Dr

240*320*2000*2*30 / (1000

Gated has 40 codes, Proposed

SBR == 0.5, Mean photon

Sing
ing pixe
that can
picosec
based 3
pulse b
tamps.
based c
ceed th
One mc
ation is
lution a
we prop
CSPHs
high-res
photon
coding
operatic
3D ima
backgro
tups. O
sistently
out com

1. Intr

Sing
technol
photons
SPCs b
can also
high (s
SPAD-b
tary me
process
resoluti
these capabilities, SPAD-based SPCs are gaining popularity in various imaging applications including 3D imaging [14, 48, 49, 55], passive low-light imaging [3, 38, 51], HDR imaging [29, 30], non-line-of-sight (NLOS) imaging [5, 36, 58], fluorescence lifetime imaging (FLIM) microscopy [61], and diffuse optical tomography [37, 60].

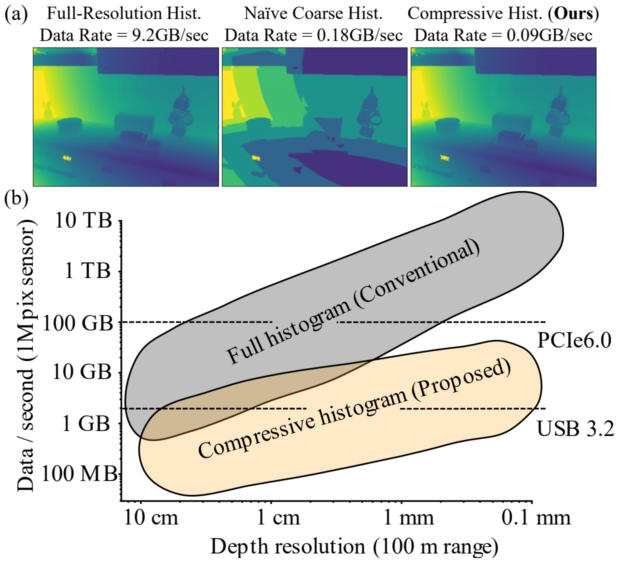


Figure 1. **Compressive Single-Photon 3D Imaging.** (a) Example depth maps with conventional (full histogram) capture, coarse resolution capture and our method with compressive capture. In this simulation, our method generates lower data, yet generates depth maps that are visually indistinguishable from the conventional method. (b) With conventional acquisition schemes, data bandwidth requirements scale linearly with the desired depth resolution. Our proposed compressive acquisition does not scale as strongly with depth resolution, keeping the output data rates manageable with existing data transfer standards like USB and PCIe.

Unlike a conventional camera pixel that outputs a single intensity value integrated over micro-to-millisecond timescales, a SPAD pixel generates an electrical pulse for each photon detection event. A time-to-digital conversion circuit converts each pulse into a timestamp recording the time-of-arrival of each photon. Under normal illumination conditions, a SPAD pixel can generate millions of photon timestamps per second. The photon timestamps are often captured with respect to a periodic synchronization signal generated by a pulsed laser source. To make this large volume of timestamp data more manageable, SPAD-based SPCs build a *timing histogram* in-sensor instead of transferring the raw photon timestamps to the processing chip. The histogram records the number of photons as a function of the time delay with respect to the synchronization pulse.

Consider a megapixel SPAD-based 3D camera. For short range indoor applications (up to tens of meters), a millimeter depth resolution would be desirable. For longer range outdoor applications (hundreds of meters), centimeter level depth resolution would be desirable. Assuming state-of-the-art sub-bin processing techniques [22], this corresponds to histograms with thousands of bins per pixel. Moreover, the rate at which these histograms are acquired can vary from tens of frames per second (fps) for low speed applications to hundreds of fps for, say, an automotive application where objects may be moving at high speeds. Even a conservative estimate of a 30 fps megapixel camera leads to a large data-rate of 10^6 pixels/frame $\times$ 1000 bins/pixel $\times$ 2 bytes/bin $\times$ 30 fps $=$ 60 GB/sec. As shown in Fig. 1(b), the amount of data generated by this conventional full histogram capture method varies linearly with the desired depth resolution and exceeds the bandwidth of state-of-the-art data-transfer busses (like USB and PCIe) by orders of magnitude.

In this paper we propose a bandwidth-efficient acquisition strategy called *compressive single-photon histograms* (CSPH). Instead of capturing the full timing histogram in each pixel, a CSPH is constructed by mapping the time bins of the full histogram onto multiple "compressive bins" through an encoding step. We consider a family of compressive encoders that are linear, which means they can be represented as a simple matrix operation. Therefore, they can be implemented efficiently using vector addition operations that can be computed on-the-fly, as each photon arrives, without the need to store large arrays of photon timestamps in-sensor. CSPHs decouple the dependence of output data rate on the desired depth resolution. While a full histogram would require more time bins to achieve higher depth resolution, a CSPH can represent them using (almost) the same number of compressive bins. As illustrated in Fig. 1(a), CSPHs can reduce the required data rate by 1-2 orders of magnitude compared to the full histogram case.

We design and evaluate various CSPH coding schemes for SPAD-based 3D cameras. We propose a general decoding algorithm that directly estimates per-pixel scene depths from a CSPH, based on recent work in structured light [39]. Our method also accounts for different laser pulse shapes and pixel response characteristics, i.e., the system's impulse response function (IRF). We perform extensive simulations over a wide range of background and signal powers, laser waveform shapes and illumination schemes. We also evaluate depth reconstruction accuracy of our compressive acquisition method with real-world data captured using a single-pixel raster-scanning hardware prototype.

2. Related Work

Coarse in-pixel histogramming is one common strategy to reduce data rates in SPAD-based 3D cameras [8, 22, 27, 28, 50]. Despite the low time resolution in coarse histograms, it is possible to achieve high depth resolution by using wide pulses [22], pulse dithering [47], or with coarse-to-fine histogram architectures [59]. In this paper, we show that coarse histogramming is sub-optimal compared to other compressive histogramming strategies. Additional data reduction strategies, such as motion-driven operation [8] or multi-photon triggering [28], have also been proposed. Moreover, in the context of scanning-based systems, adaptive sampling methods have been proposed to reduce sampling rates and consequently data transfers [4, 24, 46]. These more complex methods can be used in a complementary manner with CSPHs to further reduce data rates.

Recently, *Fourier-domain histograms* (FDHs), were proposed for fast NLOS reconstruction [35, 41] and for single-photon 3D imaging [52, 53]. FDHs are one type of CSPHs that can achieve significant compression over regular histogramming [52]. In this paper, we present CSPH strategies that are not only more efficient than FDH for 3D imaging, but are also robust to diffuse indirect reflections commonly found in flash illumination systems.

Coding matrix design for 3D imaging has been studied in the context of correlation-based ToF (C-ToF) [18, 19, 21, 31, 32, 56] and structured light (SL) [6, 16, 17, 39]. In particular, there are interesting similarities between CSPH and C-ToF coding matrices, since both 3D cameras are based on the time-of-flight principle. Nonetheless, C-ToF coding is fundamentally different. In C-ToF, each coded *measurement* is captured sequentially with a coded light signal and sensor, making the noise statistics across measurements independent. In CSPHs, all coded *projections* are performed simultaneously on the same signal, making the noise statistics across projections dependent.

