Luminance-depth reconstruction from compressed time-of-flight histograms

Valentin Poisson, Van Thien Nguyen, William Guicquero, Gilles Sicard

Luminance-depth reconstruction from compressed Time-of-Flight histograms

Valentin Poisson, Thien Van Nguyen, William Guicquero and Gilles Sicard

***Abstract*—Single photon avalanche diodes (SPADs) combined with high-frequency time-to-digital converters (TDCs) enable the estimation of photon Time-of-Flight (ToF) for active 3D-depth imaging. Nevertheless, SPAD sensors still face hardware limitations due to a complex pixel readout design and a large amount of data collected by way of pixel-wise histograms. The intrinsic high background illumination (BI) also remains a challenging issue for the related depth reconstruction algorithms. Using a physically-relevant SPAD sensor model, this work tackles these issues by implementing a pixel-wise ToF histogram compressive sensing (CS) with a specific deep generative model based reconstruction. It demonstrates a possible reduction of hardware design constraints while reaching a depth inference Root Mean Square Error below 16 centimeters regardless of BI (50 – 1050 $\mathrm{W/m^2}$) and distance (20 m), at a compression ratio (CR) of 10% (32 CS measurements). In addition, this paper introduces a novel multimodal reconstruction from SPAD data, enabling joint depth and luminance estimations. Indeed, since ToF histogram raw data gathers multiple physical scene characteristics, we propose a two-part DGM capable of inferring Super-Resolved depth maps and normalized luminance images, independently from the average scene BI. Our key contributions related to the DGM topology design are proper normalization layers with a learned pile-up effect compensation, multidimensional-multiscale filtering and concatenations of Softmax-ReLU activation functions to capture both peak-position and relative amplitude features. Numerically, depth and luminance maps reconstructions of natural scenes respectively reach more than 30dB and 25dB PSNRs for any CR higher than 2.5%.**

***Index Terms*—3-D imaging, computational imaging, LiDAR, low-light imaging, poisson noise, single-photon detection, Time-of-Flight imaging, Super-Resolution, statistical compressive sensing, multi-modal, deep generative model.**

I. Introduction

Since the invention of SPAD CMOS sensors in 2003 [1], several specifications such as power consumption, timing resolution, dynamic range, and pixel-level processing notably improved the performances of these sensors. Indeed, it enabled their use for a wide range of emerging imaging application domains like, medical imaging [2], astronomical imaging [3], lidars [4] and light communication [5]. To address mass-market industrial sectors (*e.g.*, automotive [6]) with high-end performances, the current trend is to develop larger SPAD arrays [7] benefiting from advanced technology nodes, using 3D-stacking (3D-IC) [8] to enhance capabilities at a reduced power consumption and a smaller footprint than 2D-IC [9].

This paper will focus on direct time-of-flight (D-ToF) imaging which consists in measuring the round trip time of a light pulse whose laser source is synchronized with the sensor [10]. Unlike Indirect Time of Light (I-ToF) in which the depth measurement is performed by estimating the phase shift of a continuously modulated light wave [11]–[13]. However, although SPAD based D-ToF sensors now offer a competitive depth accuracy compared to the alternatives(*e.g.*, 1.6mm at 3m [14]) , they still face several hardware constraints which limit higher image formats (*i.e.*, number of pixels) with a reduced pixel pitch, better timing accuracy (*i.e.*, depth precision) and increased temporal resolution (*i.e.*, frame rate).

A SPAD-based Time-correlated single photon counting (TCSPC) sensing generally consists in detecting incident photons and storing measured times of arrival in a histogram, in order to estimate the statistical distribution over a large number of pulse periods. Depending on incident photon flux, the histogram may include a lot of photon arrival times due to ambient light, in addition to the ones from the gated laser pulses. A major constraint is thus related to photons from the background illumination (BI) which are not issued by the synchronized active illumination. In other words, the extraction of useful information in such a high BI turns out to be complex. Indeed, in some cases more than 99% of the detected photons may come from BI [15].

Furthermore, SPAD sensors are mainly limited by the memory requirements related to the ToF histograms. Let's take the typical example of a 10-bit coded ToF value, the histogram would therefore be composed of 1024 bins providing measurements with a resolution of 1cm for a depth range of 10m. To retrieve a relevant distribution of ToF values, each bin must have a counter with a sufficient bit-depth (*e.g.*, 10bit). In this example, the per-pixel memory is therefore of 10kb, thus involving a large silicon surface at pixel-level. Finally, the overall sensor data throughput remains another crucial point, which also relates to spatial resolution. The current trend is to increase the spatial resolution to reach the so called Megapixel resolution or beyond. However, with only a VGA pixel matrix (640×480), it would already require more than 3Gb per frame. For a frame rate of 30fps, this would imply a 92Gb/s chip data throughput, which can be put into perspective in relation to the equivalent RGBZ data stream with the same image format and bit-depth that would be lower than 400Mb/s. So, even assuming high-speed chip I/O pads capable of 300Mb/s it would still require more than 300 pads! Canonical histogram-based SPAD architectures therefore seem going to a dead-end,

Manuscript received February 10, 2021; revised June 29, 2021; accepted Month X, 2020. Date of publication Month X, 2021; date of current version June 29, 2021.

Valentin Poisson, Thien Van Nguyen, William Guicquero, and Gilles Sicard are with the CEA-LETI, University Grenoble Alpes, F-38000 Grenoble, France (e-mail: valentin.poisson@cea.fr; Vanthien.NGUYEN@cea.fr; william.guicquero@cea.fr; gilles.sicard@cea.fr).

hence motivating the use of a histogram compression scheme.

The goal of this paper is thus twofold, firstly to demonstrate the need for pixel-wise SPAD histogram compression at sensor level and secondly to provide a processing approach involving multimodal reconstruction with possible Super Resolution.

A. SPAD Imaging Prior Work

1) Recent advances on SPAD imagers: Single photon avalanche diode (SPAD) has recently emerged as a key technology in numerous single photon imaging applications. Contrary to photomultiplier tubes (PMTs) [16] and microchannel plates (MCPs) [17], [18] that previously were sensors of choice in single photon detectors, SPAD-based imagers have the great advantage to enable embedded digital processing thanks to the use of a standard CMOS process. The last two years of SPAD development took advantage of several hardware components upgrades, enabling sensors power consumption reduction and spatio-temporal resolution improvements [13], [19]. In particular, we can identify two recent outstanding SPAD developments. Firstly, [7] finally succeeded to present the world's first MegaPixel SPAD sensor, in spite of numerous hardware-related design constraints. Secondly, the high-end device proposed by [9] is composed of a 256×256 SPAD CMOS sensor integrated into 3D-stacked $90\mathrm{nm}$ with a reconfigurable operating mode allowing variable trade-off between spatial and timing resolutions (*i.e.*, spatial resolution of 64×64 to 256×256 and timing resolution of $560\mathrm{ns}$ to $0.56\mathrm{ns}$). Future works will undoubtedly have the same tendencies by aiming at increasingly targeting higher spatial and temporal resolutions, implemented in even smaller device units and dedicated to high volume of industrial applications (*e.g.*, automotive and Internet-of-Things sectors).

2) Unconventional sensing strategies: Besides technological trends, we review in the literature various unconventional system design to overcome SPAD limitation such as pixel-wise memory needs, sensor data throughput or even noise photon detection filtering.
For example, [15] attempts to optimize generated SPAD data, by proposing two methods which consist in dividing the histogram construction into subparts to overcome the previously mentioned hardware limitations. Unfortunately, although reducing hardware constraints, these methods decrease the sensor acquisition performances in a way. Likewise, [20] replaces the TDC-per-SPAD scheme with a shared TDC architecture. It thus allows to optimize the use of pixel area despite of inducing a lower spatial resolution.

As previously mentioned, SPAD imaging is subject to high photon noise, modeled by Poisson process, due to the photons coming from background light. To prevent this, [21] implements photon coincidence detection in order to perform noise-filtering. Shared counters across pixel subgroups yet cause spatial resolution reduction due to the complex embedded processing that they require [9]. Another way to attenuate noise distortion, presented in [22], is to misalign SPAD listening windows with various shifts in order to spread out the noise effect and mitigate the pile up effect [23]. This method implies a possible binary code reduction of histogram bins and thus lower pixel memory needs. Nevertheless, it may be only applied to specific SPAD operating systems at one photon per pulse detection. Finally, [24] proposed to shorter time sensitive windows via time-gating. In spite of noticeable benefits (*e.g.*, noise filtering, minimizing storage/readout data), a clear limitation of this is related to the need for either a proper prior selection of this sensitive period or an increase of the overall amount of collected data.

Apart from sensor-centric acquisition scheme variants, single pixel camera [25] approach has already been investigated in the context of SPAD pixels. Thanks to CS theoretical results on l1 regularization-based reconstruction and matrix mutual coherence, this approach does not require any additional sensors to reconstruct depth maps from a single photosensitive component. For example, [26]–[29] achieve high spatial resolution by using a Spatial Light Modulator (Digital Micro-mirror Device, DMD) placed in front of a single photon counting detector or SPAD array to perform CS projections in the optical domain. Unfortunately, in practice, in addition to involving the use of a bulky DMD, this kind of system implies unacceptable drawbacks related to its calibration setup and proper optical alignments.

3) SPAD data processing: Besides to filtering and optimizing measurements system for SPAD data generation, processing approaches are essential for features extraction with high accuracy. Indeed, even though peak detection is the most commonly used approach to get depth information [30], statistical based inference have been proposed to improve performances (*e.g.*, Gaussian fit methods [31]).

Furthermore, extensions of statistical inference have been developed in order to extract more features or even to reduce data throughput. In fact, SPAD sensors provide multimodal information, where the amplitude of the histogram is related to the reflectivity of the scene, and depth estimation is related to a peak position. This way, [32]–[36] developed reconstruction methods to estimate depth and reflectivity from SPAD raw data using a maximum likelihood approach in low and high flux imaging. More recently, [37] proposed a histograms compression framework via sketching, using compressive learning [38].

