Compressive Single-Photon 3D Cameras

Felipe Gutierrez-Barragan, Atul Ingle, Trevor Seets, Mohit Gupta, Andreas Velten
University of Wisconsin-Madison
{fgutierrez3, ingle, seets}@wisc.edu, mohitg@cs.wisc.edu, velten@wisc.edu

Abstract

Single-photon avalanche diodes (SPADs) are an emerging pixel technology for time-of-flight (ToF) 3D cameras that can capture the time-of-arrival of individual photons at picosecond resolution. To estimate depths, current SPAD-based 3D cameras measure the round-trip time of a laser pulse by building a per-pixel histogram of photon timestamps. As the spatial and timestamp resolution of SPAD-based cameras increase, their output data rates far exceed the capacity of existing data transfer technologies. One major reason for SPAD's bandwidth-intensive operation is the tight coupling that exists between depth resolution and histogram resolution. To weaken this coupling, we propose compressive single-photon histograms (CSPH). CSPHs are a per-pixel compressive representation of the high-resolution histogram, that is built on-the-fly, as each photon is detected. They are based on a family of linear coding schemes that can be expressed as a simple matrix operation. Our results show that a well-designed CSPH can consistently reduce data rates by 1-2 orders of magnitude without compromising 3D imaging performance.

1. Introduction

Single-photon cameras (SPC) are an emerging sensor technology with ultra-high sensitivity down to individual photons [3, 4]. In addition to their extreme sensitivity, SPCs based on single-photon avalanche diodes (SPADs) can also record photon-arrival timestamps with extremely high (sub-nanosecond) time resolution [19]. Moreover, SPAD-based SPCs are compatible with the complementary metal-oxide semiconductor (CMOS) photolithography process which can enable fabrication of kilo-to-mega-pixel resolution SPAD arrays [5, 18] at low costs. Due to these capabilities, SPAD-based SPCs are gaining popularity in various imaging applications including 3D imaging [20, 22], low-light and HDR imaging [1, 13, 16, 21], and more [15, 24, 25].

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 cap-

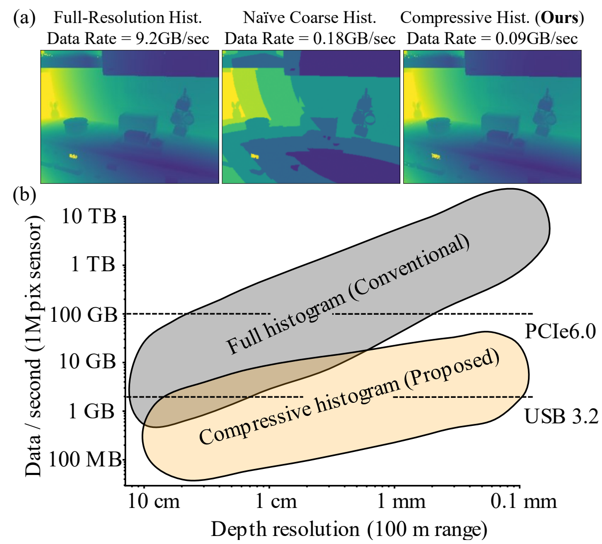

Figure 1. **Compressive Single-Photon 3D Imaging.** (a) Example depth maps with conventional (full histogram) capture, coarse resolution capture and our method with compressive capture. In this simulation, our method generates $100\times$ lower data, yet generates depth maps that are visually indistinguishable 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.

tured 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.

Consider a megapixel SPAD-based 3D camera. For short range indoor applications, a millimeter depth resolution would be desirable. For longer range outdoor applications, centimeter level depth resolution would be desirable. Assuming state-of-the-art sub-bin processing [12], this corresponds to histograms with thousands of bins. Moreover, the rate at which these histograms are acquired can vary from tens of frames per second (fps) to hundreds of fps for, say, an automotive application with high-speed object motion. 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 current data-transfer busses by orders of magnitude.

Here 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 also evaluate 3D reconstruction accuracy of our compressive acquisition method with real-world data from a hardware prototype. Please refer to the full paper for an extensive evaluation that covers various illumination conditions, scene complexity, noise, and laser pulse waveforms [10].