3. Single-Photon 3D Image Formation

Single-photon 3D cameras consist of a SPAD sensor and a periodic pulsed laser that illuminates the scene. Assuming direct-only reflections, the returning photon flux signal that will be captured by a SPAD pixel can be written as:

$$\Phi(t) = ah(t - t_z) + \Phi^{\text{bkg}} = \Phi^{\text{sig}}(t) + \Phi^{\text{bkg}} \tag{1}$$

where $h(t)$ is the system's IRF which accounts for the pulse waveform and sensor IRF, a represents the returning signal photon flux, t_z is a time shift proportional to distance, and Φ^{bkg} corresponds to the background photon flux. Although simple, Eq. 1's model is a valid approximation in a wide variety of active illumination scenarios, in particular, for scanning systems [1].

SPAD-based 3D cameras sample $\Phi(t)$ using time-correlated single-photon counting (TCSPC) [43, 59]. The SPAD pixel, once triggered, starts acquiring photons. After detecting one photon, the photon timestamp is recorded, and the SPAD is inactive for a time period called the dead time (∼50ns). As shown in Fig. 2, the above process is repeated for M cycles, and a histogram of the timestamps is constructed which approximates $\Phi(t)$. If the photons are time-tagged with a resolution, Δ, we can write the mean

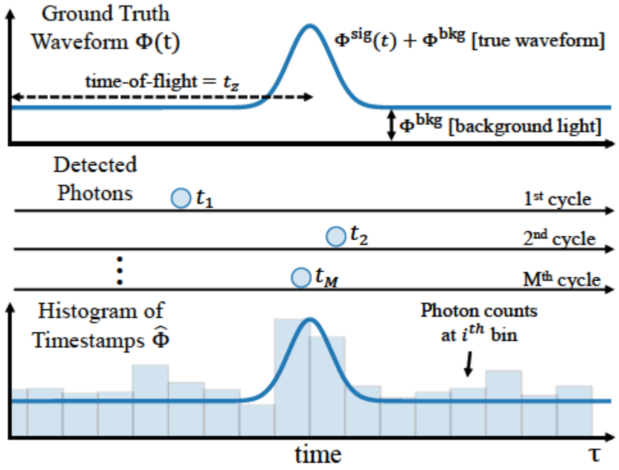


Figure 2. **Single-Photon Histogram Formation.** SPAD-based 3D cameras estimate distances by building a per-pixel histogram of the detected photons time-of-arrival. The histogram is a discrete approximation of the photon flux waveform incident on the pixel, which encodes distances in the time shift (t_z) of the pulse.

photon flux at histogram bin i as:

$$\Phi_i = \Phi_i^{\mathrm{sig}} + \Delta\Phi^{\mathrm{bkg}} \tag{2}$$

The vector, $\Phi = (\Phi_i)_{i=0}^{N-1}$, is the photon flux waveform histogram, where $N = \tau/\Delta$, and τ is the timestamp range which often equals the laser pulse repetition period. Here we assume that the SPAD sensor is being operated in asynchronous mode [14] or is capable of multi-event timestamp collection [22], which minimizes pile-up distortions [15, 25, 42], and guarantees that Φ_i is an appropriate approximation of $\Phi(t)$.

The histogram formation process generates a 3D histogram image, one histogram per pixel. In emerging megapixel SPAD arrays with picosecond time resolutions, building the histogram image off-sensor requires transferring thousands of timestamps per-pixel, leading to TB/s data rates. Moreover, building in-pixel histograms, would still require transferring the 3D data volume off-sensor for processing, which still results in impractical data rates of tens of GB/s. Overall, data bandwidth is an important practical challenge for emerging single-photon 3D cameras.

4. Compressive Single-Photon Histograms

In general, we *could* compress the 3D histogram image effectively *if* we had the entire histogram image. However, building and transferring the histogram image off the sensor is expensive. This raises the question, can we compress a histogram without ever explicitly constructing it? Recall that these histograms are created one photon at a time, raising the follow-up question: Can we compress the histogram in an online fashion where we see a photon (and its timing information) only once? This is challenging because compression schemes often require having access to the entire data before performing compression.

To answer the above question, we make two key observations. First, there is a class of linear compression techniques which can be expressed as a simple matrix-vector multiplication. Specifically, the compressed representation is the product of a $K \times N$ *coding matrix*, C, and the $N \times 1$ histogram Φ. The effectiveness of C can be measured by the compression ratio (N/K) that is achieved, while preserving down-stream task (e.g., depth estimation) performance.

Second, we observe that the entire histogram can be written as the sum of several one-hot encoding vectors, each vector representing one timestamp. Formally, let $t_j = (t_{j,i})_{i=0}^{N-1}$ be the one-hot encoding vector of the jth photon timestamp (T_j) detected, where all elements are 0 except for $t_{j,l} = 1$, in which $l = \lfloor \frac{T_j \mod \tau}{\Delta} \rfloor$. As illustrated in Fig. 3, the measured histogram, $\widehat{\Phi}$, can be written as:

$$\widehat{\Phi}_i = \sum_{j=0}^{M-1} t_{j,i} \tag{3}$$

where M is the total number of detected photons.

Given these observations, we can design an online histogram compression algorithm by simply multiplying the coding matrix with the one-hot encoding timestamp vector:

$$\widehat{B}_k = \sum_{i=0}^{N-1} C_{k,i}\widehat{\Phi}_i = \sum_{i=0}^{N-1}\sum_{j=0}^{M-1} C_{k,i}t_{j,i} \tag{4}$$

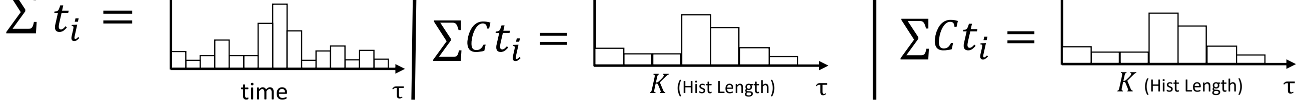
$\widehat{B}$ is the compressive single-photon histogram (CSPH), whose elements are coded projections of $\widehat{\Phi}$.

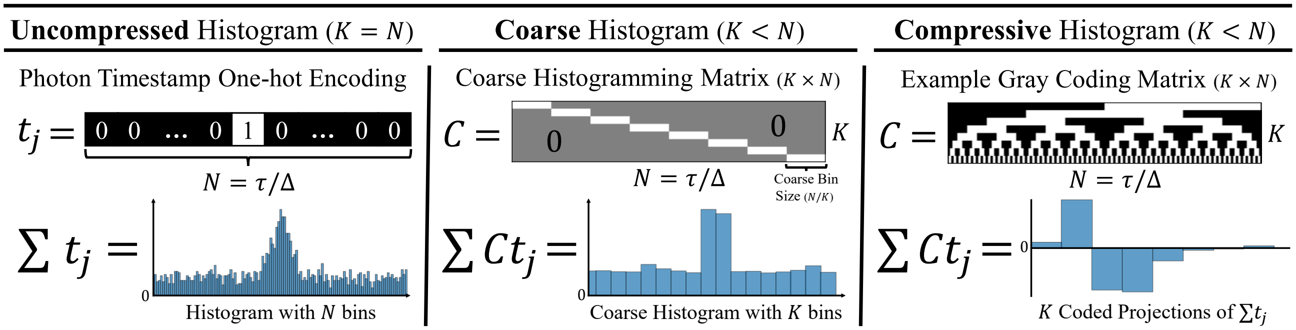
Although, Eq. 4 is expressed as a matrix-vector multiplication, a practical CSPH implementation would perform low-compute operations, per-photon, like conventional histograms. One possible implementation is to store C as a *lookup table* shared across pixels. For each new t_j with $t_{j,l} = 1$, the l^{th} column of C is added to the per-pixel CSPH ($\widehat{B} = \widehat{B} + C_{:,l}$). Note that we never need to store the timing information explicitly, nor need to create the histogram. The only data stored and output by the compressive SPAD pixel is $\widehat{B}$. Given this on-the-fly compression method, a natural question is, what are good coding matrices for compressive single-photon 3D cameras?

