# Depth and Transient Imaging with Compressive SPAD Array Cameras

Qilin Sun[1] Xiong Dun[1] Yifan Peng[2,1] Wolfgang Heidrich[1]
[1]KAUST [2]University of British Columbia

## Abstract

*Time-of-flight depth imaging and transient imaging are two imaging modalities that have recently received a lot of interest. Despite much research, existing hardware systems are limited either in terms of temporal resolution or are prohibitively expensive. Arrays of Single Photon Avalanche Diodes (SPADs) promise to fill this gap by providing higher temporal resolution at an affordable cost. Unfortunately SPAD arrays are to date only available in relatively small resolutions.*

*In this work we aim to overcome the spatial resolution limit of SPAD arrays by employing a compressive sensing camera design. Using a DMD and custom optics, we achieve an image resolution of up to 800×400 on SPAD Arrays of resolution 64×32. Using our new data fitting model for the time histograms, we suppress the noise while abstracting the phase and amplitude information, so as to realize a temporal resolution of a few tens of picoseconds.*

## 1. Introduction

Time-resolved imaging of light propagation effects has in recent years become a major research direction. From the now ubiquitous continuous wave ToF and RGB-D cameras, to more complex analysis of transient effects such as light-in-flight imaging [1, 47, 23], looking around the corner [21, 46, 24], or imaging in scattering media [14, 34, 25] the applications of high speed light transport analysis are vast.

Unfortunately, work on these problems has been hampered by hardware limitations. On the consumer end, inexpensive time of flight cameras are available for a few hundred dollars, but they provide a limited temporal resolution owing to modulation frequencies of only 10-130 MHz. On the high end, the combination of fast lasers and streak cameras provides temporal resolutions in the single digit picosecond range, but at a cost that is 2-3 orders of magnitude higher. Recently, single photon avalanche diodes have started to emerge as an alternative hardware solution that fills the gap both in terms of resolution and cost. SPADs are also attractive for their ability to detect very small signals, i.e. single photons. Unfortunately, image sensors built around SPAD technologies currently still suffer from low spatial resolution (e.g. $64 \times 32$ pixels).

In our work we aim to overcome the spatial resolution limit of SPAD arrays while preserving and even *enhancing* the time resolution by developing a compressive sensing SPAD camera. Since SPADs work with small signals that are inherently noisy, and compressive sensing is well known to amplify noise, devising such a compressive SPAD camera not only requires significant innovation on the camera hardware, but also a number of algorithmic contributions. Specifically, our contributions include

- the design and prototyping of a compact optical system containing imaging and re-imaging optics, as well as a DMD-based active modulation component.
- fill factor improvements of conventional SPAD sensors to enable the compressive sensing scenario by embedding a diffractive micro lens array (DMLA) in front of the bare sensors.
- inverse problem formulation to reconstruct high resolution 3D volumes from captured SPAD data.
- proposing a temporal PSF model based on the RC gate switching behavior of the electronics, and using it to deconvolve and sharpen the time profile of the transient image, achieving a temporal resolution in the range of tens of picoseconds.
- analysis and correction for phase distortions on the SPAD array caused by on-chip signal propagation differences.

## 2. Related Work

**Time-of-Flight Cameras** Continuous wave ToF cameras use the correlation between a sinusoidally modulated illumination source and a reference signal to measure the time delay due to light propagation [41, 38]. These types of cameras have been widely adapted in computer vision in the form of depth and RGB-D cameras. Recently, researchers have found new and exciting uses of this type of hardware beyond simple depth imaging, including transient imaging

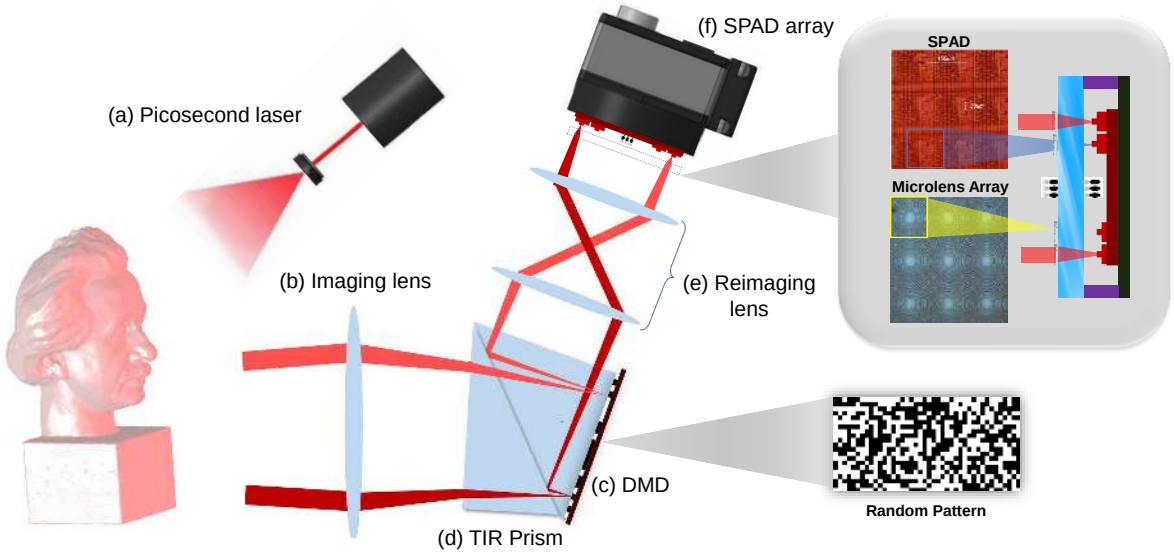


Figure 1. **Schematic of our compressive transient imaging system.** The scene is illuminated by a picosecond laser with a wavelength of 655 nm (a). Light reflected from the scene is imaged onto a digital micromirror array (c) using an imaging lens (b) with a focal length of 85 mm. A TIR prism (d) spatially separates the reflected light from the incident light, and an inverted 0.9x telecentric lens (e) is used to re-image the DMD onto a 64×32 SPAD array (f). An additional diffractive microlens array is employed in front of the SPAD array to overcome its limited pixel fill ratio. The focal length of the microlens array is designed as 1.035 mm for the illumination wavelength.

(e.g. [28, 23, 33]) and other uses (e.g. [44, 22, 43, 32]). As the modulation frequencies are usually in the range of 10-130 MHz, time resolutions and depth resolution are limited to the range of nanoseconds and centimeters, respectively.