On the other hand, deep learning considered as a data-driven method used to infer features from observations, meets all the requirements for processing SPAD data. Therefore, many deep learning models emerged to estimate intensity-depth maps [39], [40]. Indeed, data processing methods will most likely play an important role in overcoming aforementioned SPAD hardware limitations (*i.e.*, data throughput, data storage, photon detection efficiency). For example, such a processing method would involve estimating a high-resolution image from a single or multiple low-resolution images, thereby relaxing hardware constraints. This processing task is referred as Super Resolution (SR) reconstruction (see [41] for an overview of deep learning techniques related to SR). Focusing on depth

map SR, the most common approach is image-guided depth up-sampling [42]–[45] which consists in reconstructing a high resolution depth map from a full-scale RGB frame combined with its corresponding low resolution depth map.

B. Main contributions

1) Compression: This paper first presents a novel pixel-wise compressive sensing (CS) approach which appears to be efficient in various SPAD operating modes, in single photon per pulse detection as well as in multiple photon detection per pulse regardless of Poisson noise level. Thanks to a custom hardware implementation, this method would allow to reduce the data memory needs without major loss of reconstruction accuracy. To support this, we constructed a "pseudo-realistic" SPAD simulation model to demonstrate that SPAD performances can still be maintained (*i.e.*, frame rate, dynamics range,...). Applying this compression at a CR of 20% still enables a PSNR of the depth reconstruction of at least 40dB in all BI conditions.

2) Reconstruction: We exploit all modalities provided by a SPAD sensor measurements (*i.e.*, intensity and depth), to improve features reconstruction performance. To do so, we use a multimodal Deep Generative Model architecture to capture correlations across modalities in order to jointly enhance intensity-depth map reconstructions (*e.g.*, depth discontinuities and edges of the object in intensity maps). In addition, this reconstruction approach is straightforwardly compatible with intensity-depth maps SR when considering each ToF histogram acquired by a pixel as being generated by a subgroup of virtual pixels.

C. Outline

The remainder of this paper is organized as follows. We first present a physical model to simulate a SPAD operating system in "Real World" conditions by generating relevant photon times of arrival depending on physical quantities (*i.e.*, background illumination level, object albedo and depth). In the second part, we address the SPAD sensor data-deluge issues (*e.g.*, pixel memory footprint, chip I/O data throughput) by taking advantage of a pixelwise histogram dimensionality reduction through a randomly generated projection matrix (as CS [46] would do) and combine it with a Deep Generative Model (DGM) [47] to reconstruct the depth map. In the third part, the acquisition/reconstruction problem is revisited in order to further exhibit the relevance of a multimodal reconstruction, also enabling Super Resolution.

II. SPAD Imaging - System Model

In the context of active reflection imaging (see Fig. 1), the D-ToF method consists in measuring the propagation time of a light pulse, emitted by the transceiver (*i.e.*, LASER), and then received by a synchronized receiver (*i.e.*, SPAD). Thanks to a high frequency Time-to-Digital Converter, the single photon detection sensor sequentially records every photon arrival time. This data is then collected through a per-pixel histogram, where the number of photon arrivals over each measured time interval is accumulated in the corresponding bin of the histogram, $\mathbf{h}$, as depicted in Fig. 1.

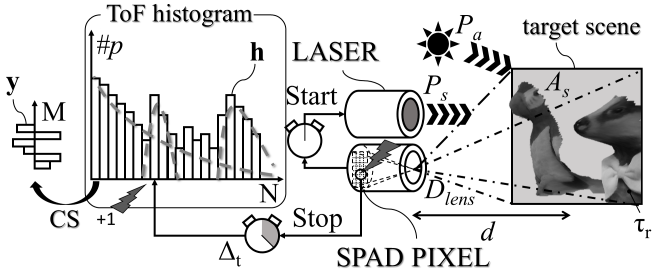


Fig. 1: SPAD Operation System Overview: direct time-of-flight measurement (D-ToF) of a light pulse reflected by a target using a time-correlated single photon counting (TCSPC) system.

This section presents an active 3D imaging model (inspired by [21]) to generate synthetic SPAD raw data by taking into account both the physical phenomena related to the environment as well as the hardware characteristics of the transceiver and the receiver. This model will later be used to evaluate our acquisition/reconstruction methods.

A. Active D-ToF imaging model

To generate more realistic SPAD data, we developed an end-to-end physical simulation model except that it focuses only on direct and not global illumination (*i.e.*, due to inter-reflections between objects of the scene) because of its second-order impact. The direct illumination of the scene towards the sensor is thus from the reflection of two light sources, which are a laser source (with power P_s [W]), and an ambient light source (with intensity P_a [W/m^2]). For the sake of physical simulation relevance, several configurations will be taken into account, regarding the hypothesis of ambient light intensity, P_a. For an indoor scene, ambient light illumination is of the order of 5 to 35klux ($P_a < 350\mathrm{W/m^2}$) and between 50 - 100klux ($P_a < 1050\mathrm{W/m^2}$) for an outdoor scene. Besides, we denote P_{AS} the solar radiation integrated over the laser-illuminated surface (cf. [21]),

$$P_{AS} = P_a \cdot A_s = P_a \cdot 4d^2 \cdot \tan\left(\frac{\theta_H}{2}\right) \cdot \tan\left(\frac{\theta_V}{2}\right), \quad (1)$$

where d is the distance to the object and θ_V, θ_H are the aperture angles of the laser. We assume that the illumination of the scene by the laser is uniform due to the invariance of the electric field in the orthogonal plane to the direction of wave propagation. Therefore the pixel-wise light power related to the laser pulse reflection is estimated (cf. [21]) by,

$$P_s^* = P_s \cdot \tau_r \cdot \frac{2}{\pi}\left(\frac{D_{lens}}{2d}\right)^2 \cdot \tau_o \cdot \delta_{bw} \frac{FF \cdot PDP}{\mathrm{H \cdot W}}. \quad (2)$$

And the pixel-wise light power due to ambient light (cf. [21]),

$$P_a^* = P_{AS} \cdot \tau_r \cdot \frac{2}{\pi}\left(\frac{D_{lens}}{2d}\right)^2 \cdot \tau_o \cdot \delta_{bw} \frac{FF \cdot PDP}{\mathrm{H \cdot W}}. \quad (3)$$

D_{lens} and τ_o are the optical parameters determining the diameter of the lens and the efficiency of the optical system. The integral of the light power considering the efficiency of the bandpass filter along all wavelengths is approximated by a single coefficient noted δ_{bw} (where the BI power density is considered uniform over the scene). Note that the pulse laser power reflected to the SPAD is subject to the inverse square law (*i.e.*, inversely proportional to the square of the flight distance) whereas BI is independent to the object distance d (due to the simplification of d while expanding P_{AS} in Eq. 3). A pixel Fill Factor (FF), a pixel array size (H $\times$ W) and a probability of photon detection (PDP) need to be defined too, as described in [48]. Despite the invariance of the scattering intensity related to the solid angle of the reflecting object (specular reflectance [49] is not considered), the reflectance intensity level of each object of the scene may still vary as illustrated by different shades of gray in the Fig. 1. For the sake of simulation purposes – which is not strictly following laws of physics – the albedo τ_r (for the considered spectral band) is modeled as the luminance component, Y, of its related RGB image. From those light powers, P^* (P_s^*, P_a^*), photon rates, R (R_s, R_a) are estimated,

$$R = \frac{\lambda P^*}{hc}, \tag{4}$$

where the Planck constant h allows to relate the energy of a photon to its frequency $f = \frac{c}{\lambda}$.

B. SPAD histogram generation process

For each of these pixel-wise photon flux, inter-photon waiting times $\{x[i]\}_{i\in\mathbb{N}}$ respect the memoryless property, *i.e.*, the waiting time between two successive photons is exponentially distributed and independant to an arbitrary starting time $\omega \geq 0$ as stated in [50],

$$\mathbb{P}(X > \omega + x | X > \omega) = \mathbb{P}(X > x), \tag{5}$$

$$\mathbb{P}(X > x) = \exp(-Rx). \tag{6}$$

In order to further simulate SPAD measurements, we first aim at generating synthetic random inter-photon arrival time samples $\{x[i]\}_{i\in\mathbb{N}}$. To this end, we thus performed a Monte Carlo sampling strategy using an exponential random number generator with a rate parameter equals to the photon flux R.

Then, lets denote $\{t[i]\}_{i\in\mathbb{N}}$ the theoretical photon arrival time sequences indexed i for a unique light source,

$$t[i] = t[i-1] + x[i]. \tag{7}$$

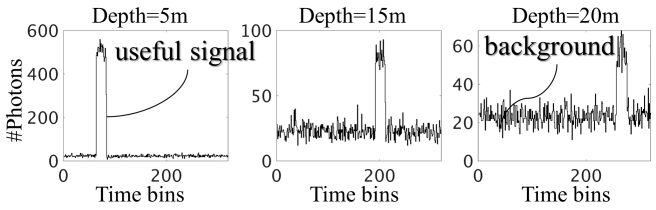
$\{t_a^f[i]\}_{i\in\mathbb{N}}$, $\{t_s^f[i]\}_{i\in\mathbb{N}}$ and $\{t_{dcr}^f[i]\}_{i\in\mathbb{N}}$ are respectively representing the sequences of arrival times of the three main contributors: BI, laser signal and dark count rate (DCR) at inter-frame f.

Note that one specificity of SPADs is that they are reverse-biased at a higher voltage than the breakdown voltage (*i.e.*, Geiger Mode [51]), thus a single charge carrier can trigger a self-sustaining avalanche. This operating mode implies a said "Dead Time" (DT) [10], that corresponds to the SPAD quenching time to be raised again above breakdown voltage to detect another photon. Measured SPAD photon arrival times sequences $\{t_{SPAD}[i]\}_{i\in\mathbb{N}}$ regardless inter-frame f, thus cannot be the straightforward concatenation of the three sequences. This is performed by an iterative process sequentially retaining the first arrival time and then removing photons arrival times from the three sequences that are below $t_{SPAD}[i] + D_t$ (see Alg. 1), where D_t refers to the dead time value. Every arrival times in all inter-frame (the number of inter-frames is equal to the ratio between the sensor frame rate and the laser repetition rate, $\lfloor \frac{fps}{Fe} \rfloor$) are consequently at least one DT apart.

