Reconstructing Transient Images from Single-Photon Sensors

Matthew O'Toole, Felix Heide, David B. Lindell, Kai Zang, Steven Diamond, Gordon Wetzstein
Stanford University

Abstract

Computer vision algorithms build on 2D images or 3D videos that capture dynamic events at the millisecond time scale. However, capturing and analyzing "transient images" at the picosecond scale—i.e., at one trillion frames per second—reveals unprecedented information about a scene and light transport within. This is not only crucial for time-of-flight range imaging, but it also helps further our understanding of light transport phenomena at a more fundamental level and potentially allows to revisit many assumptions made in different computer vision algorithms.

In this work, we design and evaluate an imaging system that builds on single photon avalanche diode (SPAD) sensors to capture multi-path responses with picosecond-scale active illumination. We develop inverse methods that use modern approaches to deconvolve and denoise measurements in the presence of Poisson noise, and compute transient images at a higher quality than previously reported. The small form factor, fast acquisition rates, and relatively low cost of our system potentially makes transient imaging more practical for a range of applications.

1. Introduction

In 1978, Abramson recorded the first holographic image of a light pulse propagating through a scene [1]. Each pixel in this so-called "transient image" is a time impulse response function τ, representing the amount of light received at time t in response to a pulse of light emitted into the scene at time $t = 0$ (Figure 1). While the transient nature of light is hidden in images captured by conventional cameras or seen by the naked eye, these transient images unveil the sequence of complex scattering events that occur when light interacts with real-world environments.

Recent developments in sensing technologies have revitalized interest in transient imaging and, more importantly, the use of transient information for scene analysis [18]. For example, transient images provide a way to measure the 3D shape and albedo of objects [20, 21], to distinguish between objects composed of different scattering properties [29, 32], or measure the reflectance and scattering properties of materials [13, 23]. Transient imaging has also enabled unprecedented imaging capabilities such as looking-around-corners [5, 30], seeing through diffusers and thick scatterers [19, 26], or even lensless imaging [33].

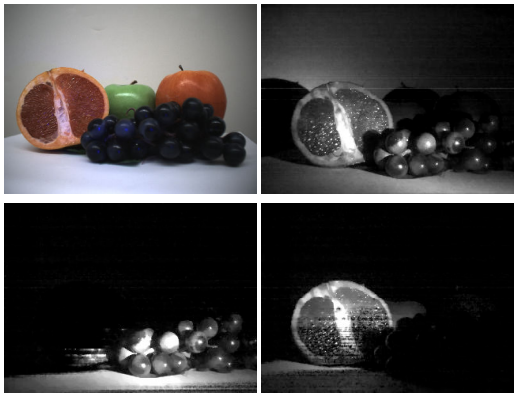


Figure 1. **Row 1:** Images of a scene captured with regular vision camera under ambient lighting (left) and a single-photon sensitive SPAD array under diffused laser illumination (right). **Row 2:** The temporal response approximately 703 ps (left) and 1171 ps (right) after emitting a pulse of light into the scene.

While several imaging technologies now exist for capturing transient images (*e.g.*, holography [1], streak camera [30, 31], photon mixer device (PMD) [15, 19], optical interferometry [12]), each approach has weighty practical limitations as outlined in Table 1. Although transient images record high-speed optical events, the physical process to capture these transient images may require upwards of several hours. In addition, these devices are often mounted onto vibration isolated tables (*e.g.*, [12]), or are prohibitively expensive (*e.g.*, [30, 31])—factors that prevent their use in everyday scenarios.

An emerging sensing technology known as single-photon avalanche diodes (SPAD) is a silicon-based photodetector capable of detecting and timestamping individual photon events (Figure 2). A SPAD pixel collects a histogram of photon arrival times (*i.e.*, a discretized version of the time impulse response $\tau[t]$) by repeatedly measuring the scene's response to a sequence of light pulses being emitted at MHz rates (Figure 3). SPAD pixels detect timestamped photon events at picosecond timescale resolutions,

	Acquisition time	Temporal resolution	Processing requirements	Digital image
Holography [1]	12 s	800 ps	N/A	✗
Streak camera [30]	>1 h	0.3-5.7 ps	Minimal	✓
PMD sensor [15]	90 s	1000 ps	High	✓
Optical interferometry [12]	>1 h	0.033 ps	Moderate	✓
SPAD (ours)	64 s	300 ps	High	✓

Table 1. Comparison of select transient imaging technologies. For a thorough review of available transient imaging technologies, see [18]. Our RAW measurements have a temporal resolution of 300 ps; after processing, the resolution is approximately 100 ps.

corresponding to a light pulse length of a few centimeters. A variety of SPAD architectures now exist, including 1D SPAD arrays with 256 pixels [4] and 2D SPAD arrays with 32×32 pixels [24].

Gariepy *et al*. [10] recently used this 2D SPAD array to capture a transient image of a light pulse bouncing around an environment. As transient imaging is an inherently noisy procedure, Gariepy *et al*. proposed a three-stage Gaussian-fitting procedure to denoise, temporally deconvolve, and interpolate the measured histograms. This procedure makes the strict assumption that the histograms contain a single pulse (*i.e*., a single light path), and therefore ignores the contribution of secondary pulses (*e.g*., the indirect path shown in Figure 3). The experimental setup achieves a temporal resolution of 500 ps and forms transient images over a relatively-short 10 min acquisition period.

We present the first SPAD-based system for *multi-path* transient imaging. Our reconstruction procedure denoises and deconvolves histograms *without* assuming single-path transport events. This enables our system to capture complex multi-path transport effects like subsurface scattering and interreflections—an important requirement for many transient imaging applications. In comparison to Gariepy *et al*. [10], our system also increases the acquisition speed, spatial resolution, and temporal resolution of the transient images.

Our aim is to make transient imaging more accessible to the computer vision and graphics community by overcoming the traditional cost, long acquisition times, and limited spatio-temporal resolution of existing systems.