**Transient and Light-in-Flight Imaging** Transient imaging started with the work by Abramson [1, 3], who utilized holographic technique to optically record and reconstruct the wavefront of a picosecond pulse. Due to the essential limitation of a holographic setup, only simple wavefronts were captured, but the methods were able to visualize effects such as reflection, interference, and focusing [2, 3] in 2D planes.

In recent years, ultra-fast cameras (e.g. streak camera) have been applied in the field of transient imaging for the investigations of looking around corners [21, 46] and the femto-photography [47]. Unfortunately, this setup suffers from very high cost, bulky hardware, and very long acquisition times. Low-budget transient imaging solutions [23, 33], using the continuous wave ToF cameras mentioned above, operate at the opposite end of the spectrum with significantly lower time resolution and cost.

**SPAD Arrays** Since ultrafast processes usually suffer from low photon counts, highly light sensitive, very fast camera equipment is needed for imaging such phenomena. In recent years, single-photon avalanche diodes (SPADs) have emerged as a viable technology. Although more expensive than ToF cameras, they are still an order of magnitude less expensive than streak cameras. To enable a SPAD array to record an ultrafast process, time-correlated single photon counting (TCSPC) [35, 39, 31] is introduced [17, 18, 42, 37], and has been widely applied in the area of fluorescence lifetime imaging. The idea behind this technique is similar to the equivalent-time sampling principle used in oscilloscopes. By repetitively measuring the time duration between a laser excitation pulse and the corresponding transient photon arrival, it is able to achieve typically sub-nanosecond resolution.

One limitation of current SPAD arrays is the low spatial resolution; the best available cameras today are in the low kilopixel range. To overcome this limitation, researchers have proposed to use a 2D translation setup to shift a 2D SPAD array with a fixed lens [42], or use a galvo mirror scenario to scan a 1D line SPAD camera [37]. Although these state-of-the-art methods yield a reasonable spatial resolution, their systems are bulky and the processes are time consuming. Thus, we seek to investigate in a computational imaging solution that not only maintains the advantages of SPAD sensor — the ultra high temporal resolution and the single photon level sensitivity, but also enhances its spatial resolution.

**Compressive Sensing and Imaging** After being first proposed in the 1970s [8], compressive sensing (CS) theory was gradually developed over the next few decades [29, 19, 40]. The mathematical foundation of CS was first laid out by Donoho [10] and Candes et al. [6] and can be traced fur-

ther back to sparse recovery works [13, 12, 11]. After the single pixel camera [15] had been invented, it became possible to replace conventional sampling and reconstruction operations with a more general linear measurement schemes coupled with optimization methods [9, 16, 36]. CS approaches are particularly beneficial in scenarios where high resolution sensor arrays are either technologically infeasible or prohibitively expensive, as is the case for SPAD sensors. For instance, a mask shifting camera [48] and a CS-based infra-red camera [7] have been realized. These works inspire us to apply CS approach into transient or depth imaging to acquire high frequency spatial information.

## 3. System Overview

Our imaging system is an integration of optics, mechanics, electronics, and computation, as illustrated in Figure 1. The key components of the the hardware setup are a SPAD array with a resolution of ($n$=64)×($m$=32) pixels that provides a high temporal resolution in combination with illumination from a picosecond laser, as well as a Digital Micromirror Device (DMD) in combination with re-imaging optics that achieve improved spatial resolution through a compressive sensing approach. In the following we explain the high level operation of the system by examining the spatial and temporal resolution characteristics.

**The Temporal Resolution** of the system is determined by the ability to measure time delays between the emission of a pulse by the the picosecond laser, and the time the reflected light is received at the SPAD array. For a given exposure interval, called the *gate width*, which can be as short as 200ps, each SPAD pixel reports whether or not a photon was received during this time interval. Moreover, the start time of the exposure interval can be shifted with a precision of 20ps. By repeatedly emitting pulses from the laser while adjusting this time offset, we can sample the transient propagation of light in the scene. In our case, the SPAD array was operated in TCSPC mode with 20ps shift per circle and 830ps gate width to improve the light efficiency of the system compared to the shortest gate width of 200ps. Although this gate width should normally limit the temporal resolution of the system, we show in Section 4.3 how to suppress the low-pass effect of the gate signal and recover a temporal resolution in the 10s of picoseconds. We note that, since each measurement is a binary event, measurements need to be repeated multiple times to reduce noise and obtain estimates of intensity for each phase delay.

**The Spatial Resolution** of the system is primarily determined by the resolution of the DMD instead of the much lower resolution of the SPAD array. Due to the re-imaging optics shown in Figure 1, the DMD is in a conjugate plane with the image sensor, so that a block of neighboring DMD pixels is mapped onto a single SPAD pixel. By cycling through random binary patterns on the DMD, we can implement a compressive sensing super-resolution scheme where each SPAD pixel can be interpreted as a single pixel camera [5] responsible for a narrow part of the total field of view. Due to some perspective foreshortening, as well as incomplete utilization of the DMD area, the final spatial resolution of our system is $N = 800 \times M = 400$.

Note that in the above description, we assume that each SPAD pixel integrates over a large block DMD pixels. This only holds if the fill factor of the SPAD array approaches one, i.e. if the light sensitive area of each pixel is as big as the pixel spacing. Unfortunately, this does not hold true for commercially available SPAD arrays today. Our SPAD array has a pixel spacing of $150\mu$m, but the active area is only $30\mu$m squared. To overcome this challenge, we designed a diffractive microlens array that focuses the light from a block of DMD pixels onto the active area of a SPAD pixel.

## 4. Model and Optimization

### 4.1. Observation Model

As mentioned in the previous section, we assume a SPAD array with $n \times m$ pixels and seek to reconstruct spatially super-resolved transient or depth images with a resolution of $N \times M$ with $N = pn, M = pm$. At each pixel, the measurement procedure cycles through $T$ phase offsets, and for each offset $K$ measurements are taken. With this setup, each SPAD pixel is essentially a $p \times p$ single pixel camera similar to the work by Duarte et al. [15]. As such, the measurements from each SPAD pixel could in principle be reconstructed independently into a transient superpixel, although tiling artifacts must then be addressed.

Thus, we can represent the captured data of a high resolution depth image or a transient frame as follows:

$$\boldsymbol{Y} = \boldsymbol{\Psi}(\boldsymbol{X}), \tag{1}$$

where $\boldsymbol{Y} \in R^{K \times T \times n \times m}$ is the observed 4D data after modulation, $\boldsymbol{X} \in R^{T \times N \times M}$ is the 3D signal under evaluation, and $\boldsymbol{\Psi}$ is an operator that maps the random patterns to individual pixels at each layer.