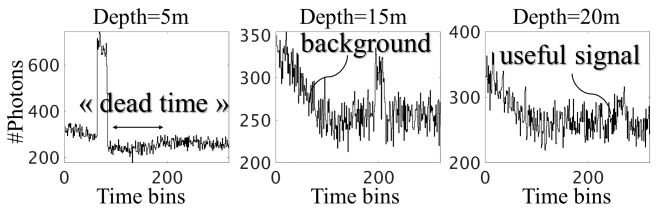
Algorithm 1: Generation of synthetic SPAD measurements from simulated sequences of photon time arrivals, for three independent light sources

```
for f = 1 : ⌊fps/Fe⌋ do
    while min(t_a^f, t_s^f, t_dcr^f) < 1/Fe do
        t_SPAD[i] ← min(t_a^f, t_s^f, t_dcr^f)
        t_a^f ← t_a^f \ {∀j, t_a^f[j] ≤ t_SPAD[i] + D_t}
        t_s^f ← t_s^f \ {∀j, t_s^f[j] ≤ t_SPAD[i] + D_t}
        t_dcr^f ← t_dcr^f \ {∀j, t_dcr^f[j] ≤ t_SPAD[i] + D_t}
        i++
    end
end
h ← hist(t_SPAD)
```

Besides, other sensor defaults such as afterpulsing [52], or crosstalk [53] are not considered because of their small influence. Fig. 2 illustrates simulated SPAD histograms, $\mathbf{h}$, of $\mathbf{t}_{SPAD}$ under typical scene conditions.



(a) $P_a = 50\mathrm{W/m^2}$ BI.



(b) $P_a = 1050\mathrm{W/m^2}$ BI.

Fig. 2: Several examples of per pixel Time-of-Flight histograms corrupted by a low BI (a) and a high BI (b), for three different object distances (5m, 15m and 20m). The laser Pulse Width is of 5ns while the bin width is of 260ps and the SPAD DT equals to 27ns.

Contrary to the SPAD model depicted in [39], it therefore provides more complex data without neglecting DT that in-

duces histogram distortions (*e.g.*, pile-up effect and rebounds as illustrated in Fig. 2). This model also takes into account photon efficiency variation from objects albedo which will allow us to demonstrate the use of a multimodal DGM from compressed histograms. Our generated histograms typically well correlate with the ones reported in the white paper [54], especially for a configuration with a wide laser pulse width and a high BI. Note that, as shown in Fig. 3 that describes our main contributions, the rest of this article will rely on this histogram modeling to derive data-driven based approaches, used for constructing both training and testing databases.

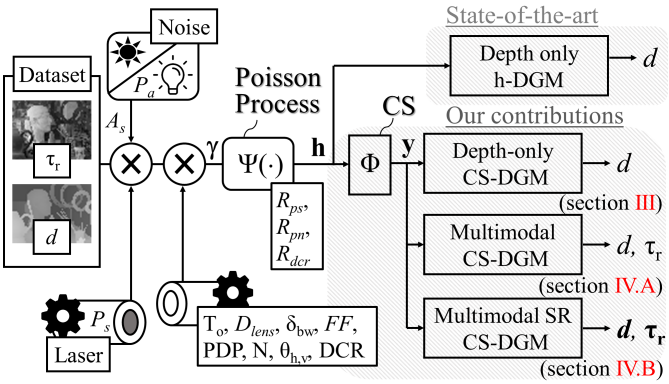


Fig. 3: Schematic description of our proposed SPAD simulation model and the contributions of this paper.

III. Depth from compressed ToF histograms

A. Histogram compressive sensing

To jointly address the aforementioned Hardware implementation issues, we introduce a pixel-wise histogram compression. Compared to previous works [55], [56], the proposed compression scheme is inspired from CS [57], which only consists of performing a set of pseudo-random projections of $\mathbf{h}$. Under the light of recent CS-imager works in various field such as, low power [58], high frame rate [59], embedded inference [60], if properly designed, this proposed approach seems compatible with a future on-line, in-pixel implementation without the need for bulky and complex embedded processing. Generally speaking, it has been proven that CS can efficiently sense sparse signals $\mathbf{s} \in \mathbb{R}^{\mathrm{N}}$ [61], [62] thanks to a sensing matrix $\Phi \in \mathbb{R}^{\mathrm{M}\times\mathrm{N}}$, this without major loss of information while providing a number of samples far below the original signal dimensionality (*i.e.*, M << N). CS theoretical foundations are then twofold, first it assumes that the sensing matrix is said incoherent with the sparsity representation basis [57], it then supposes that the signal is sufficiently sparse with respect to the targeted Compression Ratio (CR=M/N) [63]. Furthermore, in the name of a universal sensing strategy, it has been demonstrated that a pseudo-randomly generated Φ [64] enables a sufficient incoherence with any sparse representation basis. Then, although conventional CS assumes signal sparsity (*i.e.*, there exists a representation basis in which the signal exhibits very few non-zero coefficients), in practice the noisy case generalization is rather considered where the signal is said "compressible" [65] (*i.e.*, most of the coefficient modules are below a certain low threshold).

However, there exists no simple sparse representation of $\mathbf{h}$, at least at first glance (cf. Fig. 2). We can yet mention that in the specific case of [66], CS principles are extended to deal with the acquisition of a signal following a mixture of distributions model and not a sparsity-based model. Moreover, recent advances on CS reconstructions now show the interest of using Deep Generative Models to infer signals from their CS measurements [67] instead of finding the best sparse solutions using an iterative regularization-based algorithm [68]. In addition to providing a better signal rendering [69], this reconstruction strategy have two other main advantages compared to a conventional one. Firstly, it no more requires to define a specific sparse representation by hand [70]. Secondly, it enables the use of more complex, non-linear acquisition schemes [71]. Indeed, thanks to a proper DGM topology design and training, the algorithm would be capable of finding the best non-linear mapping between collected data (*i.e.*, in a latent space) and its restoration [47], [72].

In our specific case, this work aims at taking advantage of those recent advances on CS applied to SPAD histogram compression, with a limited degradation of the accuracy in terms of depth estimation. Mathematically speaking, for the sake of formalism, let's denote Ψ a generative function that takes as input a limited set of parameters $\gamma \in \mathbb{R}^{\Gamma}$ ($\Gamma = 3$: depth, albedo and background illumination) to provide a pixel-wise ToF histogram ($\mathbf{h} = \Psi(\gamma)$) being almost equivalent to a mixture of distributions [73] with a limited set of parameters. Hence, we can state that the histogram has an intrinsic latent representation γ while being corrupted by noise due its stochastic nature. Ψ thus represents the simulation model operator described in section II. Inferring γ from $\mathbf{h}$ could thus be tackled by a deep learning strategy since the network topology is capable of approximating this inverse mapping Ψ^{-1} and trainable to perform this specific task. Actually, this kind of deep learning approach has already drawn attention from the SPAD scientific community [39], [40]. Because, Γ is small compared to the dimensionality of a N-bin histogram $\mathbf{h}$, we can reasonably expect that $\mathbf{h}$ is of a compressible nature. We therefore propose to use a universal sensing strategy to capture most of the information with the smallest number of measurements, hence providing the so called "measurement vector" $\mathbf{y} \in \mathbb{R}^{\mathrm{M}}$:

$$\mathbf{y} = \Phi\Psi(\gamma) = \Phi\mathbf{h}. \tag{8}$$

Furthermore, regarding in-pixel hardware implementation, because it is unwanted to store all coefficients of Φ on-chip as it would be necessary if Φ was considered as an encoder in an autoencoder framework; Φ needs to be generated from an embedded pseudo-random (but deterministic) process, for instance following a Rademacher distribution as in [65]. Note that such a binary distribution can be efficiently generated either by a Linear Feedback Shift Register (LFSR) [74] or by a cellular automaton with chaotic behaviour [75], [76]. Such a system could be efficiently implemented by replacing

the standard one-hot encoding combined with unsigned N counters by sequentially generated rows of Φ combined with signed M counters. Managing $\mathbf{y}$ instead of $\mathbf{h}$ at pixel level will thus highly relaxes hardware constraints. All the remainder of this paper deals with this kind of CS-SPAD data (*i.e.*, using the same ± 1 Rademacher sensing matrix with zero-mean Φ, for all pixel-wise CS measurements).

B. Compressive sensing deep generative model

In the context of single photon detection, depth reconstruction from a photon count histogram aims at operating under a high ambient illumination conditions. For instance, [32], [77] attempts to estimate depth with a customized maximum likelihood method. On the other hand, [22] proposed an asynchronous detection scheme in order to compensate for the pile-up effect using a log-match filter to make noise uniform and to easily estimate depth from peak detection. Finally, [39] introduce deep learning approaches to single photon detection imaging for depth estimation. Consistently with these recent works, our depth reconstruction made from SPAD-CS data $\mathbf{y}$ instead of the histogram $\mathbf{h}$ itself is also based on a specific deep learning topology as illustrated in Fig. 4.

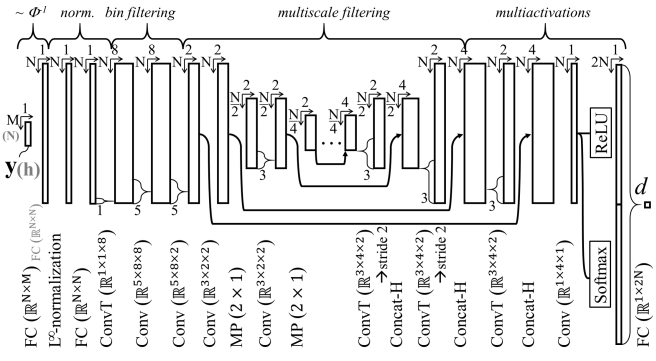


Fig. 4: Deep Generative Model topology for pixel-wise depth reconstruction. Note that the only difference between reconstruction from histogram inputs and from CS is the first Fully Connected (FC) layer. Concat-H corresponds to horizontal concatenation and MP corresponds to Max Pooling.