2. Related work

Transient Imaging Technologies Select methods for capturing the transient image of a scene include those listed in Table 1, and a detailed overview is provided by Jarabo *et al*. [18]. These methods are all active, sending a light signal into an environment and computing the time required for that light to return back to a sensor. Generally speaking, all such devices interfere or couple two signals together to recover the time-of-flight of light, using one of three interferometric approaches: the optical, optoelectronic, or electronic interference of two signals.

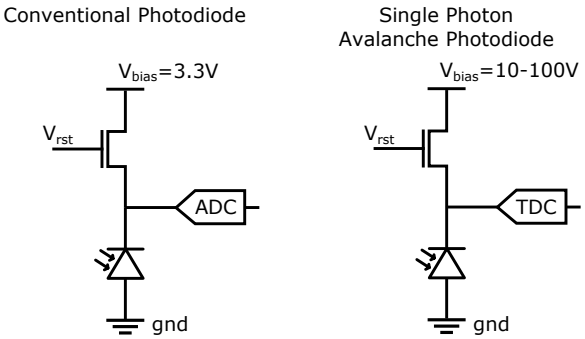


Figure 2. Simple readout circuit of a conventional photodiode (left) and a single photon avalanche diode (SPAD, right). Instead of digitizing collected photoelectrons with an analog-to-digital converter (ADC), a SPAD digitizes the time-of-arrival of individual photons with a time-to-digital converter (TDC). These events are detectable because each photon creates an electron avalanche, which amplifies the signal. While the two circuits are conceptually very similar, SPADs are operated well above their breakdown voltage to enable avalanching. The extreme voltages necessary for single photon detection require very different diode architectures in practice.

Abramson's holographic technique optically interferes light from a scene with a reference beam [1]. This analog holography technique, however, is a costly procedure compared to digital sensing technologies. Inspired by an imaging technique known as optical coherence tomography (OCT), Gkioulekas *et al*. [12] digitally measured optical interference to recover a transient imaging. Though the temporal resolution achieved is extremely high, the acquisition period is long and the system is sensitive to small vibrations.

A streak camera is an example of an optoelectronic transient imaging technology that couples an incident optical signal with a time-modulated electronic one [30, 31]. The streak camera uses the electronic signal to control an electric field that deflect photoelectrons (*i.e*., photons converted into electrons) onto a digital sensor. By changing the electric field as a function of time, the deflection amount indicates a photoelectron's time of arrival. Unfortunately, streak cameras remain prohibitively expensive and bulky.

More recently, solid-state sensing technologies have used the electronic interference of two signals to capture transient optical phenomena. Besides SPADs, photonic mixer devices (PMDs) have recently emerged as a low-cost alternative to measuring depth maps through time-of-flight. These sensors work by continuously modulating a light source, and electronically interfering the light signal reflected back in response with the original reference signal. By performing multiple measurements using different electronic signals, PMD sensors can recover a transient image of a scene [15, 19]. PMD sensors overcome some of the barriers to transient imaging, but are fundamentally limited by their low temporal resolution.

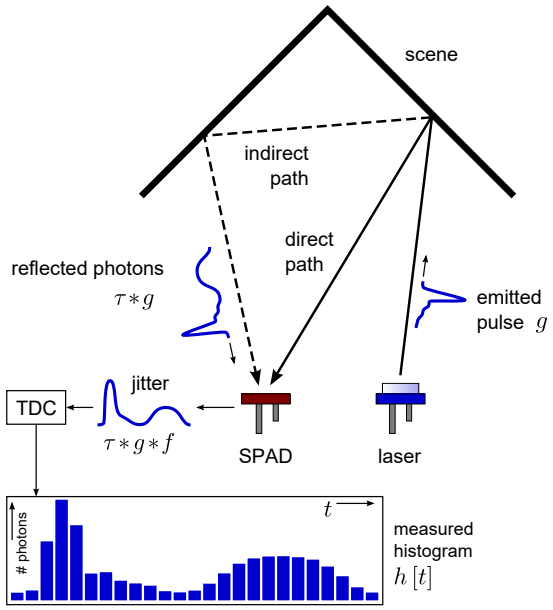


Figure 3. A short pulse g is emitted by a laser and scattered towards a SPAD detector. This scene exhibits global illumination effects, modeled by its time impulse response τ. Temporal jitter of the SPAD adds further uncertainty to the time stamping process. A time-to-digital converter (TDC) discretizes the time-of-arrival of a detected event. Repeating the experiment N times allows for a histogram h to be accumulated containing at most N event counts.

Single-photon avalanche diode (SPAD) Although avalanche diodes were first implemented in the 1960s [6, 14], SPADs have only recently become accessible through CMOS technology [25]. With only a single photon detected per pixel, avalanche diodes can recover the depth and albedo from the environment [21]. Other applications of SPADs include seeing through murky water [22], and capturing the motion [11] or shape [5] of objects hidden around a corner.

3. Single Photon Imaging

3.1. Single Photon Avalanche Diodes (SPADs)

Assume that a short laser pulse is emitted into a scene and some of the light is scattered back towards the detector. The laser pulse is described by a temporally-varying distribution of photons g. Then, the photon flux r incident on the detector during a discrete time interval t is the following [27]:

$$r\,[t] = (\tau * g)\,[t] + a\,[t]\,, \tag{1}$$

where $*$ is the 1D convolution operator, $a\,[t]$ is the ambient photon flux and τ is the temporal impulse response of the scene. The impulse response models photon travel time, albedo, and any other light transport effect. For a scene that exhibits only direct light transport, *i.e.*, light bounces off some object precisely once before returning to the detector, τ is a discrete Dirac function. In the presence of global light transport such as caustics, interreflections, or subsurface scattering, τ models the scene's response to an infinitely short laser pulse. Note that t in this notation represents the relative time w.r.t. the emission time of the laser pulse.