4.1. How to design a coding matrix for 3D Imaging?

In theory, C can be chosen to be any set of K linear projections. In practice, however, we can define certain properties C should have to achieve high compression rates while preserving 3D imaging performance.

The ith column of C can be interpreted as a code word of length K that represents the ith time bin. We can view this code word vector as a point in a K-dimensional space. Moreover, consider the curve that is traced by the N points





Uncompressed Histogram ($K = N$)
Coarse Histogram ($K < N$)
Compressive Histogram ($K < N$)
Photon Timestamp One-hot Encoding
$t_j =$ 0 0 ... 0 1 0 ... 0 0
$N = \tau/\Delta$
$\sum t_j =$
0
Histogram with N bins
Coarse Histogramming Matrix ($K \times N$)
$C =$ 0 0 K
$N = \tau/\Delta$
Coarse Bin Size (N/K)
$\sum Ct_j =$
0
Coarse Histogram with K bins
Example Gray Coding Matrix ($K \times N$)
$C =$ K
$N = \tau/\Delta$
$\sum Ct_j =$
0
K Coded Projections of $\sum t_j$

Figure 3. **On-the-fly Histogram Formation.** Single-photon timestamp histograms are often generated on-the-fly, as each photon timestamp comes in. The left column shows how a histogram, whose bin width matches the timestamp resolution ( ), is formed as the sum of timestamps represented as one-hot encoded vectors. Transferring such a large histogram for every pixel can be impractical. By multiplying ser bins, the size of the histogram can be reduced at the cost compressive histogram can be created by multiplying each mp comes in. A well-designed coding matrix can efficiently

Coding Matrices V2

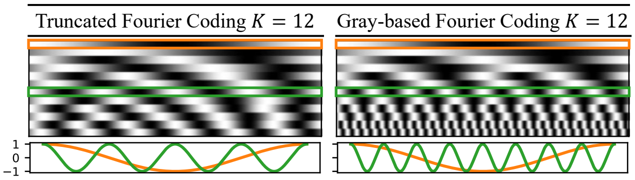


Figure 4. **Example Coding Matrices.** Truncated Fourier and Gray-based Fourier matrices with 12 codes (rows). The odd and even rows of the truncated Fourier matrix are given by and , respectively, where , , and is the column. The odd and even rows of the Gray-based Fourier matrix are given by and , respectively, and for the rows are generated using codes from the truncated Fourier matrix that have not been used.

(columns of), denoted as the coding curve $\mathcal{C}$. In this section, we make use of the coding curve concept to define desirable properties of . We would like to note that the coding curve construct has been previously used for other active imaging systems including: C-ToF [19, 21], SL [17], and FLIM [33]. A “good” compressive single-photon 3D imaging coding matrix will have the following properties:

Uniqueness Property: Each point in $\mathcal{C}$ is unique, i.e., $\mathcal{C}$ should be non self-intersecting. This guarantees that two different bins are not represented by the same code word.

Robustness Property: $\mathcal{C}$ should be locality preserving. Meaning that if we add a small perturbation to a point on the curve, it should map to neighboring points along the curve, which correspond to code words with similar time bins.

Indirect Reflections Property: The coding functions (rows of) encode information of the photon flux waveform in a similar way as correlation functions do in C-ToF imaging [18, 19]. In the presence of diffuse indirect light reflections, the direct-only model from Eq. 1 becomes invalid. Diffuse indirect reflections arise when imaging concave geometries or in the presence of volumetric scattering, and is particularly problematic in flash illumination systems. Gupta et al. [18] showed that diffuse indirect reflections appear in as smooth band-limited signals. Therefore, the direct-only model will still apply for coding functions with frequencies above a scene dependent threshold.

Band-limit Property: The coding functions should not all be high-frequency functions when the goal is to encode a photon flux waveform with a smooth system IRF ( ). This is because if a given coding function is composed of frequencies that are above the bandwidth of , then the expected encoded value will be 0 (see supplement). The orange and green lines of Fig. 5 show example of smooth IRF's. This means that, although, high frequencies can mitigate indirect reflections, very high frequency codes may not be useful if they are outside of bandwidth.

4.2. CSPH Coding Schemes

In this paper, we analyze and evaluate the following CSPH coding matrices:

1. **Coarse Histogram**: is a downsampling matrix, where each row is a rectangular window with length , as shown in Fig. 3. This matrix does not fulfill the uniqueness property because each window maps multiple bins to the same code word vector. This type of CSPH is equivalent to the commonly used coarse in-pixel histograms [8, 27, 28, 50].
2. **Truncated Fourier** [52]: is made up of the first rows of the discrete Fourier transform matrix, skipping the zeroth harmonic. This matrix fulfills the uniqueness property, and at higher it may contain high frequency codes that can mitigate indirect reflections.
3. **Continuous Gray (Proposed)** : Also known as Hamiltonian codes [19], the Cont. Gray coding curve is a Hamiltonian cycle on a -dimensional hypercube, which is provably locality preserving (robustness prop-

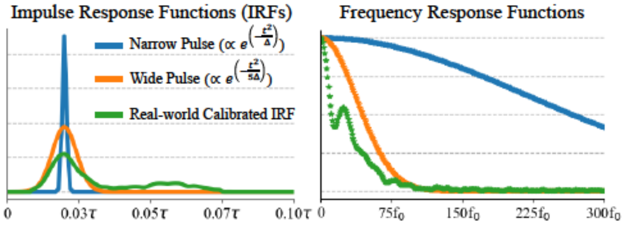


Figure 5. **System Impulse Response Functions ($h(t)$).** Gaussian pulses are commonly used approximations used for the system IRF because the single-photon 3D cameras use pulsed illumination signals. Although this approximation may be valid for some systems, the experimental system used for the results in Sec. 6.1 exhibited an IRF with a long-tail (green line). Analyzing the frequency-domain representation of the IRF is a useful step when designing the coding matrix of a compressive single-photon 3D camera.

erty) [11]. The rows of C are generated by constructing K-bit Gray code [13], where each code will have length 2^K, and linearly interpolating them to have length N. For a histogram of length $N = 2^K$ this coding matrix becomes fully binary. Fig. 3 shows a Gray C with $K = 8$.

4. **Gray-based Fourier (Proposed)**: For N histogram bins, Gray coding is only valid for $K \leq \log_2(N)$ because the higher-order coding functions start aliasing. However, in lower SNR scenarios it is sometimes desired to increase K to preserve depth precision. To this end, we design a new coding scheme that combines properties of Gray and Fourier coding. For the first $k \leq 2\log_2(N)$ rows, Fourier components are sampled using the frequency doubling pattern observed in the Gray coding matrix in Fig. 3. For the remainder $2\log_2(N) < k < K$ codes, we revert back to a truncated Fourier sampling using the remaining frequencies.

In the supplement we present further analysis on the properties of these coding matrices, and evaluate additional coding schemes including: Hadamard and short-time Fourier.