Instead of assuming sparsity in the transient image itself, we reasonably assume the gradient distributions to be sparse. This leads to a 3D total variation (TV) regularizer, with different weights $\lambda_{1,2}$ and $\lambda_3$ for the two spatial dimensions and the temporal dimension, respectively:

$$\hat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}} \frac{1}{2} \|\boldsymbol{\Psi}(\boldsymbol{X}) - \boldsymbol{Y}\|_2^2 + \sum_i \lambda_i \boldsymbol{D}_i(\boldsymbol{X}), \tag{2}$$

with

$$\begin{cases} \boldsymbol{D}_{1,2}(\boldsymbol{X}) = \|\nabla_s \boldsymbol{X}\|_1 \\ \boldsymbol{D}_3(\boldsymbol{X}) = \|\nabla_\tau \boldsymbol{X}\|_1 \end{cases}. \tag{3}$$

Here, $\boldsymbol{D}_{1,2}$ and $\boldsymbol{D}_3$ represent the total variation in the spatial domain and temporal domain, respectively. Although the TV regularizer ties together all the $N \times M$ pixels in the output image, we choose to split the problem such that each transient superpixel is reconstructed independently with a TV prior, and then the TV regularization across the superpixels is enforced separately. This choice allows for a highly parallelized implementation of the reconstruction algorithm.

### 4.2. Reconstruction Algorithm

We apply a modified version of TVAL3 [30] to reconstruct a spatially super-resolved depth or transient image from the captured data. Note that the same method can also be used to simply reconstruct a stationary intensity image by integrating along the time axis and setting $\lambda_3 = 0$.

Solving the optimization problem, we can rewrite (2) as

$$\begin{aligned} \hat{\boldsymbol{X}} &= \arg\min_{\boldsymbol{X},\boldsymbol{w}} \Sigma_i \|\boldsymbol{w}_i\|_1 \\ s.t. &\begin{cases} \boldsymbol{D}_i(\boldsymbol{X}) = \|\boldsymbol{w}_i\|_1 \\ \frac{1}{2}\|\boldsymbol{\Psi}(\boldsymbol{X}) - \boldsymbol{Y}\|_2^2 < \epsilon, \end{cases} \end{aligned} \tag{4}$$

which results in the following Augmented Lagrangian:

$$\begin{aligned} &\mathcal{L}\{\boldsymbol{w}, \boldsymbol{X}, \boldsymbol{\sigma}, \boldsymbol{\delta}\} = \Sigma_i \|\boldsymbol{w}_i\|_1 \\ &- \boldsymbol{\sigma}^T(\boldsymbol{D}(\boldsymbol{X}) - \boldsymbol{w}) - \boldsymbol{\delta}^T(\boldsymbol{\Psi}((\boldsymbol{X}) - \boldsymbol{Y})) \\ &+ \frac{\beta}{2}\|\boldsymbol{D}(\boldsymbol{X}) - \boldsymbol{w}\|_2^2 + \frac{\zeta}{2}\|\boldsymbol{\Psi}((\boldsymbol{X}) - \boldsymbol{Y})\|_2^2. \end{aligned} \tag{5}$$

We seek to minimize the above objective using an alternating TVAL3 solver. We alternate between updating $\boldsymbol{X}^k$, $\boldsymbol{w}^k$, and the Lagrangian multipliers $\boldsymbol{\sigma}$ and $\boldsymbol{\delta}$.

As is illustrated in Algorithm 1, we firstly fix $\boldsymbol{w}^k$ to update $\boldsymbol{X}^{k+1}$ by doing a gradient decent to $\mathcal{L}\{\boldsymbol{w}, \boldsymbol{X}, \boldsymbol{\sigma}, \boldsymbol{\delta}\}$. Hence, $\boldsymbol{X}^{k+1}$ is updated as:

$$\boldsymbol{X}^{k+1} = \boldsymbol{X}^k - \alpha \partial_{\boldsymbol{X}} \mathcal{L}\{\boldsymbol{w}, \boldsymbol{\sigma}, \boldsymbol{\delta}, \boldsymbol{\beta}, \zeta\}. \tag{6}$$

Where $\alpha$ is obtained by Amijo's line search[20]. We then further derive $\boldsymbol{w}^{k+1}$ by a shrinkage process with $\boldsymbol{X}^{k+1}$. The Lagrangian multipliers $\boldsymbol{\sigma}$ and $\boldsymbol{\delta}$ are updated by:

$$\begin{cases} \boldsymbol{\sigma}^{k+1} = \boldsymbol{\sigma}^k - \beta(\boldsymbol{D}(\boldsymbol{X}^{\boldsymbol{k+1}}) - \boldsymbol{w}^{k+1}) \\ \boldsymbol{\delta}^{k+1} = \boldsymbol{\delta}^k - \zeta(\boldsymbol{\Psi}(\boldsymbol{X}^{\boldsymbol{k+1}}) - \boldsymbol{Y}). \end{cases} \tag{7}$$

This process is iterated until convergence.

For details of minimizing the Augment Lagrangian objective using our modified TVAL3 algorithm, refer to the supplementary document. We further use a VST-based denoising method [4] to denoise the results and correct the final depth image or transient frames. On average, the reconstruction takes around 10 minutes. Notice that this correction uses a calibrated phase delay map which is to be discussed in next section.

**Algorithm 1:** Reconstruction Algorithm

**Input:** $\Psi$, $\boldsymbol{Y}$, opts
**Result:** $\hat{\boldsymbol{X}}$
**while** $\|\boldsymbol{X}_p - \boldsymbol{X}\|_2 > tol$ **do**
  **Step 1**. $\boldsymbol{X}_p = \boldsymbol{X}^k$
  **Step 2**. Fix $\boldsymbol{w}^k$, do Gradient Descent
  to $\mathcal{L}\{\boldsymbol{w}^k, \boldsymbol{X}, \boldsymbol{\sigma}, \boldsymbol{\delta}\}$
    a) compute step length $\tau > 0$ by **BB** rules
    b) determine $\boldsymbol{X}^{k+1}$ by
      $\boldsymbol{X}^{k+1} = \boldsymbol{X}^k - \alpha\tau\partial_{\boldsymbol{X}}\mathcal{L}\{\boldsymbol{w}^k, \boldsymbol{X}^k, \boldsymbol{\sigma}^k, \boldsymbol{\delta}^k\}$
  **Step 3**. compute $\boldsymbol{w}^{k+1}$ by shrinkage
    $\boldsymbol{w}^{k+1}$ = shrink$(\boldsymbol{D}(\boldsymbol{X}^{\boldsymbol{k+1}}) - \boldsymbol{\sigma}/\beta, 1/\beta)$
  **Step 4**. update Lagrangian Multipliers by
    $\boldsymbol{\sigma}^{k+1} = \boldsymbol{\sigma}^k - \beta(\boldsymbol{D}(\boldsymbol{X}^{\boldsymbol{k+1}}) - \boldsymbol{w}^{k+1})$
    $\boldsymbol{\delta}^{k+1} = \boldsymbol{\delta}^k - \zeta(\boldsymbol{\Psi}(\boldsymbol{X}^{\boldsymbol{k+1}}) - \boldsymbol{Y})$

