize the arrival time t_{start} of the direct component. Step 3 finds the apex of the direct component at time t_{middle}. Step 4 returns the departure time t_{end} of the direct component, assuming the apex occurs at the center of the direct impulse. The profile values $P(t)$ for $t \notin [t_{start}, t_{end}]$ are global component values; step 5 predicts the unknown global values $G(t)$ for $t \in [t_{start}, t_{end}]$ through interpolation. In the following, we detail the interpolation algorithm to get the global component values.

Interpolation Algorithm The method of interpolation of $G(t)$ for $t \in [t_{start}, t_{end}]$ determines the way direct and global components are decoupled. Direct interpolation based on the start and end points as well as their C^1 and C^2 continuities, gives results as the green curve in Fig. 4d, which is not physically valid as the separated global component rises earlier than the direct.

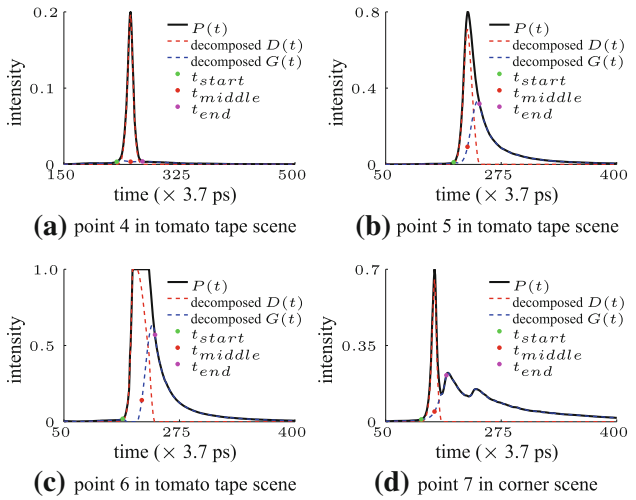
In order to impose physically-based constraints in the interpolation, we add an optimized control point at t_{middle}. The value of the control point is obtained by searching in $[0, 1]$ for the value that minimizes the penalty function for

the interpolated $D(t)$ and $G(t)$ describing the physical constraints as follows:

$$f = \sum_{i=1}^{M} \|\lambda_i \sum_{t_{start}}^{t_{end}} f_i\|_0 + \sum_{i=M+1}^{N} \|\lambda_i \sum_{t_{start}}^{t_{middle}} f_i\|_0 \quad (7)$$

where $M = 4, N = 7, f_1 : D(t) < 0$ and $f_2 : G(t) < 0$ represent the requirement of non-negativity for both $D(t)$ and $G(t)$, $f_3 : P(t) < D(t)$ and $f_4 : P(t) < G(t)$ represent the energy conservation constraint, $f_5 : D'(t) < G'(t)$, $f_6 : D'(t) < 0$ and $f_7 : G'(t) < 0$ require the direct and global to both be rising and $D(t)$ be rising faster than $G(t)$ before $D(t)$ reaches its peak. We use MATLAB's `fminsearch` to minimize Eq. 7 and obtain t_{middle}, and cubic spline interpolation to ensure C^1 and C^2 continuities. The typical values for λ_1-λ_M are 10^5, and for λ_M-λ_N they are 10^7; these values are used in all examples and scenes shown in this paper. The optimized interpolation result for the global component compared with interpolation only based on continuity constraints is shown in Fig. 4d, the former is in better agreement with the ground truth. We therefore use this optimized control point based interpolation step in step 5 of our decomposing algorithm. Lastly, step 6 subtracts the global component from the time profile to produce the direct component, as in Figs. 5d and 7.

The interpolation method is more robust to noise and saturation than the model-based fitting. Model fitting does not always work well in the presence of noisy or saturated data, or when σ is unknown and cannot be well estimated, as shown in Fig. 5a–c (and discussed in Sect. 3.3.1). Meanwhile, the interpolation method (Fig. 5d) can handle this data. Further, using interpolation solves another stated limitation of the model-based fitting, which is not being able to handle cases with complex global components such as those originating from intricate interactions of subsurface scattering and interreflections (see Fig. 6). We also apply our interpolation-based direct and global separation algorithm to points with different light transport properties, as in Fig. 7. The result for a point with only a direct component is shown in Fig. 7a, dominant subsurface scattering effects in the global part without saturation in Fig. 7b, and with serious saturation in Fig. 7c, as well as a point in the corner with complex short range diffuse interreflections in Fig. 7d. In all the above cases, our algorithm achieves reasonable and robust results.



(a) point 4 in tomato tape scene **(b)** point 5 in tomato tape scene

(c) point 6 in tomato tape scene **(d)** point 7 in corner scene

Fig. 7 Direct and global decomposition results on real data. The illustrated time profiles represent: (**a**) the direct component, (**b**) direct and subsurface scattering (without saturation), (**c**) direct and subsurface scattering (with saturation), and (**d**) direct and complex interreflection in the corner. Time profiles correspond to points 4–7 in Fig. 13

3.4 Subsurface Scattering and Interreflection

The ToF images allow to further separate the global component. Our interreflection/subsurface scattering separation algorithm takes as input the global time profile $G(t)$ and the departure time of the direct component t_{direct}. Our separation algorithm localizes and extracts the interreflection components from the global time profile.