5. When is isometric compression achieved?

A CSPH achieves isometric compression when its performance is within a specified margin of the uncompressed full-resolution histogram (FRH). Specifically,

$$\text{IsometricCompression}(\epsilon) := \varepsilon_{\text{diff}} \leq \epsilon \tag{5}$$

where $\varepsilon_{\text{diff}} = |\varepsilon_{\text{FRH}} - \varepsilon_{\text{CSPH}}|$, ε_{FRH} and $\varepsilon_{\text{CSPH}}$ are the performance metrics for a CSPH and a FRH, and ϵ is the desired performance difference margin.

To quantify 3D imaging performance, we compute the relative mean depth errors (MDE) over the depth range through Monte Carlo simulations using Eq. 1's direct-only model, as in [21] (see supplement). Consequently, to identify isometric compression we take the difference between the relative MDE of a FRH (ε_{FRH}) and the CSPH ($\varepsilon_{\text{CSPH}}$) and classify the difference into different margins. For context, in a 3D imaging scenario with a 10m depth range, a relative MDE difference of $\varepsilon_{\text{diff}} = 0.1\%$ corresponds to the CSPHs MDEs being within 1cm of the FRH MDEs.

Depth Estimation: To decode depths from a CSPH, we compute the zero-mean normalized cross-correlation (ZNCC) [39] between $\widehat{B}$ and C:

$$\widehat{t}_z \propto \arg\max_i \frac{C^h_{:,i} - \text{mean}(C^h_{:,i})}{\left\|C^h_{:,i} - \text{mean}(C^h_{:,i})\right\|} \cdot \frac{\widehat{B} - \text{mean}(\widehat{B})}{\left\|\widehat{B} - \text{mean}(\widehat{B})\right\|} \tag{6}$$

where $C^h_{:,i}$ is the ith column of C^h, and C^h is the coding matrix, C, with each row convolved with the system IRF h. To compute depths for FRHs we use matched filtering [57]. See supplement for more details on ZNCC depth estimation.

5.1. Isometric Compression Analysis

A high-performance single-photon 3D imaging system will match the laser pulse width with the SPAD sensor time resolution (Δ) [22,23]. In this section, we analyze isometric compression at a wide range of signal-to-background ratio (SBR) and photon count levels[1], for the case of a FRH with $N = 1024$ bins that records a Gaussian pulse of width Δ, i.e., $h(t) \propto \exp^{-(t)^2/\Delta}$ (blue line in Fig. 5). In addition to the CSPH coding schemes from Sec. 4.2, we also evaluate the following two baselines:

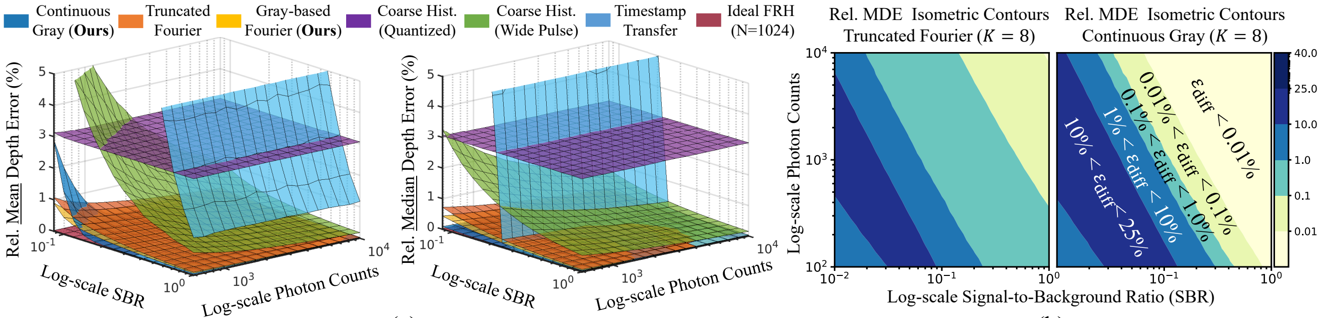
- **Coarse Hist. (Wide Pulse)** [22]: A coarse histogram C paired with a wide Gaussian pulse width that matches the window length, which enables sub-bin precision.
- **Truncated Timestamps**: A FRH constructed with *at most* K timestamps, even if the number of detected photons is $> K$. Although, not a CSPH, this is an important baseline corresponding to the simplest SPAD pixel that transfers the same amount of data as a size K CSPH.

Extreme Compression: Fig. 6a shows the relative mean and median depth errors for different CSPH with $K = 8$, resulting in a compression ratio of 128x. The ideal FRH, obtains near 0 error in the visualized SBR and photon count levels. In this extreme compression regime, Gray coding is the only CSPH that achieves an isometric compression where, $\varepsilon_{\text{diff}} \leq 0.01\%$, at various SBR and photon count levels, essentially matching FRH performance (Fig. 6b). The difference in the mean and median error trends indicates that at low SBR and low photon counts, Gray coding produces either high or near zero errors. On the other hand, the error magnitudes of truncated Fourier coding are similar across SBR and photon count levels. Unfortunately, even at high SBR levels, truncated Fourier still does not each isometric compression with $\varepsilon_{\text{diff}} \leq 0.01\%$. Gray-based Fourier coding consistently outperform truncated Fourier, and is more robust to outliers than Gray coding at low SBR. As expected, the coarse histogramming methods are either, quantization-limited due to low time resolution, or noise-limited when using a wide pulse. Finally, only transferring

[1] $\text{SBR} = \frac{\sum_{i=0}^{N-1} \Phi_i^{\text{sig}}}{N\Delta\Phi^{\text{bkg}}}$, Photon Counts $= \sum_{i=0}^{N-1} \Phi_i$

K = 8 Analysis (High Compression) V3

Isometric Compression Analysis for C-SPHs of Length $K = 8$ ($CR = 128x$)

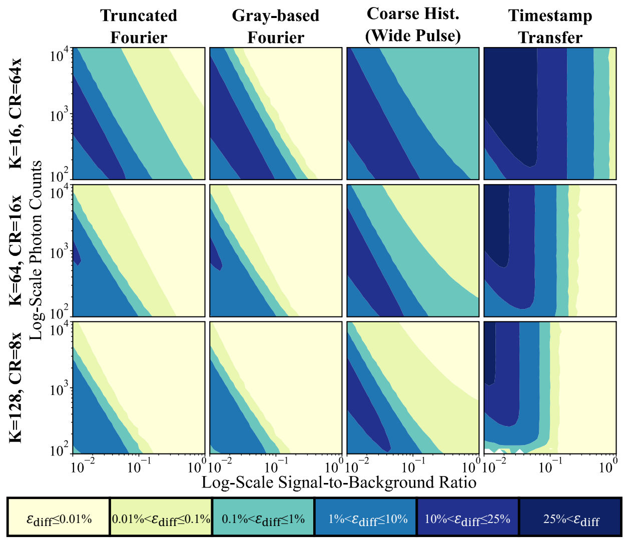


(a) (b)

Figure 6. **Isometric Compression Analysis at 128x Compression.** (a) Shows the relative mean and median depth errors computed as

Isometric Compression Contours v3

Relative Mean Depth Error Isometric Contours



$\varepsilon_{\text{diff}} \leq 0.01 \leq \%$

Figure 7. **Compression vs. Performance.** As we increase the isometric compression regions with $_{\text{diff}}$ increase for all methods. For SBR and photon count levels greater than and , Gray-based Fourier coding with performs as well as a FRH with . Truncated Fourier Coding, on the other hand, requires higher to reach a $_{\text{diff}}$ at those SBR and photon count levels. Overall, carefully designed coding matrices such as Fourier and Gray-based, consistently outperform standard approaches (coarse histograms and timestamp transfer).

timestamps, although practical, leads to poor performance at low SBR and to many outliers at high SBR.