### 4.3. Data Pre-processing

**Calibrating Hot pixels and Background Noise** Compressive sensing is well-known to act as a noise amplifier, since it exhibits more sensitivity to signal changes compared to direct scanning approaches. Although the captured 3D datasets allow for efficient processing in both spatial and temporal dimensions [18], the dark counts and the background noise will heavily affect the reconstruction quality. A part of the raw data captured with one random pattern is shown in Figure 2a. We observe that it is heavily corrupted by dark counts and ambient light, which result from the hot pixels of the sensor and the low contrast ratio of the DMD, respectively.

To solve this problem, we first obtain the average ambient light plus dark-counts rates (denoted as $\mathcal{HB}$) for all pixels as shown in Figure 2b, similar to the work done by Shin [42]. Figure 2c indicates that the level of dark counts can reach around 1/3 of the signal at a few positions of hot pixels. After subtracting $\mathcal{HB}$, the noise in temporal domain can be drastically suppressed.

**Sharpening with Temporal PSF Model** To improve the temporal resolution[45] of our reconstructions we next model the temporal PSF of the acquisition system, and use it to fit/deconvolve the sensor data *before* solving the compressed sensing problem from Section 4.2.

The laser source generates pulses with a specific time profile that is well approximated as a Gaussian function. Notice that the laser pulse duration setting (i.e. 80ps in our

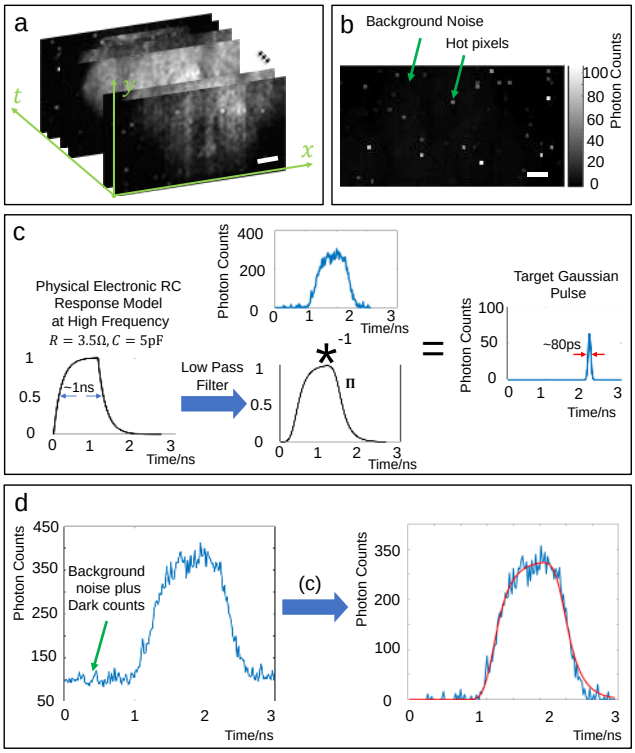


Figure 2. **Raw data and the pre-processing.** (a) Part of the raw data under the first random pattern. (b) Calibrated hot pixels plus background noise for each frame. (c) Processing of abstracting phase and amplitude information through our fitting model. (d) Time histogram of one pixel of (a) and its dark counts and background noise. After pre-processing, the clean time histogram(red curvature) is shown. Scale bars in (a) and (b): 10mm

experiments) is significantly shorter than the time bin duration $\Delta = 830$ps of the SPAD camera. Considering the distortion of the SPAD's gate signal, the time jitter of the synchronizing signal, and the convolution of the gate signal and the laser pulse, the received time histogram of the signals has a FWHM of around 1ns. Due to the high frequency of the gate signal, the response of the signal should be distorted close to a resistorcapacitor circuit (RC) response, as shown in Figure 2c. Partial high frequency components of the gate signal are cut off with a Gaussian low pass filter ($\sigma_f = 5$, kernel size=30, unit: 20ps). The parameters of $R$, $C$ and the Gaussian filter are global and manually set to obtain $\mathbf{\Pi}$(for more details, refer to the supplementary document). The final model for the temporal PSF of the system is shown in Figure 2d.

With this temporal PSF model we can remove the low-pass filtering effect of the gate signal from the raw data *before* solving the compressive sensing problem.

Specifically, we represent the sharpened sensor data $\boldsymbol{Y_i}$ for each pixel $i$ as a Gaussian $\boldsymbol{G}(t; A, \mu) = A\boldsymbol{e}^{-\frac{(t-\mu)^2}{2\sigma^2}}$, where the parameters $A, \mu$ are determined by solving the parametric deconvolution problem

$$\min_{A,\mu} \|\mathbf{G}(t; A, \mu) \star \mathbf{\Pi}(t) - \hat{\boldsymbol{Y}}\|_2^2, \tag{8}$$

and $\hat{\boldsymbol{Y}}$ is the raw sensor data. As we demonstrate in the experiments, this *parametric* deconvolution is robust despite the large size of the large duration of the gate signal compared to the pulse width of the laser, so that we can achieve temporal resolutions in the tens of picoseconds.

## 5. Hardware Details

Figure 1 illustrates the schematic of our imaging technique that mainly entails a picosecond laser source for illumination, a programmable optical device for the modulation of high frequency spatial information, multiple auxiliary lenses for imaging and re-imaging, and a SPAD sensor array for detection.