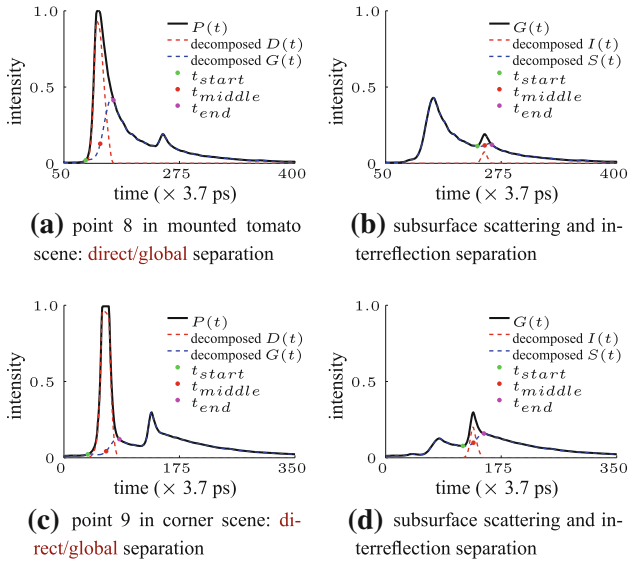
Algorithm 2 Interreflection/subsurface scattering separation

Require: global time profile $G(t)$ and end of direct component t_{direct}

1. Compute the derivative of the time profile $G'(t)$
2. Find the start of the interreflection component by solving $t_{start} = \arg\min_t G'(t) > \alpha_G$ subject to $t > t_{direct}$, for some tolerance α_G (typical value of α_G: $0.05 \max_t P'(t)$)
3. Find the apex of the interreflection component by solving $t_{middle} = \arg\min_t G'(t) < \beta_G$ subject to $t > t_{start}$, for some tolerance β_G (typical value of β_G: 0.0001)
4. Compute the end of the interreflection component $t_{end} = t_{start} + 2(t_{middle} - t_{start})$
5. Smoothly interpolate time profile values between t_{start} and t_{end} using C^1 and C^2 continuity to generate the subsurface scattering component $S(t)$
6. Extract interreflection component $I(t) = G(t) - S(t)$

Algorithm 2 finds the start of a new impulse, identifies the apex of the impulse, computes the end of the impulse, and interpolates time profile values between the start and end points. Step 5 is using the same interpolation method as described in the direct and global separation part. The right column of Fig. 8 decomposes the global time profiles from the left column of Fig. 8 into their interreflections and subsurface scattering components.

Both algorithms rely on the ability to detect individual impulses. High-order bounces have diminished intensity, and can be wider than the first bounce component, for example due to partial shallow penetration of the reflecting surface. As a result, we only demonstrate the separation of second-bounce interreflection from the global time profile.



(a) point 8 in mounted tomato scene: direct/global separation

(b) subsurface scattering and interreflection separation

(c) point 9 in corner scene: direct/global separation

(d) subsurface scattering and interreflection separation

Fig. 8 The decomposition of global components. *Each row* corresponds to a two step separation of a time profile. Time profiles correspond respectively to points 8–9 in Fig. 13. *Left column* Algorithm 1 decomposes the time profiles into their direct and global components. *Right column* Algorithm 2 further decomposes the global time profiles into the interreflections and subsurface scattering components

4 Exploiting Light Transport Components

Decomposing a single ToF image into its direct, subsurface scattering, and interreflection components has many applications in computer vision. Our paper focuses on five such applications: recovering projective depth from the direct component, labelling image regions to identify subsurface scattering objects, measuring scattering coefficients, edge detection and generating novel images.

Projective Depth from ToF Monostatic LiDAR systems use a co-located light source and sensor to recover geometry using ToF. These devices capture depth by measuring the time elapsed between emitting a pulse of light onto a point in a scene, and the backscattered light returning to the sensor. LiDAR systems typically sweep the light across an area in the scene to recover a full depth map. Our ToF imaging device for capturing depth in a single image is similar to a bistatic LiDAR system (Repasi et al. 2009), in that it requires no co-location between sensor and light source.

In a camera system, image pixels (x_{pix}, y_{pix}) are related to 3D coordinates $W = (w_x, w_y, w_z)$ by

$$x_{pix} = \alpha_x \frac{w_x}{w_z} + x_0 \tag{8}$$

$$y_{pix} = \alpha_y \frac{w_y}{w_z} + y_0 \tag{9}$$

where (x_0, y_0) is the image center, $\alpha_x = fk_x, \alpha_y = fk_y$, f is the depth of the image plane (focal length), and k_x and k_y are scaling factors in the x and y directions respectively. Each camera pixel (x_{pix}, y_{pix}) thus maps to a ray of points W_{pix} in 3D space. ToF can provide the additional constraint of light path length, that allows us to locate a reflecting object in three dimensions. To provide this constraint, the laser illuminates an unknown point in the scene and produces a virtual light source $L = (L_x, L_y, L_z)$ (see Sect. 5). This virtual light source illuminates the scene at multiple scene points W. The time when the laser pulse reaches L is denoted as t_0. The camera is centered at point $C = (0, 0, 0)$. The first light seen by camera pixel (x_{pix}, y_{pix}) must be the light traveling directly from L to one of the associated W_{pix} and must satisfy the following equation (see Fig. 9):

$$\|W - L\|_2 + \|W - C\|_2 = \upsilon(t - t_0) \tag{10}$$

where υ is the speed of light. For each camera pixel (x_{pix}, y_{pix}), the intersection of the ellipsoid in Eq. 10 with the ray of W_{pix} defined by Eqs. 8 and 9 produces a unique depth map value, for the camera pixel. In practice, we need to calibrate the system to recover L, t_0, α. Once these constants are known, our depth recovery algorithm solves for the unknown depth value given a time t for each pixel. Results for projective depth recovery using this technique are shown in Sect. 6.

Labelling Subsurface Scattering Regions As shown in Fig. 3b, the time profile for subsurface and volumetric scattering objects after receiving an initial light pulse has a unique appearance. When light hits an object in a scene, either transmission, reflection, or absorption occurs. As we discussed in Sect. 3.1, certain objects with subsurface scattering properties will scatter the transmitted light internally and slowly release the light over time. Rather then performing a complete separation into direct and global components, it is therefore possible to employ a much simpler metric, that quantifies how long light stays at a scene point after an initial illumination pulse.