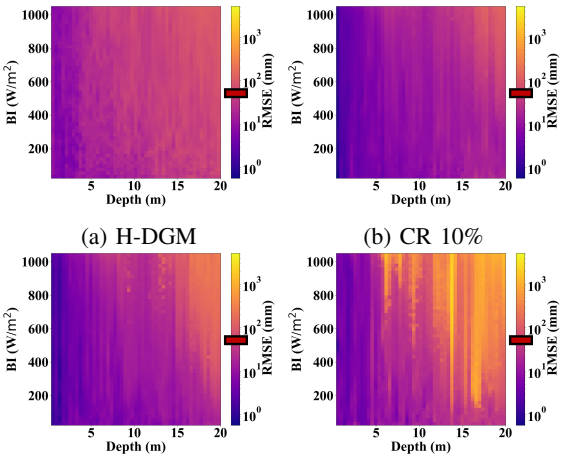
This topology is designed to reconstruct a single depth measurement per pixel, corresponding to its related CS measurement vector $\mathbf{y}$. Apart from being relevant to decompress CS data (*i.e.*, somehow inverting Φ), this unique DGM topology allows for a fair comparison between reconstructions made from histogram data (H-DGM) and compressed histograms (CS-DGM). This said pixel-wise CS-DGM embeds a Fully Connected (FC) layer to map the input to the histogram dimensionality. This FC also intrinsically performs a bin weighting to feed the L^∞ normalization, making the topology more robust against the high histogram scaling variability due to pixel-to-pixel variable reflectivities τ_r. Local bin filtering layers then enable to compensate the residual pile-up effect and facilitate the detection of specific waveforms (*e.g.*, shaped laser pulses). Multiscale filtering allows to enforce local bin collaborations to efficiently perform denoising operations. The goal of the last layer is in a way equivalent to a peak position detector as a Softargmax would be [78]. Indeed, we noticed that a Softmax followed by a learned FC performs at least at par, because of enabling more degrees of freedom during the learning stage while not implying unwanted overfitting. Finally, concatenating the ReLU output with Softmax further improves the reconstruction letting the final projection efficiently combine position and amplitude information. Note that FC, convolutional (Conv) and transpose convolutional (ConvT) layers are all followed by the rectified linear unit (ReLU) activation functions. In order to argue for future hardware optimized implementations, this subsection aims at demonstrating the efficiency of only-depth inference from SPAD-based compressed histograms in comparison with depth reconstruction from its uncompressed counterpart. We therefore evaluate our proposed acquisition scheme, with predefined parameters of the simulation model (section II) according to the ones reported in [21], [48] (cf. table I). Note that most of the techniques for accurate depth measurements only consist in histogram peak detection, this unfortunately led to the use of laser with a very short pulse duration and high illumination power (*e.g.*, 100ps of pulse duration [9]). Fortunately, our approach relies on the waveform detection instead, implying that such high performance laser is not required. This way, laser illumination pulse duration has been set to a few nanoseconds instead of hundreds of picoseconds [79]. Receiver parameters (*e.g.*, imaging range and resolutions) are thus determined according to the typical performance of a SPAD sensor [9]. Whereas environment parameters such as background light intensity are estimated through SPAD data simulation works [21], [32], [48].

Parameters	
Max range	25m
Repetition rate, Fe	12MHz
Wavelength, λ	780nm
Background light, P_a	50 - 1 050 W/m^2
Dead Time, D_t	27 ns
Object reflectivity, τ_r	1/256 - 1
Pulse Width, PW	5ns
Mean illumination power, P_s	12 mW
Field of View, (θ_H-θ_V)	30°- 40°
Diameter of collecting lens, D_{lens}	11mm
f-number	1,4
Focal length	15mm
Optical filter bandwidth, δ_{bw}	20nm
Optic efficiency, τ_o	0,7
Photon Detection Probability, PDP	0,4
Pixel Fill Factor, FF	0,7
TDC dynamic range, N	1-320
TDC Resolution, $(\mathrm{N} \times Fe)^{-1}$	260ps (78mm)

TABLE I: This table reports all simulation parameters used for subsections III and IV.

C. Pixel wise depth estimation

Since the above reconstruction model, illustrated in Fig. 4, is a pixel-wise method; the use of a realistic dataset with spatial correlation between pixels to validate our approach, is not required. Besides, we generate an unrealistic RGB-D train dataset with a random uniform generator, in order to get a higher variability of depth-luminance combination and cover all the measurement dynamic. Knowing that the reconstruction



(a) H-DGM (b) CR 10%

(c) CR 5% (d) CR 2.5%

Fig. 5: RMSE simulation results for 3D imaging at various Compression Ratio with a scale indicator of the resolution and an indicator of the simulated TDC resolution (78mm).

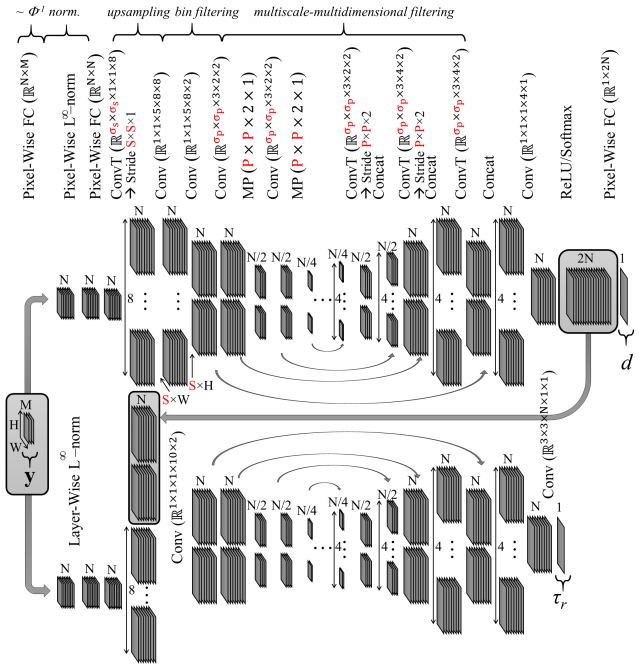


Fig. 6: Deep Generative Model topology for depth-luminance reconstruction from compressed histograms.

method is pixel-wise, Signal-to-Background Ratio (SBR) thus only depends on the BI. Indeed the SBR variation results from the linear product of the object albedo, τ and the BI, P_s, as seen in Eq. 3 and in Eq. 2. We thus estimate the depth Root Mean Square Error (RMSE) according to BI for various CR acquisition methods. Results reported in Fig. 5 were obtained by simulating multiple BI, P_s, over uniformly distributed ToF values, d (*i.e.*, 48×64 depth measurements). Each pixel of this RMSE map corresponds to the average error of an image reconstruction of 48×64 pixels with the depth value and BI value associated with it. Note that pixels coordinates of the 50×50 RMSE map correspond to the ground-truth depth and a certain BI level, where the x-axis step is of 0.4 m and the y-axis step is of 21 $\mathrm{W/m^2}$. No matter what method is used (*e.g.*, H-DGM method or the compressive sensing methods), we can observe that pixel-wise noise appears for long distance objects in case of low SBR (cf. Fig. 5, RMSE increases while increasing depth and BI). Fig. 5 shows how the performance of compression-reconstruction methods are affected by changing the depth measurements and the background illumination (BI) compared to the uncompressed counterpart. Furthermore, since the RMSE for some BI/Distance configurations is far below the simulated TDC resolution (78mm), this proposed CS acquisition system would even allow for a higher temporal resolution, without the expense of additional memory (*i.e.*, CS measurement vector $\mathbf{y}$ size is not correlated to the temporal resolution). In other words, our CS acquisition system could benefit from higher TDC clock frequency without any other hardware constraints.

IV. Deep digging into latent CS-SPAD data

A. Two-part multimodal DGM

In this subsection, we propose to build a Deep Generative Model being trained to disambiguate multiple modalities (luminance τ_r and depth d) from a latent space in the name of compressed SPAD histograms $\mathbf{y}$ while being as least as possible sensitive to BI. Our multimodal reconstruction algorithm directly takes advantage of the deep learning framework presented in the previous section. As mentioned in section III, our SPAD data model mainly takes into account the most relevant physical characteristics denoted d and τ_r. The DGM topology presented in Fig. 6 therefore aims at jointly inferring d and τ_r from $\mathbf{y}$, for all pixels simultaneously. The Depth part of this topology (*i.e.*, upper parts of Fig. 6) slightly differs from the one reported in Fig. 4 by including a local spatial collaboration during the multiscale filtering stage while using $\sigma_\mathrm{P} = 3$ and $\mathrm{P} = 2$ ($\sigma_\mathrm{S} = 1$, $\mathrm{S} = 1$) removing the strict pixel-wise characteristic of the system. For the luminance part (*i.e.*, down parts of Fig. 6), the L^∞ normalization is performed on the whole layer instead of being pixel-wise, in order to compensate for the average intensity (*i.e.*, frame-based) of the data, making the reconstruction more robust against BI changes. In addition, this luminance part also takes as input the denoised version of the histogram coming from the last part of the Depth DGM. The residual connection between depth part output and $\mathbf{y}$ input advantageously enables the inference of τ_r even in case of a small BI compensating the drawback of the direct inference that would be only performed using $\mathbf{y}$ as input (which is highly efficient in case of a high BI).