An ideal photon counter processing the stream of incident photons within a certain time window would sample the rate function λ as

$$\lambda\,[t] = \eta\,(r * f)\,[t] + d \tag{2}$$

where $\eta \in [0, 1]$ is the photon detection probability, which is comprised of the quantum efficiency and the avalanche probability of the SPAD. The dark count rate d [Hz] represents the number of false events that are detected in the absence of photons. Finally, f is the detector jitter, modeling uncertainty in the time-stamping mechanism. Jitter on the order of tens to a few hundred of ps is common for modern SPADs.

Photons that hit a SPAD create an electron avalanche that is time-stamped by the time-to-digital converter (TDC). After a detected event, however, the SPAD must be reset (or quenched) before another event can be recorded. This *dead time* is usually on the order of $100\,\mathrm{ns}$. Assuming that the length of the emitted laser pulse is significantly smaller than the SPAD's dead time, at most one of the photons in that pulse can trigger an event. Note that this is not necessarily the first photon that arrives. The presence or absence of a detected event within a short window is thus a Bernoulli trial.

Usually, the Bernoulli trial is repeated N times by firing N laser pulses at a rate that leaves sufficient time for the SPAD to reset itself between pulses. A low-level counting mechanism, for example implemented by a field programmable gate array (FPGA), accumulates events for a certain "exposure time" and stores the number of detected events per time bin in a histogram h. The probability of detecting a certain number of events in a histogram bin can be modeled as a Poisson distribution [27]:

$$h\,[t] \sim \mathcal{P}\,(N\lambda\,[t])\,. \tag{3}$$

Note, however, that this is an approximation that models individual photon events as being independent and which assumes low photon flux, which is the case in our experiments.

3.2. Capturing Spatio-temporal Volumes

Three categories of SPAD architectures exist: single pixel, 1D array, and 2D array.

Transient imaging with a single pixel requires either optical scanning or coding. Specifically, this involves raster scanning the scene with a 2-axis galvo [21], or coding the

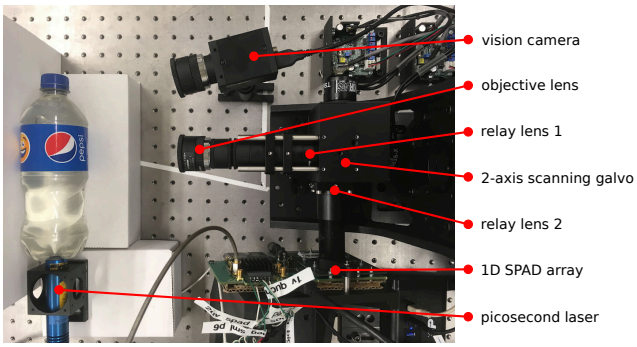


Figure 4. Transient imaging setup.

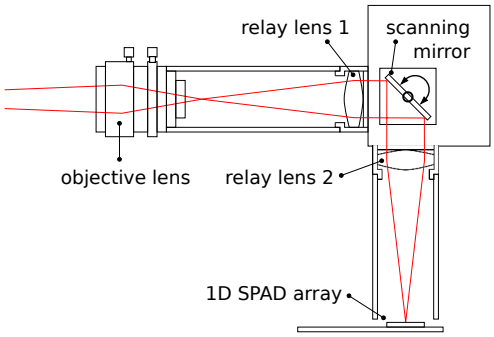


Figure 5. An illustration of the imaging arm from Figure 4 (not to scale). The components include an $8.5\,\mathrm{mm}$ focal length objective lens (Computar M8513), two $75\,\mathrm{mm}$ achromatic doublet relay lenses (Thorlabs AC254-075-A-ML), a scanning galvo system (Thorlabs GVS012), and a 1D SPAD array (Fastree3D LinoSPAD). The objective lens focuses light from the scene to an image plane, and the pair of relay lenses focuses the image onto the SPAD sensor. The galvo is placed at the Fourier plane of the system to control the scanline that the 1D SPAD images from the scene. Note that our system uses a 2-axis galvo in practice (one mirror scanning, and the second mirror fixed); the illustration shows a 1-axis galvo for simplicity.

incident illumination or detection and then applying a reconstruction procedure to decode the measurements [20]. A 2D SPAD array completely eliminates the need for optical scanning or coding [11], though the large footprint required of today's SPAD pixels severely limits resolution and fill factors. Alternatively, the 2D SPAD array can also be scanned to form a larger composite image [28].

We opt instead to scan scenes with a 1D SPAD array, as illustrated in Figure 5.

3.3. Reconstructing Transient Images

The problem of recovering transient images from noisy and blurry histograms is a deconvolution problem in the presence of Poissonian noise. To formulate this, we represent the temporal impulse response (*i.e.*, the transient image), the measured histogram, and the dark count as discrete column vectors $\boldsymbol{\tau}, \mathbf{h}, \mathbf{d} \in \mathbb{R}^{n_x n_y n_t \times 1}$. Using Equation (3), we express $\mathbf{h} \sim \mathcal{P}\left(\mathbf{A}\boldsymbol{\tau} + \boldsymbol{d}\right)$, where $\mathbf{A} \in \mathbb{R}^{n_x n_y n_t \times n_x n_y n_t}$

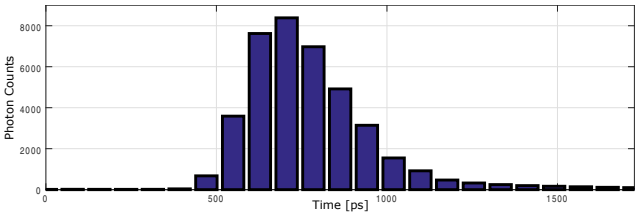


Figure 6. Part of a recorded histogram showing a temporal impulse response function $(g * f)[t]$ captured with the proposed system.