Compression vs. Performance: Fig. 7 shows how the isometric compression contours change as we increase for different coding schemes. As increases and compression decreases, the performance of all methods improves and approaches FRH performance, in particular for carefully designed CSPH coding schemes like Gray-based and truncated Fourier. At the benefits of Gray-based Fourier coding are more evident as the isometric contours where its performance matches FRH covers the biggest range of SBR and photon count levels (i.e., $_{\text{diff}}$). As continues to increase, the truncated and Gray-based Fourier coding matrices become similar, making their performance nearly identical, as observed in the case. Overall, Gray-based and Fourier-based CSPH coding, consistently outperform current photon timestamp storage and transfer approaches, i.e., coarse histograms and direct transfer of timestamps. Interestingly, transferring only timestamps significantly outperforms a coarse histogram at many SBR and photon count levels.

Isometric Compression with Wide Pulses: Here we analyzed CSPHs for a Gaussian pulse of width . As the pulse width increases, the effective time resolution of the system decreases, impacting the performance of the FRH baseline, which makes the isometric compression regions with low relative differences become larger at lower . Moreover, as seen in Fig. 5, slightly widening the pulse decreases the frequency content of the signal significantly, making compressive coding strategies more efficient because they only need to sample the non-zero frequencies. Please refer to the supplement for results and analysis using wider pulses.

6. Compressive Single-Photon 3D Imaging

In this section, we evaluate CSPH coding approaches on real-world data from a scanning-based system [14], and simulated data from a flash-illuminated system.

6.1. Real-world Scanning-based System Results

To evaluate the effectiveness of CSPHs on real SPAD timestamp data we downloaded and pre-processed the data acquired with a scanning-based system [14]. The preprocessed raw histograms have ps and

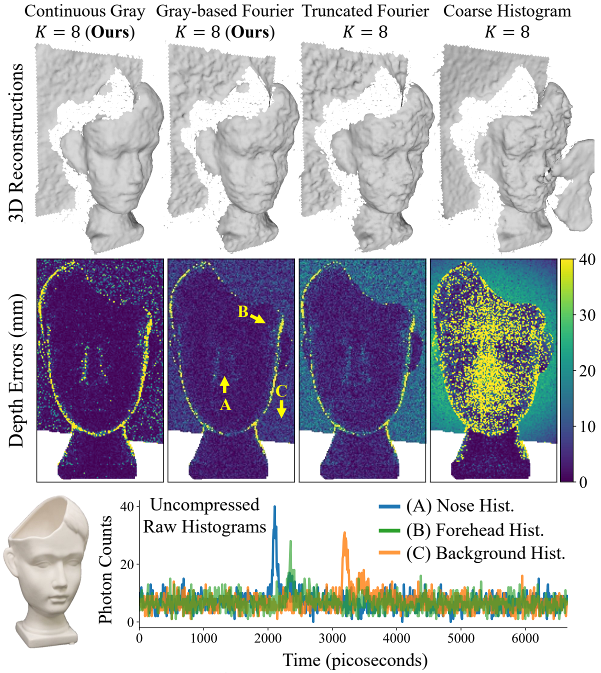


Figure 8. **Real-world Scan-based Single-photon 3D Imaging.** The depth and depth error images for different CSPH with codes. The mean and median absolute errors (in mm) achieved by each method from left to right are: [,], [,], [,], [,].

(e.g., histograms in Fig. 8). For depth estimation we extract the center pixel histogram, denoise it, and use that as the system IRF (green line in Fig. 5). Ground truth depths are obtained from the FRH with light Gaussian denoising, and we mask the pixels where even FRH had too low of an SBR to estimate reliable depths (white regions in Fig. 8 images).

Fig. 8 shows the recovered 3D reconstructions using different CSPH at a x ( ). Similar to our simulations, we find that Gray coding can essentially achieve errors for pixels with sufficient signal, while sometimes making large errors (outliers). In contrast, truncated Fourier and Gray-based Fourier are robust to outliers, but make many small and medium sized errors leading to lower quality 3D reconstructions in this example. Moreover, we found that the background wall histograms exhibited a longer tail than the foreground face histograms due to indirect reflections (see supplement). Indirect reflections cause systematic errors in truncated Fourier, while Gray-based Fourier and Gray coding are more robust to these errors since their have higher frequency coding functions, as discussed in Sec. 4.1. Finally, the wide system IRF allows a coarse histogram to achieve sub-bin precision using ZNCC decoding. Nonetheless, its performance is significantly worse than the other CSPH methods. Additional results at different and another scan can be found in the supplement.

6.2. Simulated Flash Illumination System Results

SPAD arrays are often used in flash illumination systems to achieve a fully solid-state single-photon LiDAR [22]. To evaluate CSPHs in a flash illumination system, we used physically accurate histogram images rendered with MitsubaToF [42] (ps,) obtained from [20]. To simulate the FRHs we set the mean photon count and mean SBR levels for the scene, and scale the histogram and background image (R channel of RGB) accordingly.

	In-Pixel Memory	In-Pixel Compute (per-photon)	In-Pixel Compute (per-depth)	Per-pixel Output Data Rate	Depth Precision
Full-Resolution Hist. (In-sensor Depth Est.)	High $\mathcal{O}$	Medium $\mathcal{O}$	High $\mathcal{O}$	Low $\mathcal{O}$	High
Full-Resolution Hist. (Off-sensor Depth Est.)	High $\mathcal{O}$	Medium $\mathcal{O}$	Low (None)	High $\mathcal{O}$	High
Timestamp Transfer (Off-sensor Depth Est.)	Low $\mathcal{O}$	Low (None)	Low (None)	High $\mathcal{O}$	High
Coarse Hist. (Off-sensor Depth Est.)	Medium $\mathcal{O}$	Medium $\mathcal{O}$	Low (None)	Medium $\mathcal{O}$	Low
CSPH (Proposed) (Off-sensor Depth Est.)	Medium $\mathcal{O}$	High $\mathcal{O}$	Low (None)	Medium $\mathcal{O}$	Medium

Table 1. **SPAD-based 3D Camera Design Trade-offs**. Qualitative comparison of the memory, compute, and data rate requirements for different SPAD-based 3D camera designs.

Fig. 9 shows the resulting depth images and depth errors for two different scenes. In addition to the edges where true depth is ambiguous, both scenes have regions with very low SBR where even the FRH has some depth errors (stove in kitchen, and the mat in bathroom). Due to indirect reflections, truncated Fourier makes significant systematic errors even when using codes. On the other hand, Gray-based Fourier recovers highly accurate depths like an FRH, while using - x less data. Please refer to the supplement for additional comparisons.

7. Discussion and Limitations

SPAD-based 3D cameras with high spatio-temporal resolution can produce unmanageable data rates. To reduce their data bandwidth, we proposed to capture a compressive representation (CSPH) of the high-resolution timing histogram, from which depths can be computed. The CSPH is built in an online manner by projecting each photon timestamp with a coding matrix and aggregating them. By designing the coding matrix appropriately, a CSPH can match the depth precision of a full-resolution histogram in a wide range of scenarios, while outputting significantly less data.