B. Extension to Super-Resolution reconstructions

Several works have been conducted on the use of SR approaches in the SPAD context. For example, [80] combines data acquisitions from the SPAD sensor with data from a standard CMOS camera to improve the spatial resolution of the SPAD sensor. As proposed for multimodal reconstruction from SPAD data (*i.e.*, without any additional sensor), it is relevant, in our case, to take advantage of the intrinsic nature

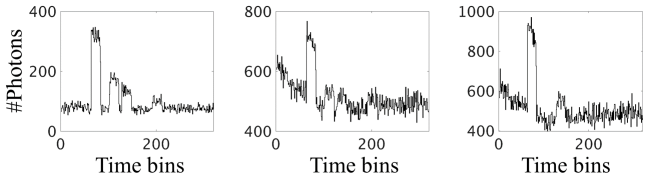


Fig. 7: Examples of ToF histogram in super resolution. The first and second histograms refer to the same objects measurements (*i.e.*, 5, 8, 10, 15 m) but with more BI for the second image. The third plot is the ToF histogram of 5, 5, 10, $10\mathrm{m}$ depth object measurements.

of the SPAD data to also perform SR at the reconstruction stage namely because SPAD imaging systems suffer from poor spatial resolution. Therefore, we decided emulate each pixel as a virtual group of pixels, thus leading to the acquisition of shared histograms. It would theoretically enable to perform a finer mutual information retrieval compared to what would happened in case of a simple averaging, this with more realistic synthetic data. However, it implies histogram peaks masking (*i.e.*, short range pixels are "shadowing" long range ones at same albedo) which makes the reconstruction more complex in practice. Fig. 7 illustrates simulated histograms where 2×2 neighbor pixels do not share same input d (with uniform τ_r). Thanks to our DGM topology preserving the overall description, to extend the DGM to SR we just added an upsampling stage using $\sigma_S = 2$ and a stride S $= 2$, as described in Fig. 6. Note that even if this SR reconstruction is not the main contribution of this paper, we still demonstrate that our proposed framework is compatible with any standard deep learning SR techniques. To improve future SR results, we can employ known deep learning based SR models surveyed in the state of the art such as image guided depth upsampling method and pixel-shuffle [81].

C. Experimental results

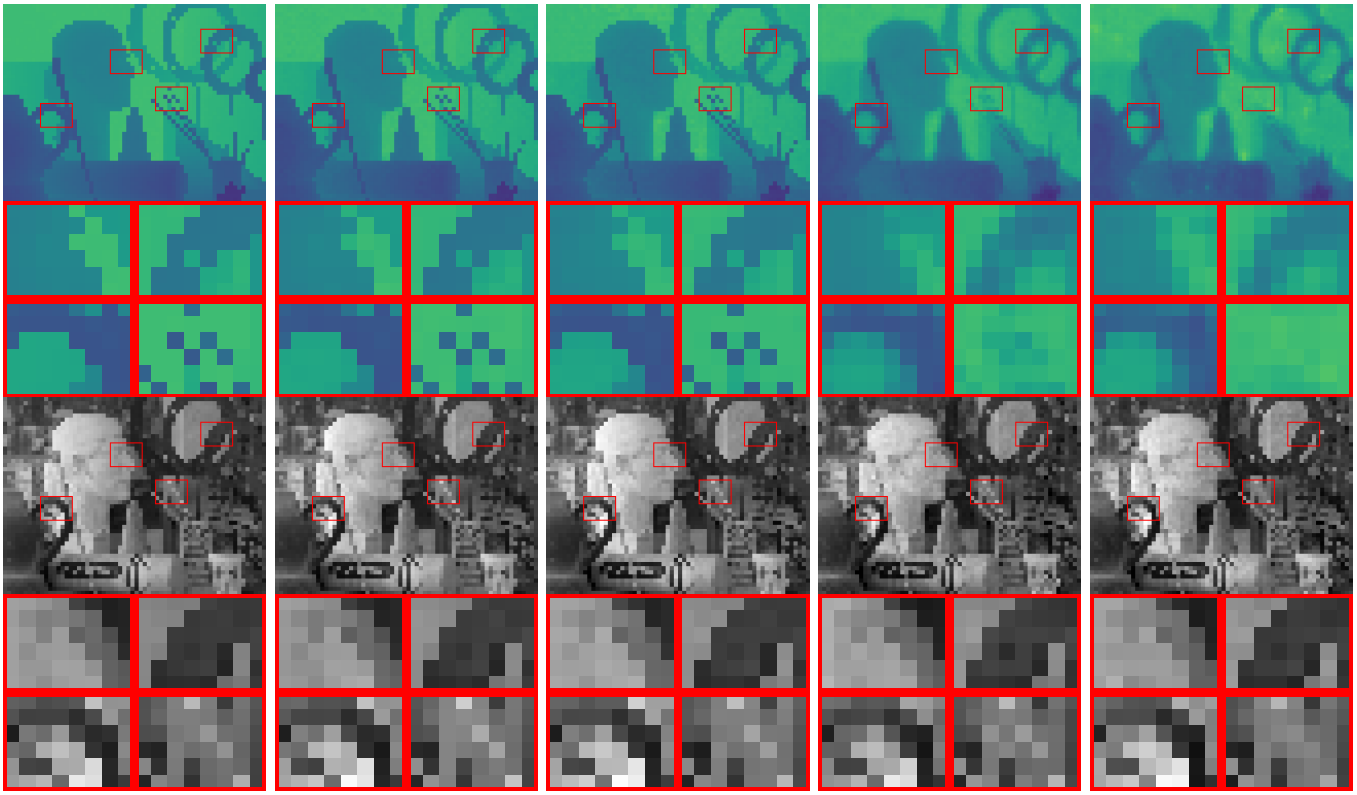
Multimodal and Super-Resolution reconstructions with NYU V2 dataset [82] as training dataset, and Middlebury dataset [83] as testing dataset are presented in Fig. 8 and in Fig. 11, 12. Because Deep Learning algorithm performances rely on the training dataset; we therefore apply data augmentation on the training and test dataset in order to limit overfitting issues. It allows us to train a model with NYU V2 dataset that would better generalize to the Middlebury test dataset. Indeed, in order to properly provide a powerful evaluation of the 12 test scenes of Middlebury reconstructions under seven different BIs (see table depicted in Fig. 10), we will take care of a relevant data augmentation. To that end, we generate three samples of every NYU V2 dataset images (*i.e.*, 1 449 $\times$ 3 RGB-D images) with different BIs randomly drawn from 50 to 1 050 $\mathrm{W/m^2}$.

As depicted in Fig. 9, two main data preprocessing stages are performed on the original training and test datasets, first related to the depth range distribution and second to the depth quantization. In order to fix the depth range distribution misalignement, we apply a generalized data augmentation, which consists in applying depth values transformation to improve measurement frequency correlation between training and test datasets over all the measurements dynamic range. Then, as depicted in Fig. 9, NYU V2 depth information is quantized in a log fashion with a limited number of bits. Therefore, a log-scaled uniform quantization noise has been added in order to avoid any unrealistic physical behavior that could limit the DGM training and testing performances by introducing reconstruction artifacts. Same correction has been applied to the Middlebury dataset but with uniform noise ranging below the quantization step. Note that Middlebury images also contains depth holes (cf. Fig. 9), we thus applied a median-diffusion algorithm [84] to fill these depth holes [85]. Since the training and test datasets must have the same scale scene, normalized depths of the training and test datasets are scaled similarly (cf., Table I) leading to a quantization step of $78\mathrm{mm}$, in order to be consistent with the simulated TDC resolution.

Finally, every images are subsampled (with border removal) to 48x64 for low resolution (LR) reconstruction, or to 96x128 for super resolution (SR) reconstruction. Models of this paper are trained using TensorFlow2 backend with a batch size of 5 frames (3 subsection IV-B) and with an amount of 40 epochs using a specific learning scheduler (lr initialized to 0.001 then rescaled each epoch by a factor of 0.85 after the first 20 epochs). Learning stages are performed thanks to an adaptive moment optimization algorithm (Adam [86]) using an output frame-based Mean Squared Error (MSE) loss function. Note that, our simulation model (*i.e.*, pseudo realistic simulation taking into account DT, albedo non-uniformity and non-narrow laser time pulses) as well as our data preprocessing (*i.e.*, CS) unfortunately does not allow for a fair and quantitative comparison with previous works of [39], [40], [42] that do not consider those considerable effects.

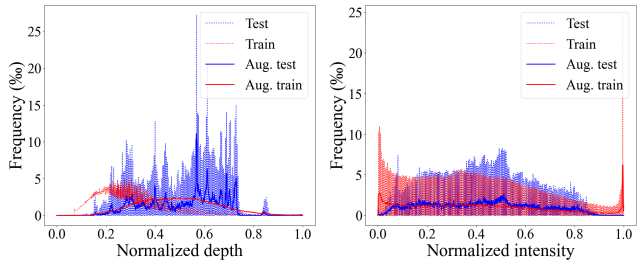
With a second MSE loss function applied on the luminance part, we obtained the average reconstruction PSNR curves reported in Fig. 10. Compared to the results for pixel-wise depth-only reconstruction (cf. Fig. 5) this multimodal collaborative reconstruction enables to further improve BI robustness, increasing the accuracy for depth estimation (cf. Fig. 10) mostly for outdoor conditions even if implying a small degradation for a low P_a. On the other hand, luminance reconstruction is optimally performed in case of indoor conditions with the light on, because both light sources (laser and BI) both participate in estimating τ_r.

Fig. 11, 12 shows examples of multimodal 2×2 SR providing 96×128 depth map and luminance images with finer spatial details from a 48×64 CS ToF histograms (M $= 64$). Reconstruction artifacts are visible (holes in circles Fig. 11 row 1 and ears details row 6) and are likely due to a combination of compression losses and background noise while long distance pixel-level details at low-resolution scale might be masked by foreground objects (with far higher contribution to low-resolved histograms).



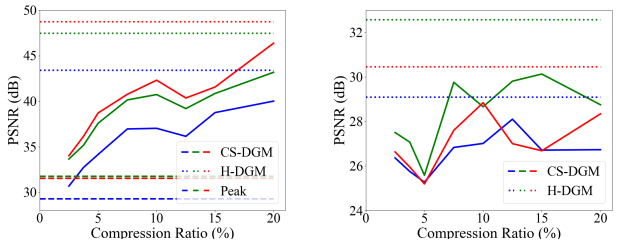
(a) GT (b) CR 20%-350W/m^2 (c) CR 20%-1050W/m^2 (d) CR 2.5%-350W/m^2 (e) CR 2.5%-1050W/m^2

Fig. 8: Depth-Luminance reconstructions from various CS inputs (*i.e.*, a CR of 2.5% for the fourth and fifth image columns, a CR of 20% for the second and third image columns) under highly enlighted indoor conditions (*i.e.*, BI of $P_a = 350\mathrm{W/m}^2$ for the second and forth image columns. BI of $P_a = 1050\mathrm{W/m}^2$ for the third and fifth image columns). First image columns are Ground truth images. Image reconstructions are issued from the DGM of Fig. 6.