encodes the convolution of the transient image with the laser pulse and SPAD jitter $(g * f)[t]$ (see Figure 6) as well as other scale factors. Each transient image has a resolution of $n_x \times n_y$ pixels and each pixel has n_t time bins. We formulate the reconstruction problem as a maximum likelihood estimation with non-negativity constraints:

$$\begin{aligned} \underset{\{\tau\}}{\text{minimize}} \quad & -\log\left(p\left(\mathbf{h}|\mathbf{A}\boldsymbol{\tau}\right)\right) + \Gamma\left(\boldsymbol{\tau}\right), \\ \text{subject to} \quad & 0 \leq \boldsymbol{\tau} \end{aligned} \tag{4}$$

where $p\left(\mathbf{h}|\mathbf{A}\boldsymbol{\tau}\right)$ is the likelihood of observing measurements $\mathbf{h}$ for a known transient image $\boldsymbol{\tau}$ and $\Gamma\left(\boldsymbol{\tau}\right)$ is an additional prior on the recovered signal. Possible choices for this include total variation (TV) [28], sparsifying convolutional filters [16], or self-similarity priors such as video block-matching and 3D filtering (VBM3D) [7].

Without loss of generality, we replace nonnegativity constraints by the indicator function $\mathcal{I}_{\mathbb{R}_+}(\cdot)$ and apply a splitting approach that represents the objective as a sum of independent penalty terms with a global consensus enforced by the constraints:

$$\begin{aligned} \underset{\{\boldsymbol{\tau}\}}{\text{minimize}} \quad & -\log\left(p\left(\mathbf{h}|\mathbf{z}_1\right)\right) + \mathcal{I}_{\mathbb{R}_+}\left(\mathbf{z}_2\right) + \Gamma\left(\mathbf{z}_3\right) \\ \text{subject to} \quad & \underbrace{\begin{bmatrix} \mathbf{A} \\ \mathbf{I} \\ \mathbf{I} \end{bmatrix}}_{\mathbf{K}} \boldsymbol{\tau} - \underbrace{\begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \mathbf{z}_3 \end{bmatrix}}_{\mathbf{z}} = 0 \end{aligned} \tag{5}$$

In this formulation, $\mathbf{z}_i \in \mathbb{R}^{n_x n_y n_t \times 1}$ are slack variables. Note that Equations (4) and (5) are equivalent. Using the Augmented Lagrangian for Equation (5), we can derive an iterative update scheme using the alternating direction method of multipliers (ADMM) [3]. This scheme is outlined by Algorithm 1. In this formulation, M is the number of iterations and $\mathbf{u} = \begin{bmatrix} \mathbf{u}_1^T & \mathbf{u}_2^T & \mathbf{u}_3^T \end{bmatrix}^T$ is the scaled dual of the Lagrange multiplier. In the following sections, we evaluate several different priors $\Gamma\left(\boldsymbol{\tau}\right)$, including an isotropic transverse 2D total variation (TV) prior, an isotropic 3D TV prior, and a video denoiser (VBM3D) [7]. We derive Algorithm 1 as well as closed-form solutions for each of the sub-problems and the priors in the Supplemental Material.

Algorithm 1 ADMM-based denoising and deconvolution

for $k = 1$ **to** M

$\boldsymbol{\tau} \leftarrow \underset{\{\boldsymbol{\tau}\}}{\arg\min}\ \frac{1}{2}\left\|\mathbf{K}\boldsymbol{\tau} - \mathbf{z} + \mathbf{u}\right\|_2^2$

$\mathbf{z}_1 \leftarrow \underset{\{\mathbf{z}_1\}}{\arg\min}\ -\log\left(p\left(\mathbf{h}|\mathbf{z}_1\right)\right) + \frac{\rho}{2}\left\|\mathbf{A}\boldsymbol{\tau} + \mathbf{u}_1 - \mathbf{z}_1\right\|_2^2$

$\mathbf{z}_2 \leftarrow \underset{\{\mathbf{z}_2\}}{\arg\min}\ \mathcal{I}_{\mathbb{R}_+}\left(\mathbf{z}_2\right) + \frac{\rho}{2}\left\|\boldsymbol{\tau} + \mathbf{u}_2 - \mathbf{z}_2\right\|_2^2$

$\mathbf{z}_3 \leftarrow \underset{\{\mathbf{z}_3\}}{\arg\min}\ \Gamma\left(\mathbf{z}_3\right) + \frac{\rho}{2}\left\|\boldsymbol{\tau} + \mathbf{u}_3 - \mathbf{z}_3\right\|_2^2$

$\mathbf{u} \leftarrow \mathbf{u} + \mathbf{K}\boldsymbol{\tau} - \mathbf{z}$

end for

4. Implementation

LinoSPAD All experiments described in this work are recorded with a 1D SPAD line sensor: LinoSPAD [4]. Histograms are recorded with 256 individual SPAD detectors using 64 TDCs. Each SPAD pixel is $24\,\mu\mathrm{m}$ in diameter, has a fill factor of 41%, and has a photon detection efficiency of 20% at $450\,\mathrm{nm}$. The median value of the dark count rate is $2.5\,\mathrm{kHz}$.

We combine this system with a picosecond laser (ALPHALAS PICOPOWER-LD-450-50) that has a wavelength of $450\,\mathrm{nm}$, peak power of $450\,\mathrm{mW}$, and a repetition rate of up to $50\,\mathrm{MHz}$. The LinoSPAD is synchronized to the reference clock generated by the picosecond laser.

The TDCs and histogram accumulation in LinoSPAD are all implemented on a Xilinx Spartan 6 FPGA. The histograms accumulated in the FPGA have exactly 280 time bins when operating the SPAD at $50\,\mathrm{MHz}$ (or a period of $20\,000\,\mathrm{ps}$ between laser pulses). Each time bin has an average length of $\frac{20000}{280} = 71.4\,\mathrm{ps}$. The full-width at half-maximum (FWHM) of a measured laser pulse is approximately $300\,\mathrm{ps}$ post jitter.