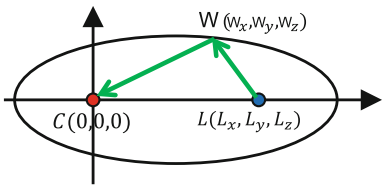


Fig. 9 The locus of scene points W from which light has traveled a constant distance (corresponding to a time t) from the source L to the camera C is an ellipsoid in 3D space defined by Eq. 10; the corresponding ellipse in 2D is depicted here

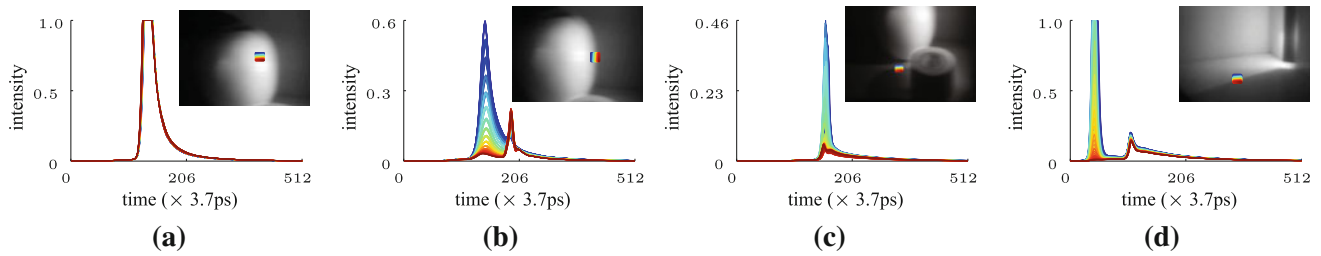


Fig. 10 The time profiles correspond to points in the following patches: (**a**) patch on tomato, (**b**) patch on perimeter of tomato, (**c**) patch on edge of shadow, (**d**) patch crossing another shadow edge

Our algorithm labels subsurface scattering regions by evaluating the following equation for each pixel (x, y):

$$T \geq Ratio(x, y) = \frac{\int_{t=0}^{T} P(t)_{x,y}}{max_{t=0}^{T} P(t)_{x,y}} \geq 1 \tag{11}$$

When the function returns a value near 1, the energy concentrates to a single instant in time within the time profile, indicating the corresponding point has no subsurface scattering properties. A larger $Ratio(x, y)$ value corresponds to distributed energy within a time profile, a characteristic of subsurface scattering points. Examples of the performance of our labelling algorithm are included in Sect. 6.

Measuring Scattering Model Parameters An important application of ToF images is material acquisition, and in particular, modelling volumetric scattering, for example from fog or other scatterers distributed across the scene volume. When a short pulse of light illuminates a volumetric scattering medium, the time profile shows an exponential decay in the measured irradiance that lasts from a few picoseconds to several nanoseconds. The time profile provides sufficient information to determine the scattering coefficient of a homogeneous medium.

Our model for volumetric scattering mediums is an exponential $\alpha e^{\gamma t} = e^{\log(\alpha)+\gamma t}$, where γ is the scattering coefficient and α is the incident illumination. Our algorithm for capturing the scattering coefficient requires two steps. First, the algorithm computes $\log(P(t))$. Second, a line of the form $\log(\alpha) + \gamma t$ is fit to the logarithmic time profile $\log(P(t))$ to produce the scattering coefficient. Given the scattering coefficient, the time profile of the volumetric scattering medium is a function of the incident illumination.

Edge Detection Using ToF The local spatial properties of the time profile can improve edge detection methods. Using conventional steady state images, it is difficult to distinguish between a depth edge and a shadow edge. The time profile of a depth edge has two distinct direct time values, whereas shadow edges have at most one distinct peak. By analyzing the time profile of a set of pixels, our algorithms distinguish between the two edges.

To identify a depth edge, we analyze the time profiles for points within a 7×7 sliding window, as illustrated in Fig. 10. For each time profile within the window, our algorithm computes the *position* and *intensity* of the maximum, and evaluates the variance of these quantities within the window, denoted as V_{time} and V_{apex} respectively. A large V_{time} indicates a high probability that the window contains a depth edge; a high value of V_{apex} indicates a high probability that the window contains a shadow edge.

Novel Images from Decompositions The decomposed light transport component can be used to generate novel images of the scene with adjusted amounts of subsurface and multibounce scattering. First, by manipulating the ratio of different components: $I_n(x, y) = \lambda_1 D(x, y) + \lambda_2 G(x, y)$, we can generate a new image with different light transport effects. For example a scene material would look more strongly scattering when increasing λ_2, and would look more opaque by increasing λ_2. We show a simple example of new rendered images using different compositions of direct and global components in Fig. 11.

Since the subsurface illumination component is modeled by a Gaussian convolved with an exponential decaying function, we can adjust the α parameter of the exponential function and thus change the object appearance, as shown in Fig. 12.

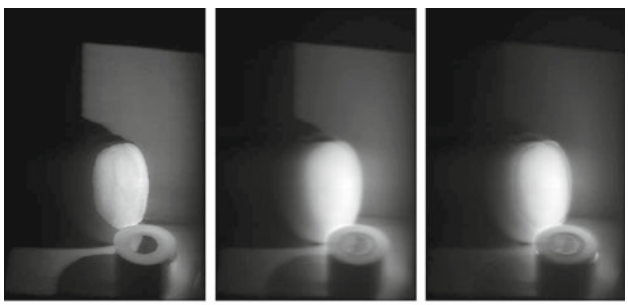


Fig. 11 Novel images based on decomposed components from tomato tape (iv) scene. (**a**) Direct component is magnified 5 times. (**b**) Simple time integration of the data. (**c**) The global component is magnified 5 times

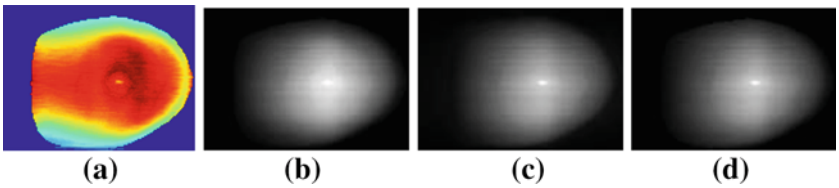


Fig. 12 Novel image based on model fitting of global light transport. (**a**) The α for each point on the tomato. (**b**) α scaled by 0.2, (**c**) original model fitting of global light transport, (**d**) α scaled by 5