(a) Depth distributions (b) Intensity distributions

Fig. 9: Depth-Intensity values distribution from images of train and test datasets (NYU V2 and Middlebury), before and after data augmentation.

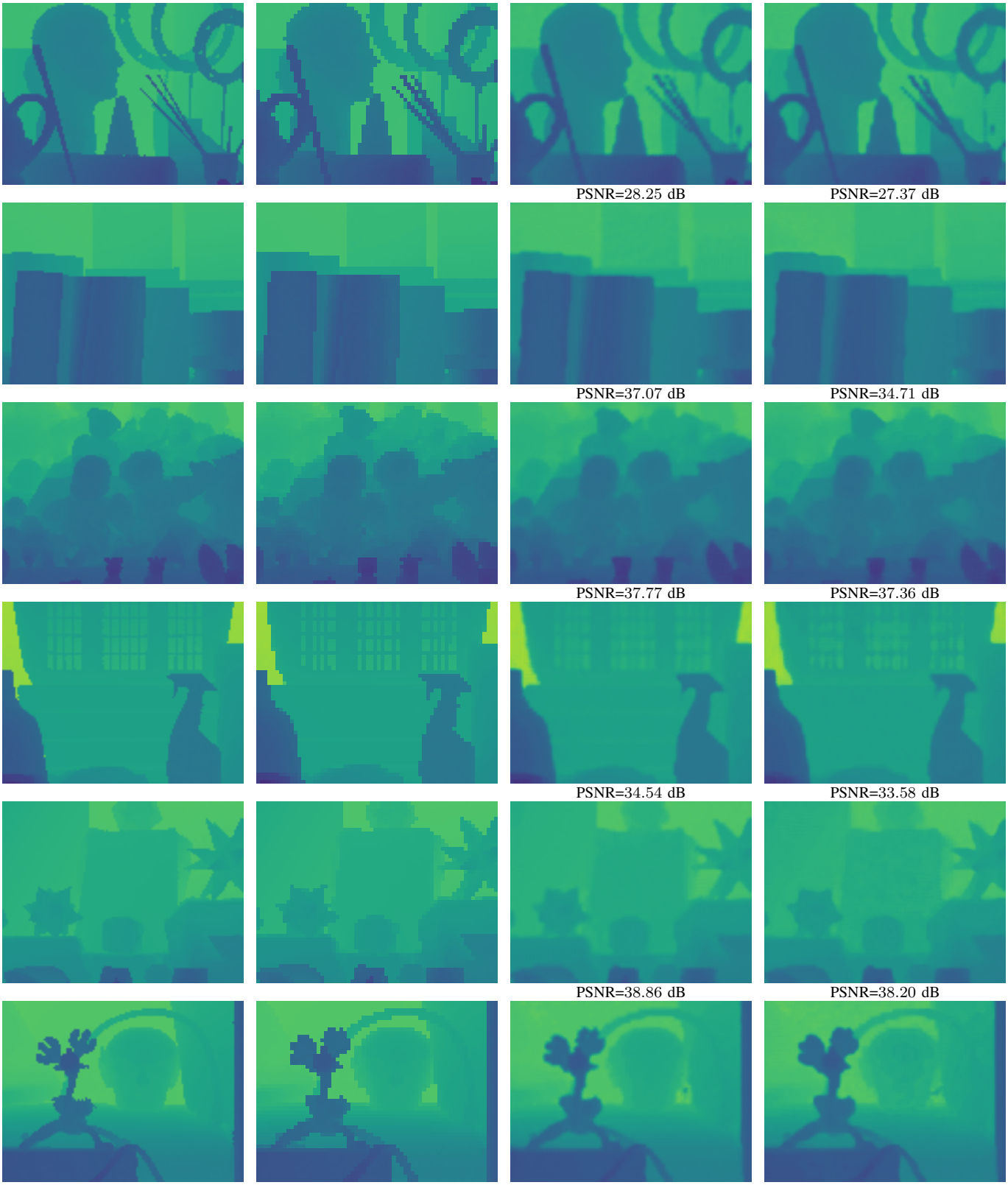


(a) Depth PSNR (b) Intensity PSNR

Fig. 10: Average depth-intensity reconstructions PSNR (over 12 images from Middlebury dataset), for intensity and depth estimation accuracy for a range of CR and three different P_a in W/m^2) (*i.e.*, indoor light-off BI (100, red curve), indoor light-on BI (350, green curve), outdoor BI (1 050, blue curve)). Horizontal dotted lines are for Histogram based reconstruction (*i.e.*, without compression, M = 320). Horizontal dashed lines are for another uncompressed reconstruction baseline (M = 320) based on a log-match-filter combined with a peak detector (*i.e.*, a learned bin-wise weighting followed by a basic peak position detector as Softargmax enabling to compensate the inherent pile-up effect).

V. Conclusion

SPAD sensor hardware characteristics are mainly limited due to the overall amount of generated data that are related to the need of managing ToF pixel-wise histograms. To tackle most of those intrinsic limitations an "in-pixel" ToF histogram CS has been proposed to facilitate future SPAD designs. This paper demonstrates that this in-pixel histogram compression can be combined with a super-resolved multimodal reconstruction thanks to a dedicated DGM. To make the reconstruction as robust as possible against background illumination variations



High Resolution
(a) GT 96 × 128

Low Resolution
(b) GT 48 × 64

PSNR=29.68 dB
(c) CR=20.0% (2 × 2 SR, M = 64)

PSNR=28.06 dB
(d) CR=5.0% (2 × 2 SR, M = 16)

Fig. 11: Depth SR reconstructions from CS inputs under highly enlighted indoor conditions ($P_a = 350\mathrm{W/m^2}$). First and second image columns are respectively groundtruth low resolved and high resolved images for depth map. While third and fourth ones are 2×2 SR reconstructions issued from the DGM of Fig. 6 at a CR of 20.0% (2×2 SR, M = 64) and 5.0% (2×2 SR, M = 16). Note that the reported PSNR are computed from the High Reolution groundtruth images.

Fig. 12: Intensity SR reconstructions from CS inputs under highly enlighted indoor conditions ($P_a = 350\mathrm{W}/\mathrm{m}^2$). First and second image columns are respectively groundtruth low resolved and high resolved images for intensity map. While third and fourth ones are 2×2 SR reconstructions issued from the DGM of Fig. 6 at a CR of 20.0% (2×2 SR, M = 64) and 5.0% (2×2 SR, M = 16). Note that the reported PSNR are computed from the High Reolution groundtruth images.

a pixel-wise/layer-wise L^∞ normalization layer is involved in the topology. Implementing depth-luminance DGM interconnections with multiple activation functions also enables a considerable improvement in reconstruction accuracy as well as making the reconstruction less sensitive to background illumination changes. Numerical results for multimodal sensing show that even for a compression using 32 CS measurements, the proposed topology enables an average PSNR depth reconstruction loss of only 7dB while its luminance counterpart is of less than 3dB, this for any background illumination levels (up to $1\,050\ \mathrm{W/m^2}$). In addition, even though being preliminary results, we demonstrated that disambiguating local ToF information collected by SPAD pixels now paves the way for even higher resolved images from D-ToF type of sensors. Future works to improve SR reconstructions will consist in including pixel-shuffle [81] with deactivable skip connections [87]. Finally, the proposed acquisition scheme may also provide input data for direct inference targeting semantic applications (*e.g.*, gesture recognition, object detection), potentially for ToF-based smart sensors.

REFERENCES

[1] A. Rochas, M. Gani, B. Furrer, P. A. Besse, R. S. Popovic, G. Ribordy, and N. Gisin, "Single photon detector fabricated in a complementary metal–oxide–semiconductor high-voltage technology," *Review of Scientific Instruments*, vol. 74, no. 7, pp. 3263–3270, Jun. 2003. [Online]. Available: https://aip.scitation.org/doi/10.1063/1.1584083

[2] A. Carimatto, S. Mandai, E. Venialgo, T. Gong, G. Borghi, D. R. Schaart, and E. Charbon, "11.4 A 67,392-SPAD PVTB-compensated multi-channel digital SiPM with 432 column-parallel 48ps 17b TDCs for endoscopic time-of-flight PET." IEEE, Feb. 2015, pp. 1–3. [Online]. Available: http://ieeexplore.ieee.org/document/7062996/

[3] L. Carrara, C. Niclass, N. Scheidegger, H. Shea, and E. Charbon, "A gamma, x-ray and high energy proton radiation-tolerant CIS for space applications," in *2009 IEEE International Solid-State Circuits Conference - Digest of Technical Papers*, Feb. 2009, pp. 40–41,41a.

[4] A. R. Ximenes, P. Padmanabhan, M. Lee, Y. Yamashita, D. N. Yaung, and E. Charbon, "A 256×256 45/65nm 3d-stacked SPAD-based direct TOF image sensor for LiDAR applications with optical polar modulation for up to 18.6db interference suppression," in *2018 IEEE International Solid - State Circuits Conference - (ISSCC)*, Feb. 2018, pp. 96–98.

[5] L. Zhang, D. Chitnis, H. Chun, S. Rajbhandari, G. Faulkner, D. O'Brien, and S. Collins, "A Comparison of APD- and SPAD-Based Receivers for Visible Light Communications," *Journal of Lightwave Technology*, vol. 36, no. 12, pp. 2435–2442, Jun. 2018.

[6] D. Bronzi, Y. Zou, F. Villa, S. Tisa, A. Tosi, and F. Zappa, "Automotive Three-Dimensional Vision Through a Single-Photon Counting SPAD Camera," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 3, pp. 782–795, Mar. 2016.