In-sensor Memory and Compute Considerations: Although, CSPHs can reduce the in-pixel memory and sensor data transmission rates, this comes at the expense of higher per-photon computations, as summarized in Tab. 1. While conventional histograms require a single addition per-photon (increment histogram bin), CSPHs perform additions per-photon. Nonetheless, it may be possible to implement these additions efficiently using SIMD processing. The next step in this line of work is to further analyze these memory, computation, and data transmission trade-offs from a hardware perspective.

Why not compute depths in-sensor? One way to reduce the data rates would be to compute the per-pixel depths in-pixel. However, due to the non-linearity of depth estima-

Kitchen-2 | SBR=0.25 | N Photons = 1e3

Bathroom-cycle-2 | SBR=0.5 | N Photons =

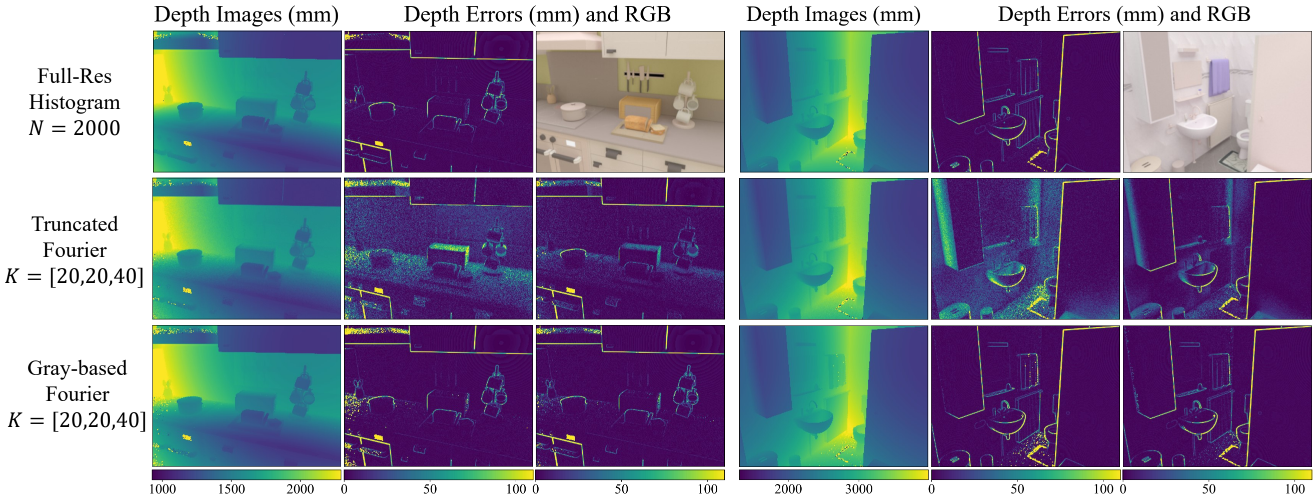


Figure 9. **Flash Illumination Compressive Single-Photon 3D Imaging.** Depth images and depth errors for CSPH methods applied to histograms of flash illuminated scenes. The kitchen and bathroom scenes were simulated with a mean photon count and mean SBR of (1000, 0.25) and (1000, 0.5), respectively. The top row has the recovered depths and depth errors of an FRH with bins, and the RGB image of the scene. The second and third rows have the CSPH depths ( ), and the depth errors for (middle column) and (right column). The mean and median absolute errors (in mm) for the kitchen scene were: FRH-2000: (14, 3), truncated Fourier-20: (26, 10), Gray Fourier-20: (22, 3), truncated Fourier-40: (16, 4), Gray Fourier-40: (14, 3). Similarly, for the bathroom scene: FRH-2000: (10, 3), truncated Fourier-20: (24, 11), Gray Fourier-20: (12, 3), truncated Fourier-40: (14, 4), Gray Fourier-40: (11, 3).

tion (e.g., via peak finding), it requires building and storing the full-resolution histograms in the SPAD sensor chip. Although, this approach could provide near-optimal compression, it requires significant in-pixel memory. Moreover, advanced depth estimation algorithms, such as matched filtering, can have computational complexities larger than the length of the histogram ( ).

Hardware Implementation: CSPHs are designed to reduce the per-pixel output data rate, which requires in-pixel implementation. Recent advances in 3D-stacking CMOS technology has enabled the in-pixel implementation of one type of CSPH, namely coarse histograms [8, 27, 28, 50]. The next step in this line of work is to explore SPAD pixel architectures that implement different CSPHs, which may impose interesting constraints on the structure of . For instance, the architecture for a binary (e.g., coarse histograms or Gray coding with $_2$) may be simpler than for a with continuous values.

Code Optimization: Instead of designing a coding matrix based on the heuristics discussed in Sec. 4.1, could be optimized. We attempted to optimize by setting it to the PCA basis learned over a dictionary of shifted Gaussian pulses. Unfortunately, this method simply leads to a Fourier matrix (see supplement). One direction for future work could explore optimizing through gradient descent [39].

Photon-starved regime: When less than 20 photons are recorded, timestamp transfer can outperform a CSPH at . At lower SBR levels all methods begin to fail and perform comparably. We present this analysis in the supplement. Ultimately, low SBR and photon-starved scenarios require denoising to recover reliable depths [34, 44]. Denoising a CSPH instead of the full 3D histogram image may provide some interesting computational benefits.

Non-linear Compression: It may be possible to design a non-linear on-the-fly compression method. One challenge in the design of such method is that it may not be able to leverage the priors that enabled the design of coding matrices robust to noise and indirect reflections. Nonetheless, exploring non-linear online compression algorithms remains an interesting direction for future work.

Efficient Depth Estimation: ZNCC depth estimation allows comparing different coding matrices under a single framework. One limitation of our ZNCC implementation is its linear computational and memory complexity. Given that ZNCC is a template matching algorithm, efficient coarse-to-fine implementations may be possible [2]. Alternatively, algorithms tailored for a particular may provide further computational benefits. For instance, Fourier coding has different analytical [10,45], optimization-based [26,52], and data-driven [54] depth decoding algorithms.

Social Implications: The work in this paper may contribute towards the deployment of SPAD-based 3D cameras. The emergence of SPAD-based NLOS imaging, raises privacy concerns as these systems become available. Additionally, 3D applications such as autonomous vehicles, may have unintended socio-economic and environmental implications.

Acknowledgments: This work was supported by the Department of Energy and National Nuclear Security Administration (DE-NA0003921), National Science Foundation (1846884, 1943149, 2107060, 2003129), and a Draper TIF grant from UW-Madison. U.S. DOE full legal disclaimer: `https://www.osti.gov/stip/about/disclaimer`.