5 Experimental and simulated data

The data acquisition setup used in this paper was presented by Velten et al. (2012, 2013). A brief summary is included here for completeness. The system is built by re-purposing modern imaging hardware, since no camera exists that can capture with such a low exposure time. At its core is a Hamamatsu C5680 streak tube (Hamamatsu 2012). This 2D sensor uses one dimension to encode space and one dimension to encode time of arrival of the impinging photons, essentially capturing x–t frames; this allows it to reach picosecond resolution, but at the cost of only recording one scanline of the scene at a time. A set of rotating mirrors enables scanning of the scene in the vertical dimension to capture the x–y–t ToF image. The illumination source is a femtosecond (fs) Kerr lens mode-locked Ti:Sapphire laser, which emits 50 fs long pulses at a rate of 75 MHz. The laser is focused with a lens onto a Lambertian diffuser, which acts as a *virtual point light source* and illuminates the entire scene with a spherically-shaped pulse. We intentionally occlude the virtual point light source within the scene, since the brightness of the virtual light source produces bleed-through and lens flare artifacts in the ToF image. The time resolution for the following datasets is 3.7 ps. The Figures show individual frames of the ToF images. All ToF images where captured as part of this work, except for scene (v), which is also used in Velten et al. (2013). Please refer to the supplementary videos for the complete ToF images visualized as videos.

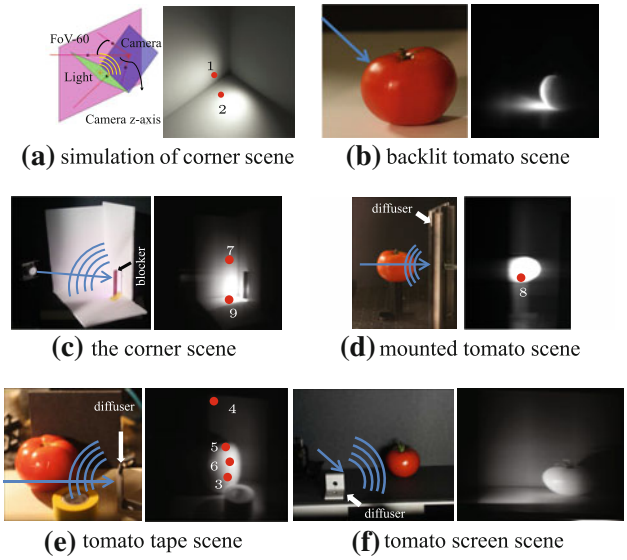


Fig. 13 Images of setup for all the datasets we are using in the paper. The *blue arrows* simulate the incident laser beam, each producing a virtual point light source within the scene. The separation plots for points 1–2 are in Figs. 4 and 5 respectively, point 3 is in Fig. 6, points 4–7 are in Fig. 7, and points 8–9 are in Fig. 8 (Color figure online)

(i) *Corner simulation: interreflections* We simulate ToF images by modifying the Monte Carlo path tracing integrator of the Physically Based Ray Tracer (PBRT). To render a ToF image, we generate a random light path connecting a camera pixel to the light source, compute the length of this light path, convert the length to the time required for light to reach the camera pixel by traveling along this light path, and store the contribution of this light path to the corresponding element in the ToF volume. This process is iterated for a large number of random light paths. The camera impulse response and quantization are applied to the simulated ToF image in MATLAB. For the corner scene demonstrated in Fig. 13a, we generate 110, 000 samples per pixel for a 512 $\times$ 512 image using the Monte Carlo path tracing integrator, a process requiring approximately 90 min when distributed across 44 3.8 Ghz Intel Xeon processors. Please refer to the supplementary video for the simulated ToF images.

(ii) *Backlit Tomato: pure subsurface scattering* The laser illuminates the backside of a tomato, as shown in Fig. 13b. The dominant transport effect within the tomato scene is subsurface scattering.

(iii) *Mounted Tomato: subsurface and simple multibounce scattering* This scene shown in Fig. 13d contains a diffuse wall and a tomato mounted on a post. The laser produces a virtual point light source on the diffuse wall, which then illuminates the tomato. This scene contains primarily direct, subsurface scattering and interreflection light transport effects.

(iv) *Corner: complex multibounce scattering* The scene shown in Fig. 13c consists of three diffuse planes arranged to form a concave corner. The laser illuminates an occluded point on the wall. Because of the large number of near discrete bounces occurring over short times and distances in this setup, decomposing the time profile into its individual components is challenging.

(v) *Tomato and Tape: complex subsurface and multibounce scattering* The scene in Fig. 13e combines several light transport effects: direct, interreflection, and subsurface scattering. The scene consists of a roll of tape, a tomato and a wall. The laser creates a virtual light source on a diffuser. Complex interreflections occur in the roll of tape, and subsurface scattering effects appear on the tomato.
(vi) *Tomato and Screen: complex subsurface and multibounce scattering* This scene, shown in Fig. 13f combines direct, interreflection, and subsurface scattering. The scene consists of a tomato, a back wall and a floor. The laser creates a virtual light source on a diffuser. The tomato, wall and floor receive direct illumination, and subsurface scattering effects appear on the tomato, complex interreflections including the subsurface scattering light from the tomato occur afterwards.

6 Results and Performance Validation

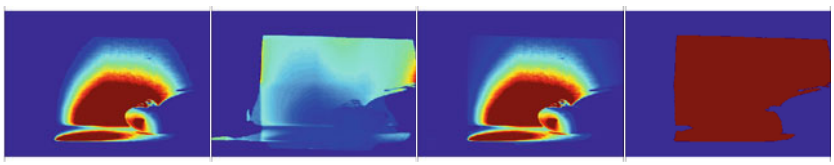
In this section we present results and practical details for both light component separation as presented in Sect. 3, and the subsequent applications described in Sect. 4.