Histogram Calibration The LinoSPAD TDCs assign timestamps to photons non-linearly. Specifically, the average number of photons detected in each histogram bin is non-uniform under ambient lighting conditions. The LinoSPAD provides a correction module on the FPGA that redistributes photon counts such that the histogram appears uniform under ambient light. This correction step brings the 280 bin histogram down to a 256 bin histogram, where each bin has a time resolution of $\frac{20000}{256} = 78.1\,\mathrm{ps}$. The histogram bins for the LinoSPAD are only approximately Poisson distributed as a result of this correction step.

Moreover, each pixel has a constant offset in the timestamps of incident photons. The arrival time of a laser pulse therefore varies across LinoSPAD pixels. To compensate for this per-pixel offset, we circularly shift the histograms such that the centroid of each pulse are aligned temporally.

Because the dark count rate for certain LinoSPAD pixels is high, we also capture a measurement under dark conditions to recover the expected dark count rate per pixel, and incorporate this information into the reconstruction procedure.

Optomechanical Components To scan spatio-temporal volumes, we integrate the line sensor into an optical assembly that sequentially scans an image by sweeping the line horizontally over the desired field of view. The prototype is shown in Figures 4 and 5. Our scanning prototype consists of the LinoSPAD mounted onto a translation stage to manually adjust focus. The scanning setup consists of the following parts: an $8.5\,\mathrm{mm}$ objective lens, two $75\,\mathrm{mm}$ Achromatic Doublets, and a 2-axis scanning galvo system. The light from the scene is focused by the objective lens onto the image plane, brought into the Fourier plane by the first achromatic doublet, and focused onto the LinoSPAD once again by the second achromatic doublet.

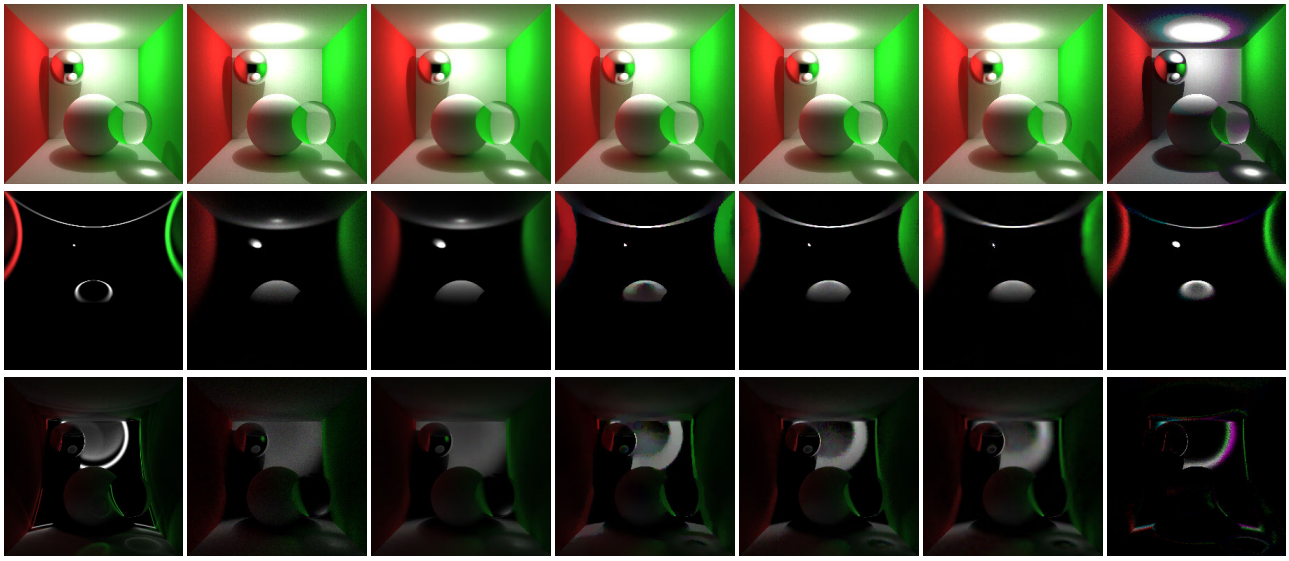
A National Instruments data acquisition device (NI-DAQ USB-6343) controls the angle of one galvo mirror $\pm 1.6°$ while keeping the second galvo mirror fixed. Each scanline takes approximately $0.2\,\mathrm{s}$ to capture. We scan 320 lines to obtain a 2D image of size 320×256 in $64\,\mathrm{s}$ of acquisition time, and remove 16 SPAD pixel values with high dark count rates to achieve a final resolution of 320×240.

Computation The optimization procedure is implemented in MATLAB, and uses the VBM3D toolbox provided by Dabov *et al*. [7, 8]. Running 200 iterations of the reconstruction algorithm requires approximately $1\,\mathrm{h}$ per transient image for 2D TV and 3D TV, and $5\,\mathrm{h}$ for VBM3D on a machine with an Intel Xeon CPU E5-2650 processor ($2.60\,\mathrm{GHz}$) and 24 GB of RAM.

5. Results

Simulations We evaluate the effectiveness of several reconstruction procedures (using different priors $\Gamma\left(\cdot\right)$ as explained in the Supplemental Material document) on a simulated transient image in Figure 7. The ground truth data was rendered using bidirectional path tracing [17]. We simulate measured data by blurring the ground truth temporally and adding Poisson noise to the result.