2. 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.

SPAD-based 3D cameras sample $\Phi(t)$ using time-correlated single-photon counting [23]. The SPAD pixel, once triggered, starts acquiring photons. After detecting one photon, its timestamp is recorded, and the SPAD is inactive for a time period called the dead time. As shown in Fig. 2, this 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, the mean photon flux at histogram bin i is:

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

The vector, $\Phi = (\Phi_i)_{i=0}^{N-1}$, is the photon flux histogram, where $N = \tau/\Delta$, and τ is the timestamp range which

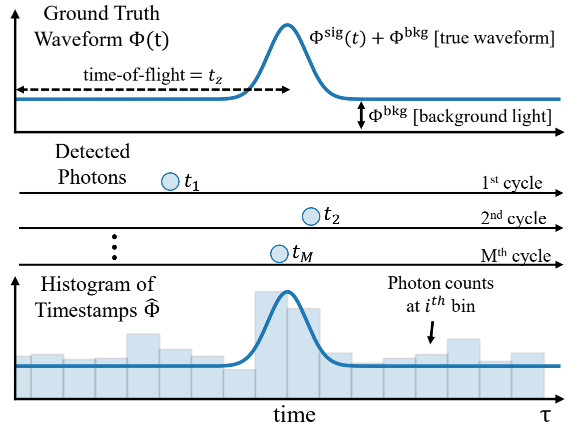

Figure 2. **Single-Photon Histogram Formation.** SPAD-based 3D cameras estimate distances by building a per-pixel histogram of the detected photon's time-of-arrival.

equals the laser pulse repetition period. Advances in SPAD operation modes [6, 12] minimize signal distortions and allow us to assume that Φ_i appropriately approximates $\Phi(t)$.

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

3. Compressive Single-Photon 3D Imaging

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. Moreover, these histograms are created one photon at a time, raising the 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 this question, we make two observations. First, there is a class of linear compression techniques which can be expressed as a 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. An effective C will have a high compression ratio (N/K), while preserving 3D imaging performance.

Second, we observe that the entire histogram can be written as the sum of 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 timestamp (T_j) detected, where all elements are 0 except for $t_{j,l} = 1$, where $l = \lfloor \frac{T_j \mod \tau}{\Delta} \rfloor$. As shown in Fig. 3, $\widehat{\Phi}$ can be written as:

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

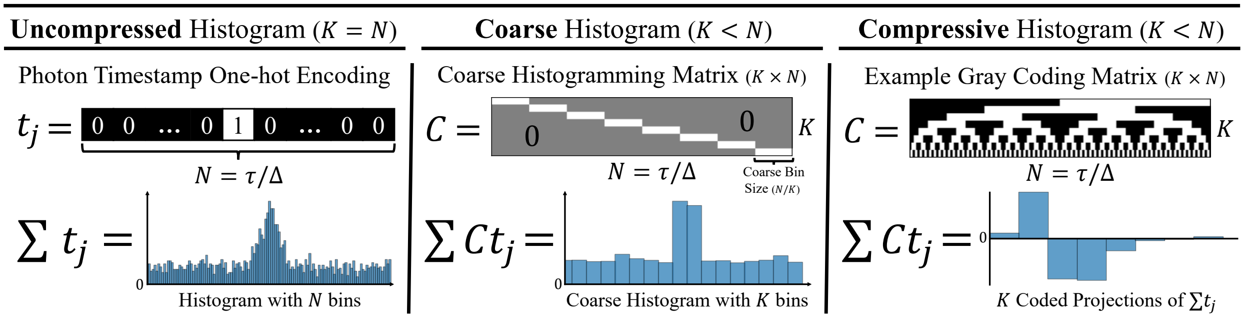

Figure 3. **On-the-fly Histogram Formation.** Timestamp histograms are often built on-the-fly, as each photon is detected. The left column shows how a histogram, whose bin width matches the timestamp resolution (Δ), is formed as the sum of timestamps represented as one-hot encoded vectors. Transferring such a large histogram for every pixel can be impractical. By multiplying each timestamp with a downsampling matrix to group timestamps into coarser bins, the size of the histogram can be reduced at the cost of resolution (middle column). Alternatively, a compressive histogram can be created by multiplying each timestamp with a coding matrix and adding them up as each timestamp comes in (right column). A well-designed C can efficiently encode the peak location from which distance can be computed.