Light Transport Component Separation For most pixels in a scene, the direct component can simply be identified as the first light arriving at the pixel. One can simply define a threshold above noise level and find the earliest element in the time profile that exceeds the threshold. This is, however, challenging in practice when signals are close to the noise level. Moreover many components of the scene simply do not have a direct component. This is why LiDAR systems often rely on the assumption that across all pixels, the weakest direct component in the scene is still stronger than the strongest global component. This is not generally true. In some cases, (1) pixels receive direct illumination but have very low intensities due to their orientations to the light and camera, material properties, or their distance from the camera. (2) Other pixels do not receive direct light, but exhibit a multibounce peak that is strong due to proximity to the camera, or favorable orientations or material properties. One can solve this problem by choosing a conservative threshold for separation, accepting that some of the direct component will be lost.

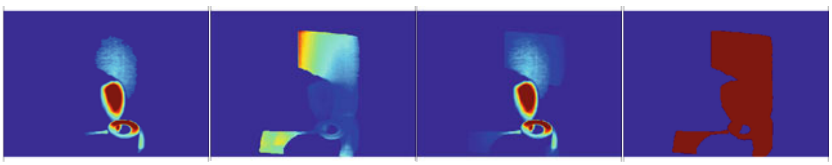
For scenes with complex light transport effects, we use a more sophisticated method to get a more accurate direct region map beyond simple thresholding. The key is to compute the inverse of the *Ratio* map in Sec. 4. The higher the value of the computed value for each pixel, the higher the probability that it receives direct illumination and nothing else. By including the pixels with high computed values, we could avoid the mis-thresholding in the cases discussed above. The illustration of the method and results for tomato screen (vi) and tomato tape (v) scenes are shown in Fig. 14.

We perform direct and global separation for the mounted tomato scene (iii), corner scene (iv), tomato and tape scene (v). We show several $x - y$ slices of the separated scenes in Figs. 15 and 16a, b respectively. In general, we observe the direct component in the earlier frames, whereas subsurface and multibounce scattering dominate the later frames.

We capture a ToF image for the mounted tomato scene (iii), and separate the image into its direct, subsurface scattering, and interreflection components. Figure 15a shows direct/global separation results on an $x - y$ image at a time when the direct component intersects the tomato. The tomato

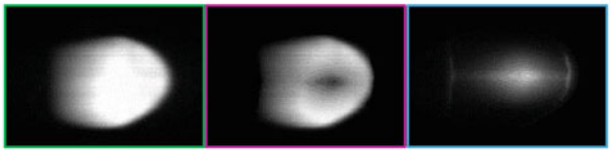


(a) illustrations for the tomato screen scene (vi)

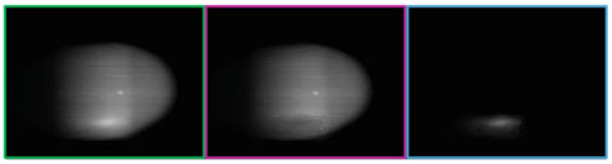


(b) illustrations for the tomato tape scene (v)

Fig. 14 Illustrations of the extraction of direct component regions, (*left*) thresholding by the peak value of the pixel time profiles, (*center left*) the inverse of the *Ratio* map reflecting the pixel containing only a direct reflection, (*center right*) combination of *left* and *center left* columns, (*right*) final region mask. Please notice for the tomato tape scene, the *bottom right part* of the back wall and the *top right part* of the floor have strong interreflections as well as reflections of subsurface scattering light from the tomato, that will significantly lower the value of the pixels. As a result, we rely on the integrated image as in Fig. 13 and incorporate regions with high intensity facing the light source



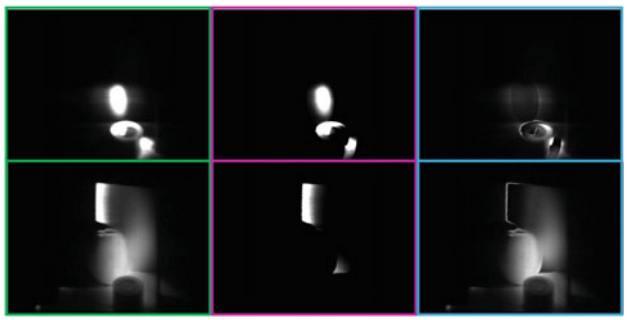
(a) direct/global decomposition



(b) subsurface scattering/interreflection decomposition

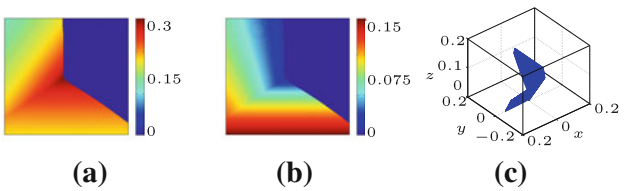
Fig. 15 Global light transport decomposition of the mounted tomato scene (iii) at two different times. (**a**) Direct/global separation. Algorithm 1 takes a ToF image (*left*) and decomposes the image into its direct (*middle*) and global (*right*) components. (**b**) Algorithm 2 takes a global image (*left*) and returns two images: the subsurface scattering component (*middle*) and the interreflection component (*right*). Light bouncing between the floor and the tomato produces an interreflection component. Note this global illumination image occurs at a time when the direct component has passed

(a) direct/global decomposition: corner scene (iv)



(b) direct/global decomposition: tomato tape scene (v)

Fig. 16 Direct/global separation results at two different times for each of (**a**) corner scene and (**b**) tomato tape scene. Our algorithm separates the original $x - y$ images (*left*) into its direct component (*middle*) and global component (*right*)



(a) **(b)** **(c)**

Fig. 17 Projective depth computation from direct component. (**a**) The projective depth of the simulated corner scene (i) in camera coordinates. (**b**) Recovered depth for the corner scene (i) in world coordinates. (**c**) 3D plot of the two walls with direct component

produces strong subsurface scattering effects as illustrated by the rightmost image. Figure 15b shows subsurface scattering/interreflection separation on an $x-y$ image at a later time when no direct component exists. The interreflection originates from light transport between the floor and the tomato.