The Anscombe VBM3D result applies a variance-stabilizing transform [2] to the measured data followed by a denoising step via VBM3D [7]. Compared with the simulated measurements, no change in the PSNR is observable. The Poisson-TV (2D) and Poisson-TV (3D) approaches simultaneously denoise and deblur the result, producing a significant increase in the PSNR. We include comparisons of the transverse 2D and the 3D total variation (TV) priors, because recent work on depth reconstruction from SPAD measurements has also employed a 2D TV prior [28]. Here,



Ground truth	Measured (22.9 dB)	Anscombe VBM3D (22.9 dB)	Poisson-TV (2D) (29.5 dB)	Poisson-TV (3D) (33.0 dB)	Poisson-VBM3D (34.9 dB)	Gaussian Fit [10] (22.2 dB)

Figure 7. Denoising & deblurring results from a simulated transient image. **Row 1:** Regular image of the scene, constructed by integrating the transient images over the temporal dimension. **Row 2-3:** Each row corresponds to a frame of the corresponding transient image. Notice the sharp wavefront constructed by the Poisson-TV (3D) and Poisson-VBM3D priors compared to the other reconstruction methods. The 2D-TV prior does not impose temporal consistency, which is even more apparent in the Supplemental Video. The Gaussian fit only preserves the direct pulse from the original transient image.

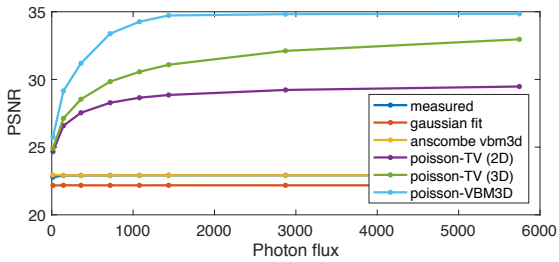


Figure 8. Evaluation of PSNR for the simulated results in Figure 7. The photon flux corresponds to the intensity of the light source used to generate the simulated measured image.

we show that for transient imaging, a 3D TV prior is superior. However, the best simulated results were achieved with a Poisson-VBM3D prior. In addition, we also implement the Gaussian fitting approach described by Gariepy *et al*. [10]; although the reconstructed direct channel matches the ground truth closely, the PSNR is worse than the original measured data because the indirect component is removed in this process. The PSNR for each reconstruction procedure for different photon flux values (corresponding to light source intensity) is shown in Figure 8.

Reconstructed Transient Images We captured 7 different transient images using our 1D SPAD array and reconstruction procedure. These images, shown in Figures 1, 9, and 11, include the following: *fruit*, *resolution chart*, *optical fiber*, *soda bottle*, *CVPR logo*, *foam box*, and *statue*. The size of each scene ranges from 25 cm wide for the *soda bottle* to 60 cm for the *CVPR logo*. Figure 9 was reconstructed with Poisson-VBM3D, and Figures 1 and 11 used Poisson-TV (3D). The same scenes viewed with a regular computer vision camera are shown in Figures 1 and 10.

Fruit: We place a diffuser in front of the laser to diffusely illuminate the fruit in Figure 1. The sliced orange shown on the left is translucent, and remains lit for a short period of time after the initial wavefront passes due to subsurface scattering. The fruit shown on the right are plastic and opaque, producing no subsurface scattering effect.

Resolution chart: The resolution chart demonstrates the spatial resolution of our scanning system and reconstruction procedure. The reconstructed frames show a sharp wavefront moving from the left side of the scene to the right.

Optical fiber: The picosecond laser sends light through an optical fiber of 2 mm in diameter. The fiber emits a small amount of light from its side, lighting up the entire optical fiber. In this scene, the length of the optical fiber is long enough that two laser pulses (20 ns apart) can be seen in the transient image at the same time. As the laser pulse propagates through the fiber, it broadens due to the dispersion property of multimode fibers.

Soda bottle: We filled a soda bottle with milky water to reproduce the bottle example of Velten *et al*. [31]. The laser

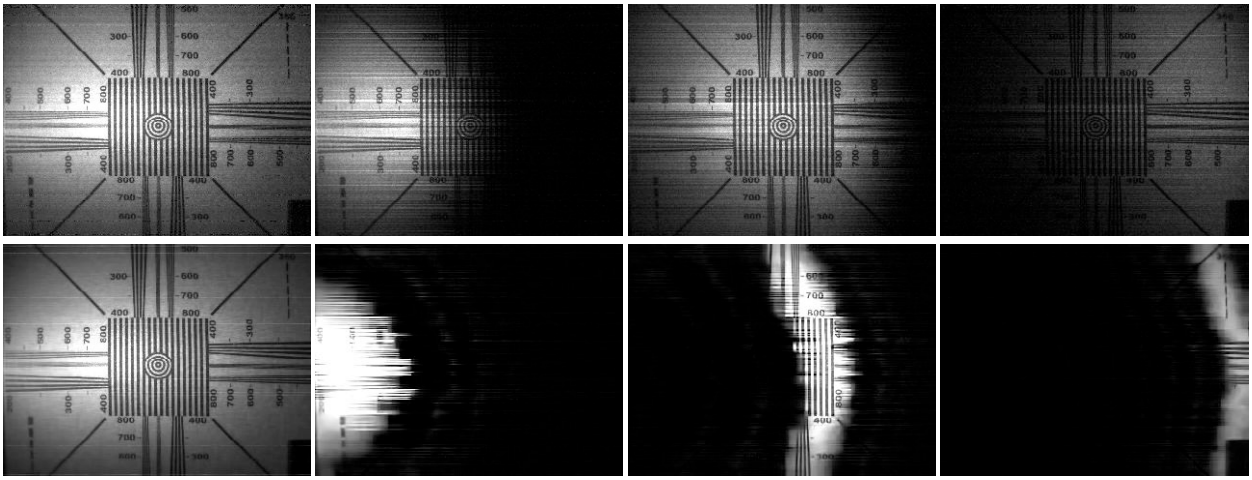


Figure 9. Comparison between original unprocessed SPAD data (Row 1) and the output of our proposed reconstruction procedure with the Poisson-VBM3D prior (Row 2). **Column 1:** Regular image of the scene under laser illumination. **Column 2-4:** Frames of the scene at different instance in time in response to a pulse of light. Our reconstructed frames show a sharp wavefront and significantly less noise when compared to the unprocessed frames.