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 $\widehat{\Phi}$. In practice, Eq. 4 need not be implemented as a matrix-vector multiplication. 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 ($\widehat{B} = \widehat{B} + C_{:,l}$). Given this on-the-fly compression method, a natural question is, what are good coding matrices for compressive single-photon 3D cameras?

Coding matrix design for 3D imaging has been studied in the context of correlation-based ToF [8, 9, 11, 14] and structured light [2, 7, 17]. Based on these works, we defined properties that C should have to achieve high compression while preserving 3D imaging performance (see [10]). Given these properties, we designed C using Gray [9] and Fourier [8] codes.

Results: To evaluate the effectiveness of CSPHs on real SPAD data we downloaded and pre-processed the data acquired with a scanning-based system [6]. The pre-processed raw histograms have $\Delta = 8$ps and $N = 832$ (e.g., histograms in Fig. 4).

Fig. 4 shows the recovered 3D reconstructions using different CSPH at 104x compression ($K = 8$). We find that Gray coding can essentially achieve 0 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

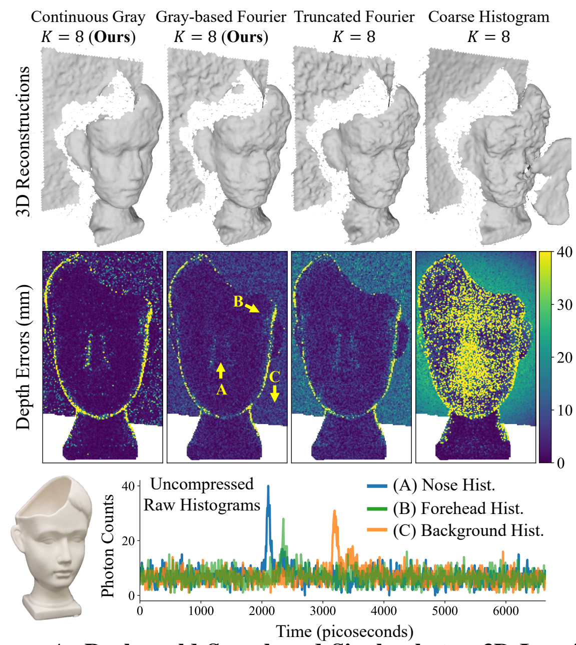

Figure 4. **Real-world Scan-based Single-photon 3D Imaging.** The depth and depth error images for different CSPH with $K = 8$ codes. The mean and median absolute errors (in mm) achieved by each method from left to right are: [7, 1], [6, 4], [9, 6], [23, 13].

histograms exhibited a longer tail than the foreground face histograms due to indirect reflections. Indirect reflections cause systematic errors in truncated Fourier, while Gray-based Fourier and Gray coding are more robust to these errors since their C have higher frequency codes, as predicted by [8]. Finally, using a conventional coarse histogram with sub-bin depth estimation leads to the worst performance.

Conclusion: High resolution SPAD-based 3D cameras can produce unmanageable data rates. To reduce their data bandwidth, we proposed to capture a compressive representation (CSPH) of the high-resolution timing histogram. An appropriately designed CSPH can preserve depth precision while outputting significantly less data.

References

[1] 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

[2] 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. 3

[3] 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

[4] 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

[5] 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

[6] 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. 2, 3

[7] 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. 3

[8] 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. 3

[9] 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. 3

[10] Felipe Gutierrez-Barragan, Atul Ingle, Trevor Seets, Mohit Gupta, and Andreas Velten. Compressive single-photon 3d cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2022. 2, 3

[11] 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. 3

[12] 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. 1, 2

[13] 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

[14] 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. 3

[15] 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

[16] 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

[17] 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. 3

[18] 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

[19] 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

[20] 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

[21] 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

[22] 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

[23] Chao Zhang, Scott Lindner, Ivan Michel Antolović, Juan Mata Pavia, Martin Wolf, and Edoardo Charbon. A 30-frames/s, 252×144 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

[24] 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

[25] 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