The corner scene (iv) in Fig. 16a contains interreflections between the three walls forming a concave corner. The time profile for pixels near the corner of the wall behaves similarly to a subsurface scattering profile, that decays exponentially over time. Our separation algorithm identifies the direct and global components, although it cannot distinguish between interreflection and subsurface scattering in this example.

We also analyze the tomato and tape scene (v) in Fig. 16b. The figure shows the direct and global components at two different time frames. The tomato produces subsurface scattering effects, and we observe interreflections in the inner part of the tape roll and between the wall and tomato. The light from the subsurface scattering glow of the tomato illuminates the rest of the scene over a prolonged period of time. The algorithm once again separates the image into its direct and global components. Please refer to the supplementary video for additional results.

Projective Depth from ToF Results Depth estimation from ToF data requires the system to be calibrated to obtain L, t_0 and α (see Sect. 4). This calibration is performed by actually measuring the real position of a number of scene points, incorporating them to Eq. 10 by means of Eqs. 8 and 9, and solving it to obtain the calibration parameters. Once these parameters are known, the same Eq. 10 can be applied to all the scene, this time being w_z the unknown. Results for simulated ToF data are shown in Fig. 17 for the corner scene (i). In real data, the accuracy of the calibration is decreased by the fact that Eqs. 8–9 assume a pinhole camera model, which is not the case of our imaging system. This hinders the accuracy of the results, which can be seen in Fig. 18.

Examples of Subsurface Scattering Labelling Our algorithm for identifying subsurface scattering materials relies on thresholding the $Ratio(x, y)$ function (Eq. 11) at each pixel. When the function returns a value larger than a particular threshold T, the algorithm identifies the pixel as a point with subsurface scattering, as shown in Fig. 19. T is 70 for Fig. 19a, c, 100 for b. There are some errors in corners, such

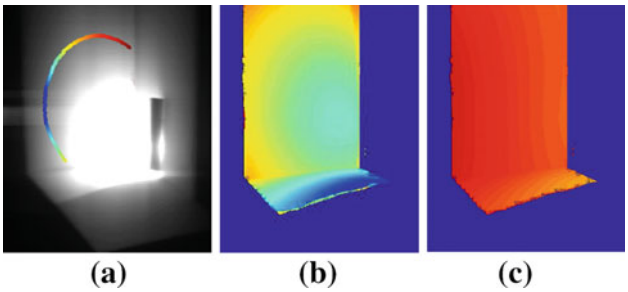


Fig. 18 Projective depth estimation for real data. (**a**) The points on a plane in the scene with the same ToF are distributed on an ellipse (see also Fig. 9). (**b**) Direct peak time map of the *back and bottom plane* in the corner scene. (**c**) The corresponding projective depth map

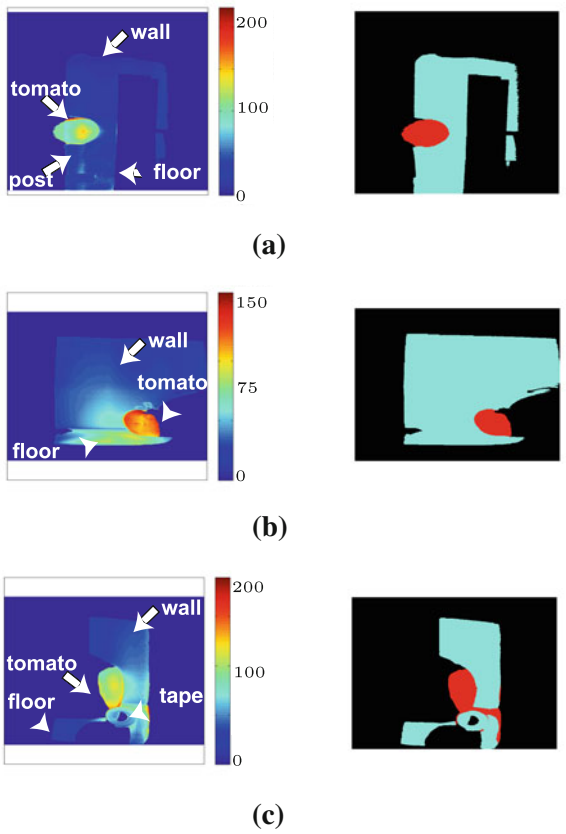


Fig. 19 Labelling subsurface scattering for the mounted tomato scene (iii), tomato screen scene (vi) and tomato tape scene (v). *Left column* the image is given by computing the ratio per pixel using Eq. 11. Note that higher values represent more subsurface scattering. *Right column* the image after thresholding. *Red* identifies subsurface scattering regions, *green* indicates no subsurface scattering, and *black* corresponds to background regions. Please notice that we exclude the regions that do not receive direct illumination from subsurface scattering (Color figure online)

as around the intersection of the wall and the floor, as well as the top edge of the tape roll, both due to strong interreflections. This is expected as multibounce scattering at time and length scales shorter than the instrument resolution behaves like subsurface scattering (see Sect. 3.1).

Characterization of Subsurface Scattering The subsurface scattering function extracted by our methods allows us to determine the scattering coefficient of scene materials and thus provides information for material identification and characterization. We create a scene material with variable scattering coefficient using a mixture of milk and water to demonstrate this ability. The milk contains floating fat bubbles and other constituents, that serve as scattering centers for the light. The subsurface scattering properties associated with different concentrations of milk in water can be characterized as a decaying exponential. We analyze subsurface scattering by first filling a tank with 3.6 liters of water, and gradually adding milk (2% reduced fat) to the water to increase the scattering properties of the medium. Our camera captures an $x - t$ image for 0, 5, 10, 20, 40, 80, and 160 ml concentrations of milk within the tank. Observing a point in the tank, the time profile of the returned light allows us to distinguish between the different milk concentrations.

A straight line closely approximates the logarithm of the observed time profile $\log(P(t))$ for different concentrations of milk in Fig. 20b. Figure 20c illustrates the exponential relationship between the decay constant and the milk concentration.