References

[1] Supreeth Achar, Joseph R Bartels, William L Whittaker, Kiriakos N Kutulakos, and Srinivasa G Narasimhan. Epipolar time-of-flight imaging. *ACM Transactions on Graphics (TOG)*, 36(4):37, 2017. 2

[2] Edward H Adelson, Charles H Anderson, James R Bergen, Peter J Burt, and Joan M Ogden. Pyramid methods in image processing. *RCA engineer*, 29(6):33–41, 1984. 8

[3] Yoann Altmann, Stephen McLaughlin, Miles J Padgett, Vivek K Goyal, Alfred O Hero, and Daniele Faccio. Quantum-inspired computational imaging. *Science*, 361(6403), 2018. 1

[4] Alexander W Bergman, David B Lindell, and Gordon Wetzstein. Deep adaptive lidar: End-to-end optimization of sampling and depth completion at low sampling rates. In *2020 IEEE International Conference on Computational Photography (ICCP)*, pages 1–11. IEEE, 2020. 2

[5] Mauro Buttafava, Jessica Zeman, Alberto Tosi, Kevin Eliceiri, and Andreas Velten. Non-line-of-sight imaging using a time-gated single photon avalanche diode. *Optics express*, 23(16):20997–21011, 2015. 1

[6] Wenzheng Chen, Parsa Mirdehghan, Sanja Fidler, and Kiriakos N Kutulakos. Auto-tuning structured light by optical stochastic gradient descent. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5970–5980, 2020. 2

[7] Sergio Cova, Massimo Ghioni, Andrea Lacaita, Carlo Samori, and Franco Zappa. Avalanche photodiodes and quenching circuits for single-photon detection. *Applied optics*, 35(12):1956–1976, 1996. 1

[8] Francesco Mattioli Della Rocca, Hanning Mai, Sam W Hutchings, Tarek Al Abbas, Kasper Buckbee, Andreas Tsiamis, Peter Lomax, Istvan Gyongy, Neale AW Dutton, and Robert K Henderson. A 128 128 spad motion-triggered time-of-flight image sensor with in-pixel histogram and column-parallel vision processor. *IEEE Journal of Solid-State Circuits*, 55(7):1762–1775, 2020. 2, 4, 8

[9] Eric R Fossum, Jiaju Ma, Saleh Masoodian, Leo Anzagira, and Rachel Zizza. The quanta image sensor: Every photon counts. *Sensors*, 16(8):1260, 2016. 1

[10] John P Godbaz, Michael J Cree, and Adrian A Dorrington. Closed-form inverses for the mixed pixel/multipath interference problem in amcw lidar. In *Computational Imaging X*, volume 8296, page 829618. International Society for Optics and Photonics, 2012. 8

[11] Craig Gotsman and Michael Lindenbaum. On the metric properties of discrete space-filling curves. *IEEE Transactions on Image Processing*, 5(5):794–797, 1996. 5

[12] Francesco Gramuglia, Ming-Lo Wu, Claudio Bruschini, Myung-Jae Lee, and Edoardo Charbon. A low-noise cmos spad pixel with 12.1 ps sptr and 3 ns dead time. *IEEE Journal of Selected Topics in Quantum Electronics*, 2021. 1

[13] Frank Gray. Pulse code communication. *United States Patent Number 2632058*, 1953. 5

[14] Anant Gupta, Atul Ingle, and Mohit Gupta. Asynchronous single-photon 3d imaging. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 7909–7918, 2019. 1, 3, 6

[15] Anant Gupta, Atul Ingle, Andreas Velten, and Mohit Gupta. Photon-flooded single-photon 3d cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6770–6779, 2019. 3

[16] Mohit Gupta, Amit Agrawal, Ashok Veeraraghavan, and Srinivasa G Narasimhan. Structured light 3d scanning in the presence of global illumination. In *CVPR 2011*, pages 713–720. IEEE, 2011. 2

[17] Mohit Gupta and Nikhil Nakhate. A geometric perspective on structured light coding. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 87–102, 2018. 2, 4

[18] Mohit Gupta, Shree K Nayar, Matthias B Hullin, and Jaime Martin. Phasor imaging: A generalization of correlation-based time-of-flight imaging. *ACM Transactions on Graphics (ToG)*, 34(5):156, 2015. 2, 4

[19] Mohit Gupta, Andreas Velten, Shree K Nayar, and Eric Breitbach. What are optimal coding functions for time-of-flight imaging? *ACM Transactions on Graphics (TOG)*, 37(2):13, 2018. 2, 4

[20] Felipe Gutierrez-Barragan, Huaijin Chen, Mohit Gupta, Andreas Velten, and Jinwei Gu. itof2dtof: A robust and flexible representation for data-driven time-of-flight imaging. *arXiv preprint arXiv:2103.07087*, 2021. 7

[21] Felipe Gutierrez-Barragan, Syed Azer Reza, Andreas Velten, and Mohit Gupta. Practical coding function design for time-of-flight imaging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1566–1574, 2019. 2, 4, 5

[22] Istvan Gyongy, Sam W Hutchings, Abderrahim Halimi, Max Tyler, Susan Chan, Feng Zhu, Stephen McLaughlin, Robert K Henderson, and Jonathan Leach. High-speed 3d sensing via hybrid-mode imaging and guided upsampling. *Optica*, 7(10):1253–1260, 2020. 2, 3, 5, 7

[23] Nathan Hagen, Matthew Kupinski, and Eustace L Dereniak. Gaussian profile estimation in one dimension. *Applied optics*, 46(22):5374–5383, 2007. 5

[24] Abderrahim Halimi, Philippe Ciuciu, Aongus Mccarthy, Stephen Mclaughlin, and Gerald S Buller. Fast adaptive scene sampling for single-photon 3d lidar images. In *2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 196–200. IEEE, 2019. 2

[25] Felix Heide, Steven Diamond, David B Lindell, and Gordon Wetzstein. Sub-picosecond photon-efficient 3d imaging using single-photon sensors. *Scientific reports*, 8(1):1–8, 2018. 3

[26] Felix Heide, Matthias B Hullin, James Gregson, and Wolfgang Heidrich. Low-budget transient imaging using photonic mixer devices. *ACM Transactions on Graphics (ToG)*, 32(4):45, 2013. 8

[27] Robert K Henderson, Nick Johnston, Sam W Hutchings, Istvan Gyongy, Tarek Al Abbas, Neale Dutton, Max Tyler, Susan Chan, and Jonathan Leach. 5.7 a 256 256 40nm/90nm cmos 3d-stacked 120db dynamic-range reconfigurable time-resolved spad imager. In *2019 IEEE International Solid-State Circuits Conference-(ISSCC)*, pages 106–108. IEEE, 2019. 2, 4, 8

[28] Sam W Hutchings, Nick Johnston, Istvan Gyongy, Tarek Al Abbas, Neale AW Dutton, Max Tyler, Susan Chan, Jonathan Leach, and Robert K Henderson. A reconfigurable 3-d-stacked spad imager with in-pixel histogramming for flash lidar or high-speed time-of-flight imaging. *IEEE Journal of Solid-State Circuits*, 54(11):2947–2956, 2019. 2, 4, 8

[29] Atul Ingle, Trevor Seets, Mauro Buttafava, Shantanu Gupta, Alberto Tosi, Mohit Gupta, and Andreas Velten. Passive inter-photon imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8585–8595, 2021. 1

[30] Atul Ingle, Andreas Velten, and Mohit Gupta. High flux passive imaging with single-photon sensors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6760–6769, 2019. 1

[31] Achuta Kadambi, Refael Whyte, Ayush Bhandari, Lee Streeter, Christopher Barsi, Adrian Dorrington, and Ramesh 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

[32] Martin Laurenzis and Emmanuel Bacher. Three-dimensional laser-gated viewing with error-free coding. *Optical Engineering*, 57(5):053103, 2018. 2

[33] Jongho Lee, Jenu Varghese Chacko, Bing Dai, Syed Azer Reza, Abdul Kader Sagar, Kevin W Eliceiri, Andreas Velten, and Mohit Gupta. Coding scheme optimization for fast fluorescence lifetime imaging. *ACM Transactions on Graphics (TOG)*, 38(3):1–16, 2019. 4