**Illumination and Detection** We used an MPD-SPC3 SPAD array as the detector. The picosecond laser source (PicoQuant LDH P-650), with the center wavelength of 655nm and a peak power of 360mW (pulse energy 40pJ), was operated at around 80ps pulse duration with 50MHz repetition rate (triggered by the SPAD camera). The SPAD array was operated in TCSPC mode with 20ps shift per cyrcle and 830ps gate width. We set the integration time as 52$\mu$s and summed up 1280 frames before read out. We in total sent around 3.3 million pulses per frame under one pattern and received a maximum of ~40-60 photons per pixel. The laser beam was scattered by a diffuser and reconcentrated by a 30mm biconvex lens to illuminate the scene. The capture process including background calibration together lasted around 9 minutes.

We note that if we use a laser source with 40nJ or higher pulse energy and with 450nm center wavelength, the capture process can theoretically be completed within 1s. Although such a high powered laser could be dangerous when collimated, the laser in our setup is diffused into a spotlight, such that with proper engineering it can be safe.

**Auxiliary Optics** We used a Canon 85mm lens as the imaging lens. At its focal plane, the images were modulated by a digital micro mirror device (DMD, TI DLP4500). We used telecentric architectures [26] and placed a TIR-prism (Young Optics) in front of the DMD. An inversed Edmund Platinum 0.9X double-side telecentric lens was applied as the reimaging lens. Eventually, we achieve a spatial resolution improvement by a factor of 12.5. For more details, please refer to supplementary document.

**Diffractive Microlens Array** Our designed DMLA has a focal length of 1.035mm and works in a finite conjugate

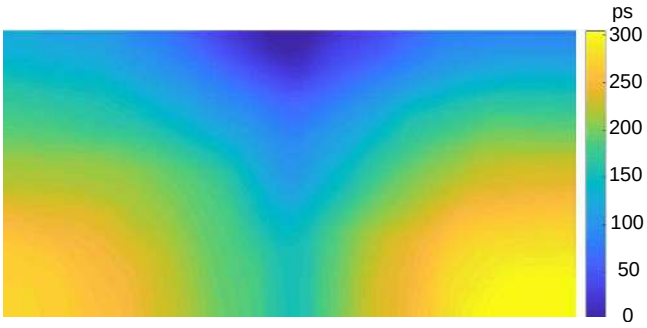


Figure 3. **Calibrated spatial variant phase delay map.**

configuration [27]. It was carefully aligned with the SPAD sensor using FINEPLACER femto. With this process, we managed to improve the fill factor from 3.14% to 52.87%. For more details, please refer to supplemental document.

**Calibrating Spatially Variant Behavior of SPAD array** We discovered that our SPAD array suffers from a spatial variant phase delay of the captured data. In other words, there exist different time delays for the gate signals corresponding to each pixel, presumably due to differences in on-chip signal travel times. One possible reason is that when the SPAD array is operated at TCSPC mode, the picosecond level phase shift per cycle results in a more sensitive behavior of the trace length in electronics(either SPAD sensor chip or peripheral circuit). We use a whiteboard that is placed ∼1m away to calibrate the spatial variant phase delay of the SPAD sensor. Figure 3 visualizes of the reconstructed high resolution spatial variant phase delay. Fortunately, the fixed spatial variant behavior will not influence the reconstruction quality and will only cause a fixed phase offset that can be subtracted from the final reconstructions.

**System Workflow** During the capture process, first a random pattern frame is sent to DLP4500, followed by a trigger signal to display the pattern on the DMD. Once the DMD shows the pattern, the SPAD array is used to capture a sequence of data frames. The SPAD array triggers the emission of laser pulses and then starts its own photon integration with a fixed time offset. After sufficient frames have been captured for a given phase offset, the SPAD camera shifts the gate window (i.e. 20ps delay regarding the rising edge of the clock signal) and captures another sequence of frames until covering all the phases. After all the data has been captured, it is transferred back to the controlling system. The same process then continues with a new DMD pattern, until all patterns have been covered (i.e. 100 patterns in our setting, CS ratio 0.64). A work flow diagram and detailed explanation are provided in the supplementary document.

## 6. Results and Analysis

**Spatial Resolution Analysis** In order to visualize the improvement of spatial resolution, we capture a stand UASF 1951 passive resolution chart that is placed ∼1m away from the imaging lens.

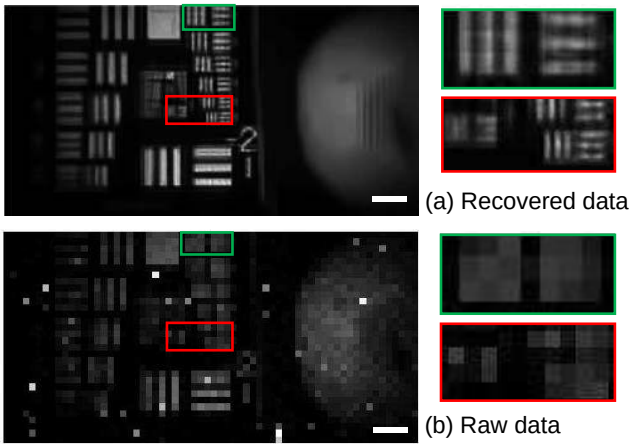


Figure 4. **Visualization of spatial resolution improvement tested on a standard resolution chart.** (a) Reconstructed intensity image. (b) The directly up-sampled summation of raw data under the first random pattern over time axis. Scale bars:10mm

Figure 4b shows that the direct up-sampled summation of raw data not only suffers from low resolution but also from heavy noise. The area marked by the green rectangle cannot be resolved. In comparison, our reconstructed intensity image indicates a significant improvement in spatial resolution at least 8 times as shown in Figure 4a, even the fine lines marked by the red rectangle can be distinguished.

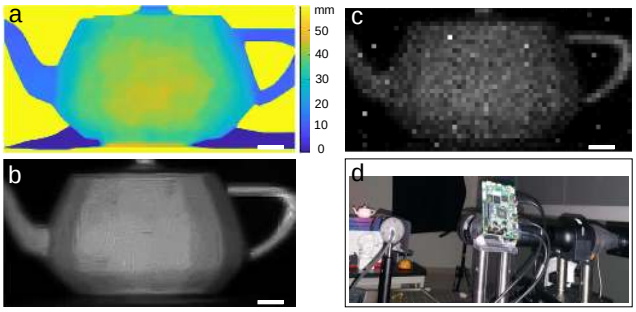


Figure 5. **Experimental reconstructed depth image and intensity image.** (a) Reconstructed depth image. (b) Reconstructed intensity image. (c) The direct up-sampled summation of raw data under the first random pattern over time axis. (d) Photograph of the setup while capturing the teapot scene. Scale bars: 10mm