Validation of Edge Detection Edge detection algorithms aim to find edges through variations in depth, texture, and lighting. Different vision applications require different types of edges. Typical image edge detection algorithms extract edges due to brightness variations, which are coupled for all depth, texture, and light edges. In our ToF images, the depth edge

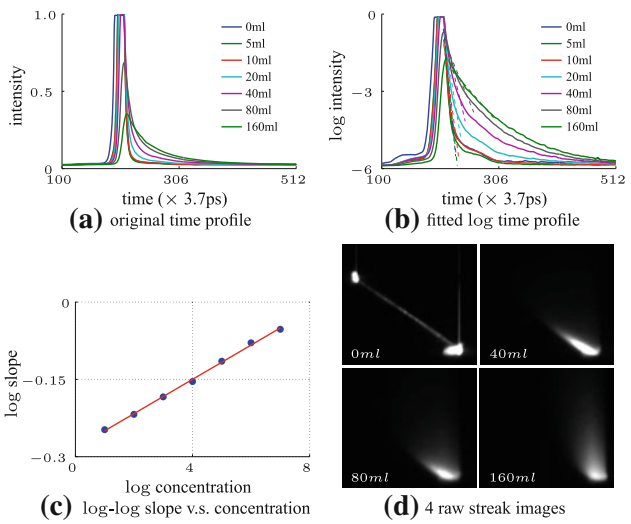


Fig. 20 Subsurface scattering in milky water. (**a**) The time profiles of the same point in 7 different concentrations of milky water. (**b**) *Lines* are fit to the log plot of the time profiles. (**c**) log–log plot for slope of lines versus concentration. (**d**) 4 raw streak images captured for milky water scene

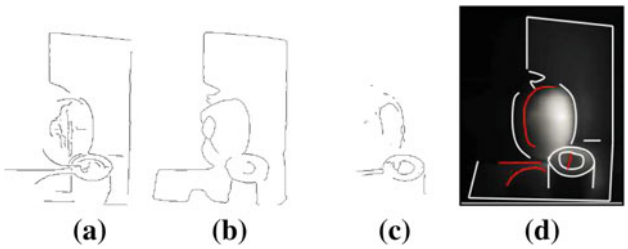


Fig. 21 Differentiation of depth edges and shadow edges. (**a**) Edge detection using Canny edge detector for the tomato tape scene. (**b**) *depth edge* from V_{time} map, (**c**) *shadow edge* from V_{apex} map, (**d**) time integrated image of the scene showing manually labeled depth edges (*white*) and shadow edges (*red*) for comparison (Color figure online)

points and shadow edge points have different local properties. We compare our edge detection based on 3D light transport with the Canny algorithm acting in the time integrated ToF image and with manually labeled edges in Fig. 21 for the tomato tape scene (v). The results show that our ToF approach to edge detection generates cleaner depth edges and shadow edges compared to the Canny algorithm: While image-based techniques do not work well with this type of data, analysis in the temporal dimension can yield more accurate detection. Still, some shadow edges remain undetected as such by our algorithm. Our current method for edge detection is based on thresholding of the computed V_{apex} and V_{time} maps, explained in Sect. 4, to obtain shadow edges and depth edges, respectively; complex cases such as parts of the scene receiving strong interreflections will "incorrectly" behave as depth edges and will not be detected as shadow edges by the algorithm.

Limitations of the Separation Method When considering separation of discrete bounces from subsurface scattering, it is important to note that these are not two completely distinct phenomena. Most surfaces contain at least some subsurface scattering. A second or third bounce striking a subsurface scatterer causes subsurface scattering just like the direct component does. Light emerging from a subsurface scatterer can afterwards undergo discrete bounces in the scene. This light is not distinguishable from immediate subsurface scattering light using our methods. The same limitation applies to related methods based on structured illumination as the mixing of the two phenomena occurs similarly in the temporal and in the spatial domain.

Subsurface scattering is accurately modelled as a series of random bounces inside the material and as such is simply multibounce scattering on time and lengths scales smaller than the resolution of the imaging system. For a physically accurate model, it is therefore expected that the appearance of multibounce scattering approaches the appearance of subsurface scattering as the path lengths between multibounce scattering events approach the resolution limit of the capture system. Thus, the time resolution dictates the ability to recover direct and multibounce components, especially in the presence of many interreflections. Lastly, a one hour capture time for a $672 \times 600 \times 512$ x–y–t ToF image limits our measurements to static scenes.

7 Conclusion

This work represents the first steps toward using ToF information in light transport analysis. We decompose and label scene elements according to the type of global light transport, measure coefficients of volumetric scattering, and produce depth maps from direct light transport.

In the future, combining ToF imaging with structured illumination could more accurately identify transport properties. Compact and low cost ultra-fast systems are becoming available due to massive commercial interest in optical communication and optical computing. We believe the addition of the time dimension in computer vision, computational photography, and active illumination techniques will enable a new range of applications.

Acknowledgments The work of the MIT affiliated coauthors was funded by the Media Lab Consortium Members, DARPA through the DARPA YFA grant, and the Institute for Soldier Nanotechnologies and U.S. Army Research Office under contract W911NF-07-D-0004. The work of the Tsinghua affiliated coauthors was supported by the National Basic Research Project (No.2010CB731800) of China and the Key Project of NSFC (No. 61120106003 and 60932007). O'Toole received the support of the NSERC PGS-D and GRAND NCE programs. Masia was additionally funded by an FPU grant, project TIN2010-21543 and an NVIDIA Graduate Fellowship.

References

Abramson, N. (1978). Light-in-flight recording by holography. *Optics Letters*, *3*(4), 121–123.

Arvo, J. (1993). Transfer equations in global illumination. *SIGGRAPH Course Notes*.

Bai, J., Chandraker, M., Ng, T., & Ramamoorthi, R. (2010). A dual theory of inverse and forward light transport. *European Conference on Computer Vision*, *2010*, 294–307.

Brooker, G. (2009). *Introduction to sensors for ranging and imaging*. Raleigh: Scitech.

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

Chen, T., Lensch, H., Fuchs, C., & Seidel, H. (2007). *Polarization and phase-shifting for 3d scanning of translucent objects. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1–8).