[34] David B Lindell, Matthew O'Toole, and Gordon Wetzstein. Single-photon 3d imaging with deep sensor fusion. *ACM Trans. Graph.*, 37(4):113–1, 2018. 8

[35] Xiaochun Liu, Sebastian Bauer, and Andreas Velten. Phasor field diffraction based reconstruction for fast non-line-of-sight imaging systems. *Nature communications*, 11(1):1–13, 2020. 2

[36] Xiaochun Liu, Ibón Guillén, Marco La Manna, Ji Hyun Nam, Syed Azer Reza, Toan Huu Le, Adrian Jarabo, Diego Gutierrez, and Andreas Velten. Non-line-of-sight imaging using phasor-field virtual wave optics. *Nature*, 572(7771):620–623, 2019. 1

[37] Ashley Lyons, Francesco Tonolini, Alessandro Boccolini, Audrey Repetti, Robert Henderson, Yves Wiaux, and Daniele Faccio. Computational time-of-flight diffuse optical tomography. *Nature Photonics*, 13(8):575–579, 2019. 1

[38] Sizhuo Ma, Shantanu Gupta, Arin C Ulku, Claudio Bruschini, Edoardo Charbon, and Mohit Gupta. Quanta burst photography. *ACM Transactions on Graphics (TOG)*, 39(4):79–1, 2020. 1

[39] Parsa Mirdehghan, Wenzheng Chen, and Kiriakos N Kutulakos. Optimal structured light à la carte. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6248–6257, 2018. 2, 5, 8

[40] Kazuhiro Morimoto, Andrei Ardelean, Ming-Lo Wu, Arin Can Ulku, Ivan Michel Antolovic, Claudio Bruschini, and Edoardo Charbon. Megapixel time-gated spad image sensor for 2d and 3d imaging applications. *Optica*, 7(4):346–354, 2020. 1

[41] Ji Hyun Nam, Eric Brandt, Sebastian Bauer, Xiaochun Liu, Eftychios Sifakis, and Andreas Velten. Real-time non-line-of-sight imaging of dynamic scenes. *arXiv preprint arXiv:2010.12737*, 2020. 2

[42] Adithya K Pediredla, Aswin C Sankaranarayanan, Mauro Buttafava, Alberto Tosi, and Ashok Veeraraghavan. Signal processing based pile-up compensation for gated single-photon avalanche diodes. *arXiv preprint arXiv:1806.07437*, 2018. 3, 7

[43] Sara Pellegrini, Gerald S Buller, Jason M Smith, Andrew M Wallace, and Sergio Cova. Laser-based distance measurement using picosecond resolution time-correlated single-photon counting. *Measurement Science and Technology*, 11(6):712, 2000. 1, 2

[44] Jiayong Peng, Zhiwei Xiong, Xin Huang, Zheng-Ping Li, Dong Liu, and Feihu Xu. Photon-efficient 3d imaging with a non-local neural network. In *European Conference on Computer Vision*, pages 225–241. Springer, 2020. 8

[45] Christoph Peters, Jonathan Klein, Matthias B Hullin, and Reinhard Klein. Solving trigonometric moment problems for fast transient imaging. *ACM Transactions on Graphics (TOG)*, 34(6):1–11, 2015. 8

[46] Francesco Pittaluga, Zaid Tasneem, Justin Folden, Brevin Tilmon, Ayan Chakrabarti, and Sanjeev J Koppal. Towards a mems-based adaptive lidar. In *2020 International Conference on 3D Vision (3DV)*, pages 1216–1226. IEEE, 2020. 2

[47] Joshua Rapp, Robin MA Dawson, and Vivek K Goyal. Dithered depth imaging. *Optics Express*, 28(23):35143–35157, 2020. 2

[48] Joshua Rapp, Yanting Ma, Robin MA Dawson, and Vivek K Goyal. High-flux single-photon lidar. *Optica*, 8(1):30–39, 2021. 1

[49] Joshua Rapp, Julian Tachella, Yoann Altmann, Stephen McLaughlin, and Vivek K Goyal. Advances in single-photon lidar for autonomous vehicles: Working principles, challenges, and recent advances. *IEEE Signal Processing Magazine*, 37(4):62–71, 2020. 1

[50] Ximing Ren, Peter WR Connolly, Abderrahim Halimi, Yoann Altmann, Stephen McLaughlin, Istvan Gyongy, Robert K Henderson, and Gerald S Buller. High-resolution depth profiling using a range-gated cmos spad quanta image sensor. *Optics express*, 26(5):5541–5557, 2018. 2, 4, 8

[51] Trevor Seets, Atul Ingle, Martin Laurenzis, and Andreas Velten. Motion adaptive deblurring with single-photon cameras. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1945–1954, 2021. 1

[52] Michael P. Sheehan, Julián Tachella, and Mike E. Davies. A sketching framework for reduced data transfer in photon counting lidar. *IEEE Transactions on Computational Imaging*, 7:989–1004, 2021. 2, 4, 8

[53] Michael P Sheehan, Julián Tachella, and Mike E Davies. Surface detection for sketched single photon lidar. *arXiv preprint arXiv:2105.06920*, 2021. 2

[54] Shuochen Su, Felix Heide, Gordon Wetzstein, and Wolfgang Heidrich. Deep end-to-end time-of-flight imaging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6383–6392, 2018. 8

[55] Julián Tachella, Yoann Altmann, Nicolas Mellado, Aongus McCarthy, Rachael Tobin, Gerald S Buller, Jean-Yves

Tourneret, and Stephen McLaughlin. Real-time 3d reconstruction from single-photon lidar data using plug-and-play point cloud denoisers. *Nature communications*, 10(1):1–6, 2019. 1

[56] Ryuichi Tadano, Adithya Kumar Pediredla, and Ashok Veeraraghavan. Depth selective camera: A direct, on-chip, programmable technique for depth selectivity in photography. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3595–3603, 2015. 2

[57] George Turin. An introduction to matched filters. *IRE transactions on Information theory*, 6(3):311–329, 1960. 5

[58] Shumian Xin, Sotiris Nousias, Kiriakos N Kutulakos, Aswin C Sankaranarayanan, Srinivasa G Narasimhan, and Ioannis Gkioulekas. A theory of fermat paths for non-line-of-sight shape reconstruction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6800–6809, 2019. 1

[59] Chao Zhang, Scott Lindner, Ivan Michel Antolović, Juan Mata Pavia, Martin Wolf, and Edoardo Charbon. A 30-frames/s, spad flash lidar with 1728 dual-clock 48.8-ps tdcs, and pixel-wise integrated histogramming. *IEEE Journal of Solid-State Circuits*, 54(4):1137–1151, 2018. 2

[60] Yongyi Zhao, Ankit Raghuram, Hyun Kim, Andreas Hielscher, Jacob T Robinson, and Ashok Narayanan Veeraraghavan. High resolution, deep imaging using confocal time-of-flight diffuse optical tomography. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. 1

[61] Vytautas Zickus, Ming-Lo Wu, Kazuhiro Morimoto, Valentin Kapitany, Areeba Fatima, Alex Turpin, Robert Insall, Jamie Whitelaw, Laura Machesky, Claudio Bruschini, et al. Fluorescence lifetime imaging with a megapixel spad camera and neural network lifetime estimation. *Scientific Reports*, 10(1):1–10, 2020. 1