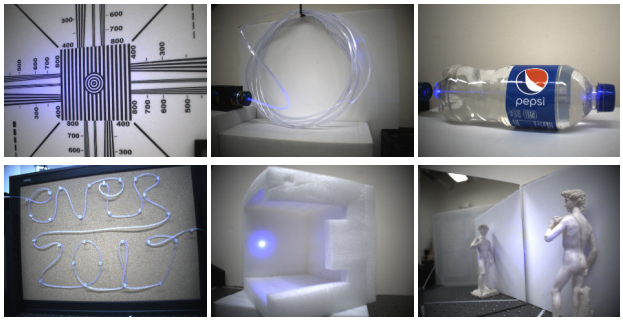


Figure 10. Scenes from Figures 9 and 11 captured with a regular vision camera under ambient lighting.

shines light at one end of the bottle, the light propagates to the cap, and a small amount of light is reflected back to the base of the bottle. Note the floor is lit indirectly by the light pulse passing through the bottle as well.

CVPR logo: The optical fiber spells "CVPR 2017" and each frame of the transient image lights up a single letter.

Foam box: This scene observes a variety of transport effects. The laser illuminates a point within the cube, producing a strong direct reflection. The translucent nature of the foam produces subsurface scattering around this direct point. The box remains lit for a long period of time due to significant interreflections occurring within the box. An artifact of the reconstruction procedure appears in one of the frames as a large black spot at the point directly illuminated by the laser.

Statue: A statue of David is positioned in front of a mirror, and diffusely lit by the laser. Light first illuminates the statue before propagating to the mirror.

6. Discussion

In summary, we demonstrate transient imaging of complex light transport effects using single photon avalanche diodes. To overcome the noise and temporal resolution limits imposed by the SPAD measurements, we propose a denoising & deblurring algorithm that effectively increases the PSNR of these transient images. We demonstrate the capabilities of our system by capturing images with a 1D SPAD array and a scanning galvo system for a variety of scenes with complex transport effects.

Limitations Though our scanned system captures high-quality data using an acquisition time of $64\,\mathrm{s}$, the scene must remain static during this acquisition period. This can be impractical for many scenarios that require imaging at video frame rates.

Although the resolution of our transient images remains far below that of regular camera sensors, significant progress is being made in developing higher-resolution SPAD sesnsors (*e.g.*, [9]).

Our system also currently requires a dark room, and only captures monochromatic transient images. The availability of lasers at other wavelengths (including NIR wavelengths) offers the possibility of multi-spectral transient imaging.

Future Work Our reconstruction procedure currently assumes a Poisson noise model. Though the detected timestamped events do follow a Poisson distribution for low flux levels, one of the calibration steps of the LinoSPAD distorts the noise model of our histograms. For future work, we will revise the reconstruction procedure to better handle the noise characteristics of the LinoSPAD, though this may not be necessary for other SPAD implementations.

Figure 11. Reconstructed transient images using the Poisson-TV (3D) prior. **Column 1**: Images of the scene under laser illumination. Each pixel value in this image is the result of integrating the number of photons in a histogram. **Column 2-4**: Frames from a transient image appearing in chronological order.

Conclusion The ability to capture transient images enables computer vision algorithms to exploit the finite speed of light. Our proposed system is a first step towards making time-resolved computational imaging practical for the computer vision community, potentially allowing for fundamental assumptions in existing algorithms to be revisited. Finally, unprecedented imaging capabilities, such as non-line-of-sight imaging and descattering, may become feasible in practical scenarios with the proposed system.

Acknowledgments The authors are grateful to Edoardo Charbon and Pierre-Yves Cattin for providing the LinoSPAD sensor used in this work. O'Toole is supported by the Government of Canada through the Banting Postdoctoral Fellowships program, and Wetzstein is supported by the National Science Foundation (NSF) CAREER award and Terman Faculty Fellowship.

References

[1] N. Abramson. Light-in-flight recording by holography. *Optics Letters*, 3(4):121–123, Oct. 1978. 1, 2

[2] F. J. Anscombe. The transformation of Poisson, binomial and negative-binomial data. *Biometrika*, 35(3-4):246–254, Dec. 1948. 5

[3] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, Jan. 2011. 4

[4] S. Burri, H. Homulle, C. Bruschini, and E. Charbon. LinoSPAD: a time-resolved 256×1 CMOS SPAD line sensor system featuring 64 FPGA-based TDC channels running at up to 8.5 giga-events per second. *Proc. SPIE*, 9899:1–10, Apr. 2016. 2, 5

[5] M. Buttafava, J. Zeman, A. Tosi, K. Eliceiri, and A. Velten. Non-line-of-sight imaging using a time-gated single photon avalanche diode. *Optics Express*, 23(16):20997–21011, Aug. 2015. 1, 3

[6] E. Charbon. Towards large scale CMOS single-photon detector arrays for lab-on-chip applications. *Journal of Physics D: Applied Physics*, 41(9):094010, 2008. 3

[7] K. Dabov, A. Foi, and K. Egiazarian. Video denoising by sparse 3D transform-domain collaborative filtering. *Proc. European Signal Processing Conference*, pages 145–149, Sept. 2007. 4, 5

[8] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans. on Image Processing*, 16(8):2080–2095, 2007. 5

[9] N. A. W. Dutton, I. Gyongy, L. Parmesan, S. Gnecchi, N. Calder, B. R. Rae, S. Pellegrini, L. A. Grant, and R. K. Henderson. A SPAD-based QVGA image sensor for single-photon counting and quanta imaging. *IEEE Trans. on Electron Devices*, 63(1):189–196, Jan. 2016. 7