Gleckler, A., & Gelbart, A. (2000). Multiple-slit streak tube imaging lidar (ms-stil) applications. *Proceedings SPIE* (Vol. 4035, pp. 266–278).

Gupta, M., Agrawal, A., Veeraraghavan, A. & Narasimhan, S. (2011). *Structured light 3d scanning in the presence of global illumination. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 713–720).

Gupta, M., Tian, Y., Narasimhan, S., & Zhang, L. (2009). *(De) focusing on global light transport for active scene recovery. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2969–2976).

Gupta, O., Willwacher, T., Velten, A., Veeraraghavan, A., & Raskar, R. (2012). Reconstruction of hidden 3d shapes using diffuse reflections. *Optics Express*, *20*(17), 19096–19108.

Hamamatsu. (2012). Guide to streak cameras. http://sales.hamamatsu.com/assets/pdf/catsandguides/e_streakh.pdf.

Heide, F., Hullin, M., Gregson, J., & Heidrich, W. (2013). Low-budget transient imaging using photonic mixer devices. *Proceedings of ACM SIGGRAPH, ACM Transactions on Graphics*, *32*(4), 146:1–146:12.

Holroyd, M., Lawrence, J., & Zickler, T. (2010). A coaxial optical scanner for synchronous acquisition of 3d geometry and surface reflectance. *ACM Transactions on Graphics*, *29*(4), 99.

Huang, D., Swanson, E., Lin, C., Schuman, J., Stinson, W., Chang, W., et al. (1991). Optical coherence tomography. *Science*, *254*(5035), 1178–1181.

Jarabo, A., Masia, B., Velten, A., Barsi, C., Raskar, R., & Gutierrez, D. (2013). Rendering relativistic effects in transient imaging. *Proceedings of CEIG*.

Kajiya, J. (1986). The rendering equation. *ACM SIGGRAPH Computer Graphics*, *20*(4), 143–150.

Liu, S., Ng, T., & Matsushita, Y. (2010). Shape from second-bounce of light transport. *European Conference on Computer Vision*, *2010*, 280–293.

Mukaigawa, Y., Yagi, Y., & Raskar, R. (2010). Analysis of light transport in scattering media. *Proceedings of Computer Vision and Pattern Recognition (CVPR)* (pp. 153–160).

Naik, N., Zhao, S., Velten, A., Raskar, R., & Bala, K. (2011). Single view reflectance capture using multiplexed scattering and time-of-flight imaging. *ACM Transaction on Graphics*, *30*(6), 171.

Nayar, S., Krishnan, G., Grossberg, M., & Raskar, R. (2006). Fast separation of direct and global components of a scene using high frequency illumination. *ACM Transaction on Graphics*, *25*(3), 935–944.

Ng, T., Pahwa, R., Bai, J., Tan, K., & Ramamoorthi, R. (2012). From the rendering equation to stratified light transport inversion. *International Journal of Computer Vision*, *96*(2), 235–251.

O'Toole, M., Raskar, R., & Kutulakos, K. (2012). Primal-dual coding to probe light transport. *ACM Transactions on Graphics*, *31*(4), 39.

Pandharkar, R., Velten, A., Bardagjy, A., Lawson, E., Bawendi, M., & Raskar, R. (2011). *Estimating motion and size of moving non-line-of-sight objects in cluttered environments. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 265–272).

Reddy, D., Ramamoorthi, R., & Curless, B. (2012). Frequency-space decomposition and acquisition of light transport under spatially varying illumination. *European Conference on Computer Vision* (pp. 596–610).

Repasi, E., Lutzmann, P., Steinvall, O., Elmqvist, M., Göhler, B., & Anstett, G. (2009). Advanced short-wavelength infrared range-gated imaging for ground applications in monostatic and bistatic configurations. *Applied Optics*, *48*(31), 5956–5969.

Seitz, S., Matsushita, Y., & Kutulakos, K. (2005). *A theory of inverse light transport. Tenth IEEE International Conference on Computer Vision*, (Vol. 2, pp. 1440–1447).

Sen, P., Chen, B., Garg, G., Marschner, S., Horowitz, M., Levoy, M., et al. (2005). Dual photography. *ACM Transactions on Graphics*, *24*(3), 745–755.

Velten, A., Willwacher, T., Gupta, O., Veeraraghavan, A., Bawendi, M. G., & Raskar, R. (2012). Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature Communication*, *3*, 745–758.

Velten, A., Wu, D., Jarabo, A., Masia, B., Barsi, C., Joshi, C., et al. (2012). Relativistic ultrafast rendering using time-of-flight imaging. *In ACM SIGGRAPH Talks*, *41*(1).

Velten, A., Wu, D., Jarabo, A., Masia, B., Barsi, C., Joshi, C., Lawson, E., Bawendi, M., Gutierrez, D., & Raskar, R. (2013). Femto-photography: Capturing and visualizing the propagation of light. *Proceedindgs of SIGGRAPH, ACM Transaction on Graphics 32*(4).

Wang, L., & Wu, H. (2007). *Biomedical optics: Principles and imaging*. Hoboken, NJ: Wiley.

Wetzstein, G., & Bimber, O. (2007). Radiometric compensation through inverse light transport. *Proceedings of the 15th Pacific Conference on Computer Graphics and Applications* (pp. 391–399).

Wu, D., Wetzstein, G., Barsi, C., Willwacher, T., O'Toole, M., Naik, N., et al. (2012). Frequency analysis of transient light transport with applications in bare sensor imaging. *European Conference on Computer Vision*, *2012*, 542–555.

Wyant, J. (2002). White light interferometry. In *AeroSense 2002* (pp. 98–107). Bellingham, WA: International Society for Optics and Photonics.

Xia, H., & Zhang, C. (2009). Ultrafast ranging lidar based on real-time Fourier transformation. *Optics Letters*, *34*, 2108–2110.

Zhang, L., & Nayar, S. (2006). Projection defocus analysis for scene capture and image display. *ACM Transactions on Graphics*, *25*(3), 907–915.