[7] K. Morimoto, A. Ardelean, M.-L. Wu, A. C. Ulku, I. M. Antolovic, C. Bruschini, and E. Charbon, "Megapixel time-gated SPAD image sensor for 2d and 3d imaging applications," *Optica*, vol. 7, no. 4, pp. 346–354, Apr. 2020. [Online]. Available: https://www.osapublishing.org/optica/abstract.cfm?uri=optica-7-4-346

[8] E. Charbon, C. Bruschini, and M. Lee, "3d-Stacked CMOS SPAD Image Sensors: Technology and Applications," in *2018 25th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, Dec. 2018, pp. 1–4.

[9] S. W. Hutchings, N. Johnston, I. Gyongy, T. Al Abbas, N. A. W. Dutton, M. Tyler, S. Chan, J. Leach, and R. 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*, vol. 54, no. 11, pp. 2947–2956, Nov. 2019.

[10] F. Arvani and T. C. Carusone, "Direct Time-of-Flight TCSPC Analytical Modeling Including Dead-Time Effects," in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2018, pp. 1–4.

[11] M. Beer, O. Schrey, B. J. Hosticka, and R. Kokozinski, "Dead time effects in the indirect time-of-flight measurement with SPADs," in *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2017, pp. 1–4.

[12] X. Chen, L. Cai, Y. Xu, J. Ou, and J. Liao, "A 180nm CMOS fully digital Indirect Time-of-Flight SOC based on SPAD," in *2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC)*, Jun. 2020, pp. 523–526.

[13] C. Zhang, S. Lindner, I. M. Antolović, J. M. Pavia, M. Wolf, and E. Charbon, "A 30-fps, 252 × 144 SPAD Flash LiDAR With 1728 Dual-Clock 48.8-ps TDCs, and Pixel-Wise Integrated Histogramming," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 4, pp. 1137–1151, Apr. 2019.

[14] M. Perenzoni, N. Massari, L. Gasparini, M. M. Garcia, D. Perenzoni, and D. Stoppa, "A fast 50 × 40-pixels single-point dtof spad sensor with photon counting and programmable roi tdcs," *IEEE Solid-State Circuits Letters*, vol. 3, pp. 86–89, 2020.

[15] I. Vornicu, A. Darie, R. Carmona-Galán, and Á. Rodríguez-Vázquez, "Compact Real-Time Inter-Frame Histogram Builder for 15-Bits High-Speed ToF-Imagers Based on Single-Photon Detection," *IEEE Sensors Journal*, vol. 19, no. 6, pp. 2181–2190, Mar. 2019.

[16] S. Cova, M. Ghioni, F. Zappa, A. Gulinatti, I. Rech, and A. Tosi, "Single photon counting detectors in action: Retrospect and prospect," in *2010 23rd Annual Meeting of the IEEE Photonics Society*, Nov. 2010, pp. 177–178.

[17] "High Resolution Photon Counting With MCP-Timepix Quad Parallel Readout Operating at - Frame Rates - IEEE Journals & Magazine." [Online]. Available: https://ieeexplore.ieee.org/document/6375792

[18] E. Charbon, "Single Photon Imaging in CMOS," in *LEOS 2006 - 19th Annual Meeting of the IEEE Lasers and Electro-Optics Society*, Oct. 2006, pp. 721–722.

[19] S. Jahromi, J. Jansson, P. Keränen, and J. Kostamovaara, "A 32 × 128 SPAD-257 TDC Receiver IC for Pulsed TOF Solid-State 3-D Imaging," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 7, pp. 1960–1970, Jul. 2020.

[20] F. Arvani, T. C. Carusone, and E. S. Rogers, "TDC Sharing in SPAD-Based Direct Time-of-Flight 3d Imaging Applications," in *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2019, pp. 1–5.

[21] P. Padmanabhan, C. Zhang, and E. Charbon, "Modeling and Analysis of a Direct Time-of-Flight Sensor Architecture for LiDAR Applications," *Sensors*, vol. 19, no. 24, p. 5464, Jan. 2019. [Online]. Available: https://www.mdpi.com/1424-8220/19/24/5464

[22] A. Gupta, A. Ingle, and M. Gupta, "Asynchronous Single-Photon 3d Imaging," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2019, pp. 7908–7917.

[23] J. Arlt, D. Tyndall, B. R. Rae, D. D. Li, J. A. Richardson, and R. K. Henderson, "A study of pile-up in integrated time-correlated single photon counting systems," Oct. 2013.

[24] H. Ruokamo, "Time-gating technique for a single-photon detection-based solid-state time-of-flight 3d range imager," 2019.

[25] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, Mar. 2008.

[26] A. Colaço, A. Kirmani, G. A. Howland, J. C. Howell, and V. K. Goyal, "Compressive depth map acquisition using a single photon-counting detector: Parametric signal processing meets sparsity," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2012, pp. 96–102, iSSN: 1063-6919.

[27] G. A. Howland, D. J. Lum, M. R. Ware, and J. C. Howell, "Photon counting compressive depth mapping," *Optics Express*, vol. 21, no. 20, pp. 23 822–23 837, Oct. 2013. [Online]. Available: https://www.osapublishing.org/oe/abstract.cfm?uri=oe-21-20-23822

[28] Q. Sun, X. Dun, Y. Peng, and W. Heidrich, "Depth and Transient Imaging with Compressive SPAD Array Cameras," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 273–282.

[29] A. Farina, A. Farina, A. Candeo, A. D. Mora, A. Bassi, A. Bassi, R. Lussana, F. Villa, G. Valentini, G. Valentini, S. Arridge, C. D'Andrea, and C. D'Andrea, "Novel time-resolved camera based on compressed sensing," *Optics Express*, vol. 27, no. 22, pp. 31 889–31 899, Oct. 2019. [Online]. Available: https://www.osapublishing.org/oe/abstract.cfm?uri=oe-27-22-31889

[30] H. Seo and J. Choi, "Histogram-based mixed-signal time-to-digital-converter array for direct time-of-flight depth sensors," *Electronics Letters*, vol. 55, no. 6, pp. 310–312, Mar. 2019. [Online]. Available: https://digital-library.theiet.org/content/journals/10.1049/el.2018.7914

[31] M. Georgiev, R. Bregović, and A. Gotchev, "Fixed-Pattern Noise Modeling and Removal in Time-of-Flight Sensing," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 4, pp. 808–820, Apr. 2016.

[32] D. Shin, A. Kirmani, V. K. Goyal, and J. H. Shapiro, "Photon-Efficient Computational 3-D and Reflectivity Imaging With Single-Photon Detectors," *IEEE Transactions on Computational Imaging*, vol. 1, no. 2, pp. 112–125, Jun. 2015.

[33] J. Tachella, Y. Altmann, X. Ren, A. McCarthy, G. S. Buller, J.-Y. Tourneret, and S. McLaughlin, "Bayesian 3d Reconstruction of Complex Scenes from Single-Photon Lidar Data," *SIAM Journal on Imaging Sciences*, vol. 12, no. 1, pp. 521–550, Jan. 2019, arXiv: 1810.11633. [Online]. Available: http://arxiv.org/abs/1810.11633

[34] J. Rapp, J. Rapp, Y. Ma, R. M. A. Dawson, and V. K. Goyal, "High-flux single-photon lidar," *Optica*, vol. 8, no. 1, pp. 30–39, Jan. 2021. [Online]. Available: https://www.osapublishing.org/optica/abstract.cfm?uri=optica-8-1-30

[35] J. Rapp, Y. Ma, R. M. A. Dawson, and V. K. Goyal, "Dead Time Compensation for High-Flux Ranging," *IEEE Transactions on Signal Processing*, vol. 67, no. 13, pp. 3471–3486, Jul. 2019.

[36] A. Kirmani, D. Venkatraman, D. Shin, A. Colaço, F. N. C. Wong, J. H. Shapiro, and V. K. Goyal, "First-Photon Imaging," *Science*, vol. 343, no. 6166, pp. 58–61, Jan. 2014. [Online]. Available: https://science.sciencemag.org/content/343/6166/58

[37] M. P. Sheehan, J. Tachella, and M. E. Davies, "A Sketching Framework for Reduced Data Transfer in Photon Counting Lidar," *arXiv.org*, Feb. 2021. [Online]. Available: https://arxiv.org/abs/2102.08732v1

[38] N. Keriven, A. Bourrier, R. Gribonval, and P. Pérez, "Sketching for large-scale learning of mixture models," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 6190–6194.

[39] J. Peng, Z. Xiong, X. Huang, Z.-P. Li, D. Liu, and F. Xu, "Photon-Efficient 3d Imaging with A Non-Local Neural Network," Aug. 2020.

[40] S. Su, F. Heide, G. Wetzstein, and W. Heidrich, "Deep End-to-End Time-of-Flight Imaging," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 6383–6392.

[41] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep Learning for Image Super-resolution: A Survey," *arXiv:1902.06068 [cs]*, Feb. 2019, arXiv: 1902.06068. [Online]. Available: http://arxiv.org/abs/1902.06068

[42] L. B, O'TooleMatthew, and WetzsteinGordon, "Single-photon 3d imaging with deep sensor fusion," *ACM Transactions on Graphics (TOG)*, Jul. 2018.

[43] T.-W. Hui, C. C. Loy, and X. Tang, "Depth Map Super-Resolution by Deep Multi-Scale Guidance," in *Computer Vision – ECCV 2016*, ser. Lecture Notes in Computer Science. Springer, Cham, Oct. 2016, pp. 353–369.

[44] D. Ferstl, C. Reinbacher, R. Ranftl, M. Ruether, and H. Bischof, "Image Guided Depth Upsampling Using Anisotropic Total Generalized Variation," in *2013 IEEE International Conference on Computer Vision*, Dec. 2013, pp. 993–1000.