[10] G. Gariepy, N. Krstaji, R. Henderson, C. Li, R. R. Thomson, G. S. Buller, B. Heshmat, R. Raskar, J. Leach, and D. Faccio. Single-photon sensitive light-in-fight imaging. *Nature Communications*, 6(6021), Jan. 2015. 2, 6

[11] G. Gariepy, F. Tonolini, R. Henderson, J. Leach, and D. Faccio. Detection and tracking of moving objects hidden from view. *Nature Photonics*, 10, Jan. 2016. 3, 4

[12] I. Gkioulekas, A. Levin, F. Durand, and T. Zickler. Micron-scale light transport decomposition using interferometry. *ACM Trans. on Graph.*, 34(4):37:1–37:14, July 2015. 1, 2

[13] I. Gkioulekas, A. Levin, and T. Zickler. An evaluation of computational imaging techniques for heterogeneous inverse scattering. *Proc. ECCV*, 9907:685–701, 2016. 1

[14] R. H. Haitz. Studies on optical coupling between silicon pn junctions. *Solid-State Electronics*, 8(4):417–425, 1965. 3

[15] F. Heide, M. B. Hullin, J. Gregson, and W. Heidrich. Low-budget transient imaging using photonic mixer devices. *ACM Trans. on Graph.*, 32(4):45:1–45:10, July 2013. 1, 2

[16] F. Heide, L. Xiao, A. Kolb, M. B. Hullin, and W. Heidrich. Imaging in scattering media using correlation image sensors and sparse convolutional coding. *Optics Express*, 22:26338–26350, 2014. 4

[17] A. Jarabo, J. Marco, A. Muñoz, R. Buisan, W. Jarosz, and D. Gutierrez. A framework for transient rendering. *ACM Trans. on Graph.*, 33(6):177:1–177:10, Nov. 2014. 5

[18] A. Jarabo, B. Masia, J. Marco, and D. Gutierrez. Recent advances in transient imaging: A computer graphics and vision perspective. *Visual Informatics*, 2017. 1, 2

[19] A. Kadambi, R. Whyte, A. Bhandari, L. Streeter, C. Barsi, A. Dorrington, and R. Raskar. Coded time of flight cameras: sparse deconvolution to address multipath interference and recover time profiles. *ACM Trans. on Graph.*, 32(6):167, 2013. 1, 2

[20] A. Kirmani, A. Colaco, F. N. C. Wong, and V. K. Goyal. CoDAC: a compressive depth acquisition camera framework. *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 5425–5428, 2012. 1, 4

[21] A. Kirmani, D. Venkatraman, D. Shin, A. Colaco, F. N. C. Wong, J. H. Shapiro, and V. K. Goyal. First-photon imaging. *Science*, 343(6166):58–61, 2014. 1, 3

[22] A. Maccarone, A. McCarthy, A. Halimi, R. Tobin, A. M. Wallace, Y. Petillot, S. McLaughlin, and G. S. Buller. Depth imaging in highly scattering underwater environments using time-correlated single-photon counting. *Proc. SPIE*, 9992:1–8, 2016. 3

[23] N. Naik, S. Zhao, A. Velten, R. Raskar, and K. Bala. Single view reflectance capture using multiplexed scattering and time-of-flight imaging. *ACM Trans. on Graph.*, 30(6):171:1–171:10, Dec. 2011. 1

[24] J. Richardson, R. Walker, L. Grant, D. Stoppa, F. Borghetti, E. Charbon, M. Gersbach, and R. K. Henderson. A 32 × 32 50ps resolution 10 bit time to digital converter array in 130nm CMOS for time correlated imaging. *IEEE Custom Integrated Circuits Conference*, pages 77–80, Sept. 2009. 2

[25] A. Rochas, M. Gani, B. Furrer, P. A. Besse, R. S. Popovic, G. Ribordy, and N. Gisin. Single photon detector fabricated in a complementary metaloxidesemiconductor high-voltage technology. *Review of Scientific Instruments*, 74(7):3263–3270, June 2003. 3

[26] G. Satat, B. Heshmat, D. Raviv, and R. Raskar. All photons imaging through volumetric scattering. *Scientific Reports*, 6, Sept. 2016. 1

[27] D. Shin. Computational imaging with small numbers of photons. *Ph.D. thesis, Massachusetts Institute of Technology*, Feb. 2016. 3

[28] D. Shin, F. Xu, D. Venkatraman, R. Lussana, F. Villa, F. Zappa, V. K. Goyal, F. N. C. Wong, and J. H. Shapiro. Photon-efficient imaging with a single-photon camera. *Nature Communications*, 7(12046):1–7, 2016. 4, 5

[29] S. Su, F. Heide, R. Swanson, J. Klein, C. Callenberg, M. Hullin, and W. Heidrich. Material classification using raw time-of-flight measurements. *Proc. CVPR*, June 2016. 1

[30] A. Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, M. G. Bawendi, and R. Raskar. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature Communications*, 3, Mar. 2012. 1, 2

[31] A. Velten, D. Wu, A. Jarabo, B. Masia, C. Barsi, C. Joshi, E. Lawson, M. Bawendi, D. Gutierrez, and R. Raskar. Femto-photography: Capturing and visualizing the propagation of light. *ACM Trans. on Graph.*, 32(4):44:1–44:8, July 2013. 1, 2, 6

[32] D. Wu, A. Velten, M. O'Toole, B. Masia, A. Agrawal, Q. Dai, and R. Raskar. Decomposing global light transport using time of flight imaging. *International Journal of Computer Vision*, 107(2):123–138, Apr. 2014. 1

[33] D. Wu, G. Wetzstein, C. Barsi, T. Willwacher, M. O'Toole, N. Naik, Q. Dai, K. Kutulakos, and R. Raskar. Frequency analysis of transient light transport with applications in bare sensor imaging. *Proc. ECCV*, 7572:542–555, 2012. 1