**Depth Imaging and Intensity Imaging** We start from a simple example of capturing a teapot, with the results presented in Figure 5. The 3D printed tea pot is placed ∼1m

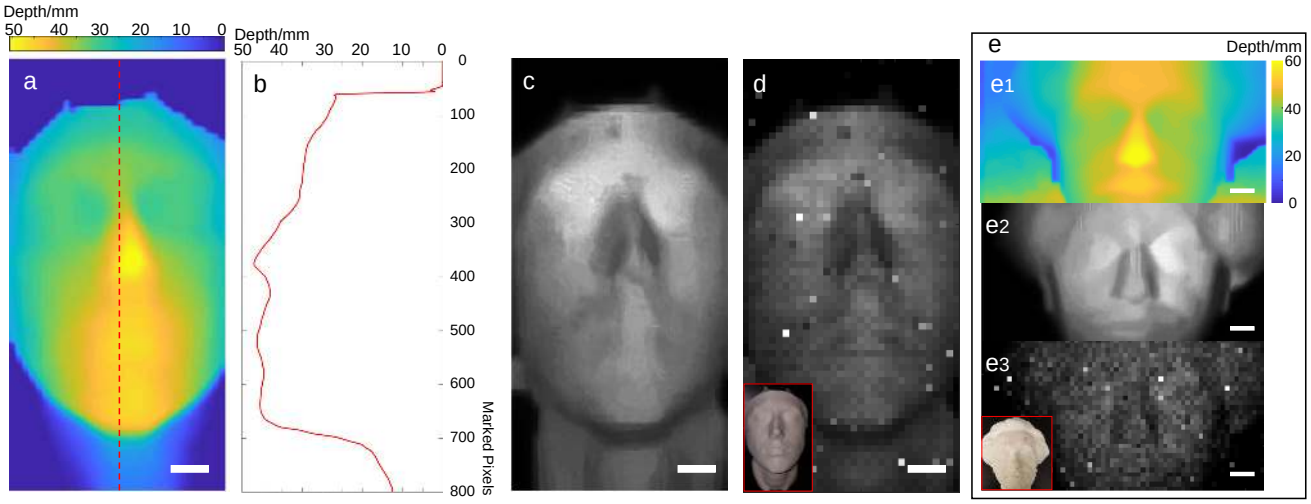


Figure 6. **Experimental reconstructed depth image and intensity image.** (a) Reconstructed depth image. (b) Vertical cross section plot of reconstructed depth image, marked with a red dash line in (a). (c) Reconstructed intensity image. (d) The direct up-sampled summation of raw data under the first random pattern over time axis and scene (e) Similar data for a different face model. Scale bars: 10mm

away and has a height of approximately 6cm, as shown in Figure 5d. As can be seen, the reconstructed depth image and intensity image exhibit details that are indistinguishable in raw data, for instance the edges of the teapot handle and the teapot body. Note that the dark area on the depth image is due to a fact that the phase of reflected light by the stand under the teapot is below the front depth threshold. Figure 6 presents the experimental results of capturing a 3D printed human face model. The visualization of the direct up-sampled summation of raw data is quite noisy and highly corrupted by dark counts. With the sharpening process from Section 4.1, we mostly suppress the noise especially the dark counts in both spatial and temporal domain.

In the experiment, we show the ability to distinguish the details of a 3D printed human face model whose total width is 6cm only. In this configuration, we have achieved a depth resolution of millimetres scale. To further explore the capability of distinguishing fine details, we use a more complex model (Einstein by Artur Loewenthal from 1930). The printed model has a total width of 12cm. The depth image and intensity image (Figure 6e) validate that after our reconstruction the details even the mustache under the nose can be visually distinguished. Note that the physical height difference between the mustache and the lip is only a few millimetres.

Figure 7 presents the results of capturing light-in-flight. As can be seen in the last image of Figure 7, milky water fills up the bottom two thirds of the glass pipe, while the top third is filled with air. We can observe an interface between the two illuminated parts. Light propagates from the left in this scattering medium and reflects both at the side walls of the pipe and in a mirror at the end of the pipe.

Throughout the sequence of time steps, we can see a main pulse propagating with a width of about 20mm. Given the speed of light in water, this corresponds to a pulse duration of ca. 80ps, which matches well with the physical parameters of the used laser source. This result shows that the sharpening process from Section 4.3, which acts only on the time axis of the transient image, also produces results that are meaningful and consistent along the spatial dimensions.

From 0ps to 200ps, the scattering effect turns more significant and the 'tail' gradually expands away from the main light beam. At the same time, the main light beam continues to decay. At 400ps, the scattered 'tail' starts reaching the interface of the milky water and the photons travel out of the milky water. The photons escaping from the milky water become invisible because the scattering of air is very small and so no photons are scattered towards the camera. At 700ps, the photons traveling out of the milky water are reflected back by the internal surface of the glass pipe. Part of the reflected light reaches the interface of milky water again and emits a bright line at 800ps. Note that the light propagates 1.33 times faster in the air than in the milky water, resulting in fact from 700ps to 800ps the light disappears faster in the air. Refer to the supplementary video and digital versions of the images for better visualization.

**Transient or Light-in-flight Imaging** Figure 8 presents another result for capturing a transient scene. A cube and a triangular pyramid are placed $\sim$1m away from the imaging system. Both of them have an edge length of 3cm. We also put two mirrors in the scene, shown in Figure 8.

From 0ps to 240ps, a wavefront passes over the objects, resulting first in a gradual increase and then a gradual de-

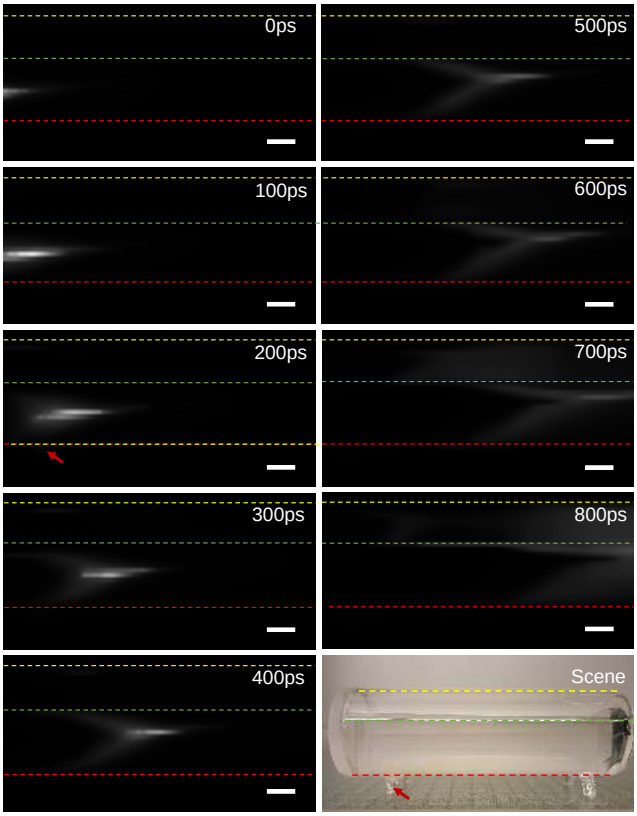


Figure 7. **Compressive light-in-flight imaging.** From left to right and from top to bottom are the selected frames of our reconstructed light-in-flight video stream. We here present one frame of every 5 frames. The light travels via a glass pipe filled with 2/3 of milky water as shown in the top view of the scene. Scale bars: 10mm