[45] A. Ruget, S. McLaughlin, R. K. Henderson, I. Gyongy, A. Halimi, and J. Leach, "Robust super-resolution depth imaging via a multi-feature fusion deep network," *arXiv.org*, Nov. 2020. [Online]. Available: https://arxiv.org/abs/2011.11444v2

[46] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006. [Online]. Available: http://ieeexplore.ieee.org/document/1614066/

[47] A. Bora, A. Jalal, E. Price, and A. G. Dimakis, "Compressed Sensing using Generative Models," Mar. 2017. [Online]. Available: https://arxiv.org/abs/1703.03208v1

[48] A. Tontini, L. Gasparini, and M. Perenzoni, "Numerical Model of SPAD-Based Direct Time-of-Flight Flash LIDAR CMOS Image Sensors," *Sensors*, vol. 20, no. 18, p. 5203, Jan. 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/18/5203

[49] T. Yoshida, V. Golyanik, O. Wasenmüller, and D. Stricker, "Improving Time-of-Flight Sensor for Specular Surfaces with Shape from Polarization," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, Oct. 2018, pp. 1558–1562.

[50] H. Mahmoudi, M. Hofbauer, B. Steindl, K. Schneider-Hornstein, and H. Zimmermann, "Statistical Study of Intrinsic Parasitics in an SPAD-Based Integrated Fiber Optical Receiver," *IEEE Transactions on Electron Devices*, vol. 66, no. 1, pp. 497–504, Jan. 2019.

[51] X. Jiang, M. Itzler, K. O'Donnell, M. Entwistle, M. Owens, K. Slomkowski, and S. Rangwala, "InP-Based Single-Photon Detectors and Geiger-Mode APD Arrays for Quantum Communications Applications," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 21, no. 3, pp. 5–16, May 2015.

[52] F. Wang, W. Chen, Y. Li, D. He, C. Wang, Y. Han, S. Wang, Z. Yin, and Z. Han, "Non-Markovian Property of Afterpulsing Effect in Single-Photon Avalanche Detector," *Journal of Lightwave Technology*, vol. 34, no. 15, pp. 3610–3615, Aug. 2016.

[53] A. Ingargiola, M. Segal, A. Gulinatti, I. Rech, I. Labanca, P. Maccagnani, M. Ghioni, S. Weiss, and X. Michalet, "Optical crosstalk in SPAD arrays for high-throughput single-molecule fluorescence spectroscopy," *Nuclear instruments & methods in physics research. Section A, Accelerators, spectrometers, detectors and associated equipment*, vol. 9, no. 12, pp. 255–258, Dec. 2018. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6586236/

[54] O. semiconductor, "Sipms in direct tof ranging applications," *White paper*, no. TND6254/D, Jun. 2020. [Online]. Available: https://www.onsemi.com/pub/Collateral/TND6254-D.PDF

[55] I. Vornicu, R. Carmona-Galán, and Á. Rodríguez-Vázquez, "Real-Time Inter-Frame Histogram Builder for SPAD Image Sensors," *IEEE Sensors Journal*, vol. 18, no. 4, pp. 1576–1584, Feb. 2018.

[56] I. Vornicu, A. Darie, R. Carmona-Galán, and Á. Rodríguez-Vázquez, "ToF Estimation Based on Compressed Real-Time Histogram Builder for SPAD Image Sensors," in *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2019, pp. 1–4.

[57] E. J. Candes and M. B. Wakin, "An Introduction To Compressive Sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[58] W. Zhao, C. Park, I. Park, N. Sun, and Y. Chae, "An Always-on 4× Compressive VGA CMOS Imager with 51pj/Pixel and >32db PSNR," in *2020 IEEE Symposium on VLSI Circuits*, Jun. 2020, pp. 1–2.

[59] F. Mochizuki, K. Kagawa, S. Okihara, M. Seo, B. Zhang, T. Takasawa, K. Yasutomi, and S. Kawahito, "6.4 Single-shot 200mfps 5×3-aperture compressive CMOS imager," in *2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers*, Feb. 2015, pp. 1–3.

[60] W. Benjilali, W. Guicquero, L. Jacques, and G. Sicard, "Hardware-Compliant Compressive Image Sensor Architecture Based on Random Modulations and Permutations for Embedded Inference," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 4, pp. 1218–1231, Apr. 2020.

[61] M. F. Duarte and Y. C. Eldar, "Structured Compressed Sensing: From Theory to Applications," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4053–4085, Sep. 2011.

[62] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-Based Compressive Sensing," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.

[63] L. Jacques and P. Vandergheynst, "Compressed Sensing: "When Sparsity Meets Sampling"," *Optical and Digital Image Processing*, pp. 507–527, Apr. 2011. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/9783527635245.ch23

[64] E. J. Candes and T. Tao, "Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006. [Online]. Available: http://ieeexplore.ieee.org/document/4016283/

[65] J. Haupt and R. Nowak, "Signal Reconstruction From Noisy Random Projections," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 4036–4048, Sep. 2006.

[66] G. Yu and G. Sapiro, "Statistical Compressed Sensing of Gaussian Mixture Models," *IEEE Transactions on Signal Processing*, vol. 59, no. 12, pp. 5842–5858, Dec. 2011.

[67] A. Mousavi, A. B. Patel, and R. G. Baraniuk, "A deep learning approach to structured signal recovery," in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sep. 2015, pp. 1336–1343.

[68] A. Beck and M. Teboulle, "A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, Jan. 2009. [Online]. Available: https://epubs.siam.org/doi/abs/10.1137/080716542

[69] S. Ravishankar, J. C. Ye, and J. A. Fessler, "Image Reconstruction: From Sparsity to Data-adaptive Methods and Machine Learning," *arXiv.org*, Apr. 2019. [Online]. Available: https://arxiv.org/abs/1904.02816v3

[70] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang, "A Survey of Sparse Representation: Algorithms and Applications," *IEEE Access*, vol. 3, pp. 490–530, 2015.

[71] T. Blumensath, "Compressed Sensing With Nonlinear Observations and Related Nonlinear Optimization Problems," *IEEE Transactions on Information Theory*, vol. 59, no. 6, pp. 3466–3474, Jun. 2013.

[72] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep Convolutional Neural Network for Inverse Problems in Imaging," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4509–4522, Sep. 2017.

[73] A. Halimi, Y. Altmann, A. McCarthy, X. Ren, R. Tobin, G. S. Buller, and S. McLaughlin, "Restoration of intensity and depth images constructed

using sparse single-photon data," in *2016 24th European Signal Processing Conference (EUSIPCO)*, Aug. 2016, pp. 86–90.

[74] W. Aloisi and R. Mita, "Gated-Clock Design of Linear-Feedback Shift Registers," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 55, no. 6, pp. 546–550, Jun. 2008.

[75] D. Jun, L. Na, G. Yixiong, and Y. Jun, "A High-performance Pseudo-random Number Generator Based on FPGA," in *2009 International Conference on Wireless Networks and Information Systems*, Dec. 2009, pp. 290–293.

[76] L. Kotoulas, D. Tsarouchis, G. C. Sirakoulis, and I. Andreadis, "1-d cellular automaton for pseudorandom number generation and its reconfigurable hardware implementation," in *2006 IEEE International Symposium on Circuits and Systems*, May 2006, pp. 4 pp.–.

[77] J. Rapp and V. K. Goyal, "A Few Photons Among Many: Unmixing Signal and Noise for Photon-Efficient Active Imaging," *IEEE Transactions on Computational Imaging*, vol. 3, no. 3, pp. 445–459, Sep. 2017, arXiv: 1609.07407. [Online]. Available: http://arxiv.org/abs/1609.07407

[78] Z. Sun, D. B. Lindell, O. Solgaard, and G. Wetzstein, "SPADnet: deep RGB-SPAD sensor fusion assisted by monocular depth estimation," *Optics Express*, vol. 28, no. 10, pp. 14 948–14 962, May 2020. [Online]. Available: https://www.osapublishing.org/oe/abstract.cfm?uri=oe-28-10-14948

[79] T. Okino, S. Yamada, Y. Sakata, S. Kasuga, M. Takemoto, Y. Nose, H. Koshida, M. Tamaru, Y. Sugiura, S. Saito, S. Koyama, M. Mori, Y. Hirose, M. Sawada, A. Odagawa, and T. Tanaka, "5.2 A 1200×900 6µm 450fps Geiger-Mode Vertical Avalanche Photodiodes CMOS Image Sensor for a 250m Time-of-Flight Ranging System Using Direct-Indirect-Mixed Frame Synthesis with Configurable-Depth-Resolution Down to 10cm," in *2020 IEEE International Solid- State Circuits Conference - (ISSCC)*, Feb. 2020, pp. 96–98.

[80] C. Callenberg, A. Lyons, D. d. Brok, A. Fatima, A. Turpin, V. Zickus, L. Machesky, J. Whitelaw, D. Faccio, and M. B. Hullin, "Super-resolution time-resolved imaging using computational sensor fusion," *Scientific Reports*, vol. 11, Jan. 2021. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7813875/

[81] C. Du, H. Zewei, S. Anshun, Y. Jiangxin, C. Yanlong, C. Yanpeng, T. Siliang, and M. Y. Yang, "Orientation-Aware Deep Neural Network for Real Image Super-Resolution," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2019, pp. 1944–1953.

[82] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor Segmentation and Support Inference from RGBD Images," in *Computer Vision – ECCV 2012*, ser. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, Oct. 2012, pp. 746–760.

[83] H. Hirschmuller and D. Scharstein, "Evaluation of Cost Functions for Stereo Matching," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2007, pp. 1–8.

[84] R. L. Biradar and V. V. Kohir, "A novel image inpainting technique based on median diffusion," *Sadhana*, vol. 38, no. 4, pp. 621–644, Aug. 2013.

[85] N. Yang, Y. Kim, and R. Park, "Depth hole filling using the depth distribution of neighboring regions of depth holes in the Kinect sensor," in *2012 IEEE International Conference on Signal Processing, Communication and Computing (ICSPCC 2012)*, Aug. 2012, pp. 658–661.

[86] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv.org*, Dec. 2014. [Online]. Available: https://arxiv.org/abs/1412.6980v9

[87] V. F. Abrevaya, A. Boukhayma, P. H. S. Torr, and E. Boyer, "Cross-Modal Deep Face Normals With Deactivable Skip Connections," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2020, pp. 4978–4988.