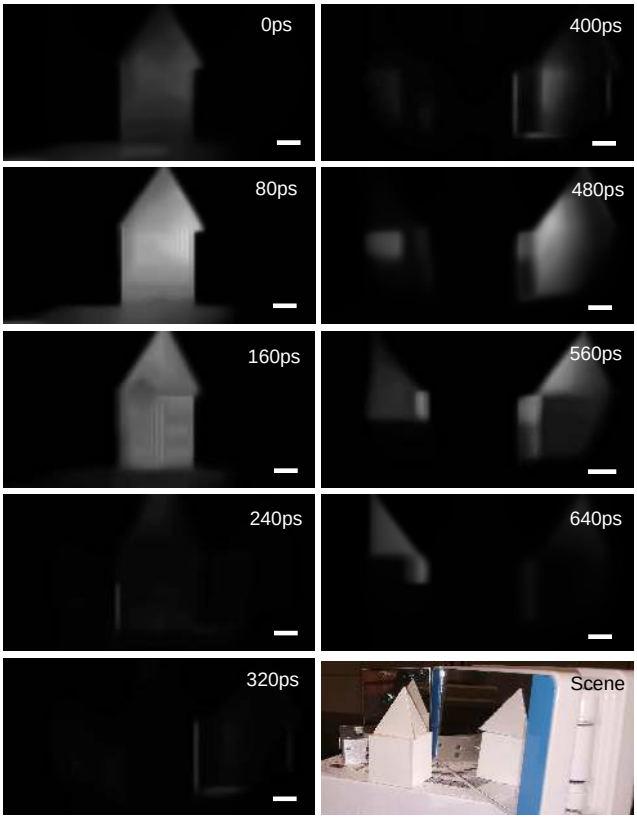


Figure 8. **Compressive transient imaging.** From left to right and from top to bottom are the selected frames of our reconstructed light-in-flight video stream. We here present one frame of every 4 frames. The light illuminates the scene and reflected by 2 mirrors as shown in the scene. Scale bars: 10mm

crease of their illumination. Later, the light propagates from the objects to the right mirror and then is reflected to be captured by the sensor. At 320ps, the light reflected by the left mirror starts reaching the sensor. As the light keeps on propagating, more light is reflected from the left mirror and the right side starts decaying. After 400ps, the light has propagated away from the right mirror and continues to hit the left mirror. From 320ps to 400ps, we can tell the distance of the object from the left mirrors is a little bit larger than the right side, which is in agreement with the scene. Overall, we have successfully captured and reconstructed the high resolution light propagation indicating the light hitting the objects and bouncing back by the mirrors at two sides. Refer to the supplementary video for better visualization.

## 7. Conclusion

In conclusion, we have demonstrated the use of SPAD arrays in a compressive sensing system for transient imaging and depth images. Our approach for the first time overcomes the spatial resolution limit of SPAD arrays with a fast capturing process without mechanical scanning. The achieved temporal resolution lies in the range of tens of picoseconds and thus surpasses that achieved with common time-of-flight cameras.

In terms of optical design, we have increased the fill factor of the SPAD array by prototyping a DMLA in front of it. In addition, the DMD's telecentric architecture is applied in our imaging path to enable a great degree of freedom to modulate the light. On the computational side, we devise a combination of compressed sensing reconstruction with model-based temporal sharpening to counteract the low-pass nature of the SPAD gate signal. We further formulate an optimization that takes account of the edge effect of volumes. The combination of these approaches yields high quality intensity image, depth image and transient frames with time resolutions of tens of picoseconds.

## Acknowledgment

This work is supported by the KAUST baseline funding. The authors thank MPD for providing supports of a research and development SPAD sensor. We thank Prof. Xiaohang Li for providing supports of experimental environment.

## References


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

[2] N. Abramson. Light-in-flight recording: high-speed holographic motion pictures of ultrafast phenomena. *Applied optics*, 22(2):215–232, 1983. 2

[3] N. H. Abramson and K. G. Spears. Single pulse light-in-flight recording by holography. *Applied optics*, 28(10):1834–1841, 1989. 2

[4] L. Azzari and A. Foi. Variance stabilization for noisy+estimate combination in iterative poisson denoising. *IEEE Signal Processing Letters*, 23(8):1086–1090, Aug 2016. 4

[5] R. G. Baraniuk. Compressive sensing. *IEEE Signal Processing Magazine*, 24(4), 2007. 3

[6] E. J. Candes, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math*, 59(8):1207–1223, 2006. 2

[7] H. Chen, M. S. Asif, A. C. Sankaranarayanan, and A. Veeraraghavan. Fpa-cs: Focal plane array-based compressive imaging in short-wave infrared. In *Proc. CVPR*, pages 2358–2366. IEEE, 2015. 3

[8] J. F. Claerbout and F. Muir. Robust modeling with erratic data. *Geophysics*, 38(5):826–844, 1973. 2

[9] W. Dai and O. Milenkovic. Subspace pursuit for compressive sensing signal reconstruction. *IEEE Transactions on Information Theory*, 55(5):2230–2249, 2009. 3

[10] D. L. Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006. 2

[11] D. L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization. *Proceedings of the National Academy of Sciences*, 100(5):2197–2202, 2003. 3

[12] D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Transactions on Information Theory*, 47(7):2845–2862, 2001. 3

[13] D. L. Donoho and P. B. Stark. Uncertainty principles and signal recovery. *SIAM Journal on Applied Mathematics*, 49(3):906–931, 1989. 3

[14] A. Dorrington, J. Godbaz, M. Cree, A. Payne, and L. Streeter. Separating true range measurements from multi-path and scattering interference in commercial range cameras. In *Proc. Electronic Imaging*, 2011. 1

[15] M. F. Duarte, M. A. Davenport, D. Takbar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2):83–91, 2008. 3

[16] J. H. Ender. On compressive sensing applied to radar. *Signal Processing*, 90(5):1402–1414, 2010. 3

[17] G. Gariepy, N. Krstajić, 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, 2015. 2

[18] G. Gariepy, F. Tonolini, R. Henderson, J. Leach, and D. Faccio. Detection and tracking of moving objects hidden from view. *Nature Photonics*, 10(1):23–26, 2016. 2, 4

[19] E. D. Gluskin. Norms of random matrices and widths of finite-dimensional sets. *Mathematics of the USSR-Sbornik*, 48(1):173, 1984. 2

[20] L. Grippo, F. Lampariello, and S. Lucidi. A nonmonotone line search technique for newtons method. *SIAM Journal on Numerical Analysis*, 23(4):707–716, 1986. 4

[21] O. Gupta, T. Willwacher, A. Velten, A. Veeraraghavan, and R. Raskar. Reconstruction of hidden 3d shapes using diffuse reflections. *Optics Express*, 20(17):19096–19108, 2012. 1, 2

[22] F. Heide, W. Heidrich, M. Hullin, and G. Wetzstein. Doppler time-of-flight imaging. *ACM Trans. Graph. (SIGGRAPH)*, 34(4):36, 2015. 2

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

[24] F. Heide, L. Xiao, W. Heidrich, and M. t. B. Hullin. Diffuse mirrors: 3D reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors. In *Proc. CVPR*, June 2014. 1

[25] 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(21):26338–26350, 2014. 1

[26] T. Instruments. Dlp system optics application note, 2015. 5

[27] G. Intermite, A. McCarthy, R. E. Warburton, X. Ren, F. Villa, R. Lussana, A. J. Waddie, M. R. Taghizadeh, A. Tosi, F. Zappa, et al. Fill-factor improvement of si cmos single-photon avalanche diode detector arrays by integration of diffractive microlens arrays. *Optics Express*, 23(26):33777–33791, 2015. 6

[28] 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 Transactions on Graphics (ToG)*, 32(6):167, 2013. 2

[29] B. S. Kashin. Diameters of some finite-dimensional sets and classes of smooth functions. *Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya*, 41(2):334–351, 1977. 2

[30] C. Li, W. Yin, and Y. Zhang. Users guide for tval3: Tv minimization by augmented lagrangian and alternating direction algorithms. *CAAM report*, 20:46–47, 2009. 4

[31] D.-U. Li, J. Arlt, J. Richardson, R. Walker, A. Buts, D. Stoppa, E. Charbon, and R. Henderson. Real-time fluorescence lifetime imaging system with a 32× 32 0.13 $\mu$m cmos low dark-count single-photon avalanche diode array. *Optics Express*, 18(10):10257–10269, 2010. 2

[32] F. Li, A. P. Huaijin Chen, C. Yeh, K. He, A. Veeraraghavan, and O. Cossairt. Cs-tof: High-resolution compressive time-of-flight imaging. *Optics Express*, 25(25):31096–31110, 2017. 2

[33] J. Lin, Y. Liu, J. Suo, and Q. Dai. Frequency-domain transient imaging. *IEEE transactions on pattern analysis and machine intelligence*, 39(5):937–950, 2017. 2

[34] 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. Graph. (SIGGRAPH Asia)*, 30(6):171:1–171:10, 2011. 1

[35] D. O'Connor. *Time-correlated single photon counting*. Academic Press, 2012. 2

[36] G. Oliveri, N. Anselmi, and A. Massa. Compressive sensing imaging of non-sparse 2d scatterers by a total-variation approach within the born approximation. *IEEE Transactions on Antennas and Propagation*, 62(10):5157–5170, 2014. 3

[37] M. OToole, F. Heide, D. B. Lindell, K. Zang, S. Diamond, and G. Wetzstein. Reconstructing transient images from single-photon sensors. In *Proc. IEEE CVPR*, pages 2289–2297, 2017. 2

[38] A. Payne, A. Daniel, A. Mehta, B. Thompson, C. S. Bamji, D. Snow, H. Oshima, L. Prather, M. Fenton, L. Kordus, et al. 7.6 a 512× 424 cmos 3d time-of-flight image sensor with multi-frequency photo-demodulation up to 130mhz and 2gs/s adc. In *Proc. ISSCC*, pages 134–135. IEEE, 2014. 1

[39] 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. In *Proc. CICC*, pages 77–80. IEEE, 2009. 2

[40] F. Santosa and W. W. Symes. Linear inversion of band-limited reflection seismograms. *SIAM Journal on Scientific and Statistical Computing*, 7(4):1307–1330, 1986. 2

[41] R. Schwarte, Z. Xu, H. Heinol, J. Olk, R. Klein, B. Buxbaum, H. Fischer, and J. Schulte. New electro-optical mixing and correlating sensor: facilities and applications of the photonic mixer device. In *Proc. SPIE*, volume 3100, pages 245–253, 1997. 1

[42] D. Shin, F. Xu, D. Venkatraman, R. Lussana, F. Villa, F. Zappa, V. K. Goyal, F. N. Wong, and J. H. Shapiro. Photon-efficient imaging with a single-photon camera. *Nature Communications*, 7, 2016. 2, 4

[43] S. Su, F. Heide, R. Swanson, J. Klein, C. Callenberg, M. Hullin, and W. Heidrich. Material classification using raw time-of-flight measurements. In *Proc. CVPR*, pages 3503–3511, 2016. 2

[44] S. Surti. Update on time-of-flight pet imaging. *Journal of Nuclear Medicine*, 56(1):98–105, 2015. 2

[45] R. Tadano, A. K. Pediredla, K. Mitra, and A. Veeraraghavan. Spatial phase-sweep: Increasing temporal resolution of transient imaging using a light source array. In *Proc. ICIP*, pages 1564–1568. IEEE, 2016. 4

[46] 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*, 3:745, 2012. 1, 2

[47] 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. Graph. (SIGGRAPH)*, 32(4):44, 2013. 1, 2

[48] X. Yuan, P. Llull, X. Liao, J. Yang, D. J. Brady, G. Sapiro, and L. Carin. Low-cost compressive sensing for color video and depth. In *Proc. CVPR*, pages 3318–3325, 2014. 3