From algorithm to hardware co-design for multimodal information retrieval of TCSPC data

Valentin Poisson

Communauté
UNIVERSITÉ Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Sciences pour l'Ingénieur**

Arrêtée ministériel : 25 mai 2016

Présentée par

Valentin POISSON

Thèse dirigée par **Gilles SICARD, PhD, HDR**

préparée au sein du **Laboratoire circuits Intégrés Intelligents pour l'Image (L3I)**
dans **l'École doctorale Électronique, Électrotechnique, Automatique, Traitement du Signal (ED EEATS)**

From algorithm to hardware co-design for multimodal information retrieval of TCSPC data

Algorithme et co-conception matérielle pour l'analyse multimodale de données TCSPC

Thèse soutenue publiquement le **15/12/2022**,
devant le jury composé de :

Frédéric Pétrot
Professeur des universités, Grenoble INP-ENSIMAG, Président

Maxime Pelcat
Maitre de conférence HDR, INSA Rennes, Rapporteur

Jean-Yves Tourneret
Professeur des universités, ENSEEIHT Toulouse, Rapporteur

Maria Trocan
Professeure, ISEP Paris, Examinatrice

Gilles Sicard
Ingénieur HDR, Directeur de thèse, CEA-Leti Grenoble, Invité

William Guicquero
Ingénieur docteur, Encadrant de thèse, CEA-Leti Grenoble, Invité

Acknowledgements

My thanks go first to my thesis advisor, William Guicquero for his support, for being available, attentive. But above all I would like to mention that I am very proud to have benefited from his teaching, both technical and human. Thanks to Gilles Sicard, my thesis director, for the long discussions we had every morning during three years.

I thank Mrs Frédéric Pétrot, Maxime Pelcat, Jean-Yves Tourneret, Maria Trocan for having given me the honor of participating in the thesis jury.

I thank Fabrice Guellec for having welcomed me in the L3i laboratory and all the members of the laboratory.

I would also like to thank my friends who have more or less contributed by participating in the proofreading of my work or by supporting me emotionally which has been a difficult task and allowed me to get there.

Thanks to all my family who even with the distance gives me the strength and support to accomplish my wishes. I can't forget to sincerely thank my parents who are always willing to give good advice, attentive and supportive. Sorry to give you nightmares about my new sports practice in Grenoble.

These three years of thesis would not be as enjoyable and delicious without you Madeline, especially thanks to your cooking!

Abstract

Since the invention of the first photon counting sensors, several technological breakthroughs have been made in the field of photon counting imaging systems. The invention of Single Photon Avalanche Diode (SPAD) sensors particularly marked in the 2000s as these sensors are priced lower and show much better performances in terms of noise and detection efficiency compared to previous technologies. Several SPAD characteristics were subsequently improved by the scientific community, including RGB-D (red, green, blue, depth) alignment, timing resolution, dynamic range and pixel-level processing, allowing SPAD sensors to be used in a wide range of emerging applications. Among all these applications, 3D imaging through the use of direct time-of-flight (D-ToF) SPAD sensors, as opposed to indirect time-of-flight (I-ToF), will be developed in this PhD thesis.

In spite of the discovery of new achievements such as the scaling of technology nodes and 3D integration, technical hurdles to SPAD development (*i.e.*, better spatial resolution [1], greater reliability in all conditions [2, 3] and achieving a more compact form [4]), limiting its ability to grow and massively penetrate the 3D imaging market in comparison to other 3D sensing technology. Indeed, compared to competing 3D sensing technologies, spatial resolution is actually one of the main challenges in the recent development of SPAD D-ToF sensors, as well as robustness to noise in outdoor conditions. In fact, the high spatial resolution inevitably leads to output data that exceeds the capacity of existing data transfer technologies and requires too much memory area at the pixel level due to an unfortunate coupling between the spatial and temporal resolution of SPAD-based cameras. Regarding the spatial resolution issue, the scientific community agrees that the time correlated single photon counting (TCSPC) format itself remains the main bottleneck. Therefore, several state-of-the-art (SoA) works are underway, either to avoid the use of TCSPC based imaging system [5] or to implement design architecture tricks in order to circumvent the TCSPC data format problem reducing the data storage [6, 7] and the data rate [8] requirements. However, some of TCSPC data format implementation tricks have been performed at the expense of other SPAD performances such as acquisition frequency, background illumination robustness, spatial resolution and temporal resolution [7, 9]. On the other hand, regarding noisy reconstruction artifacts, SoA works introduce canonical Bayesian inference [10, 11] and deep learning (DL) [12, 13] algorithmic approaches to remotely and accurately process TCSPC data in complex SPAD operating mode (*e.g.*, in outdoor conditions). The drawbacks of these works are that it only focuses on low photon flux operating mode, neglecting other SPAD configurations with pile-up effect, hardware defaults and constraints.

After a quick overview of the 3D imaging field and a presentation of the SoA works, this thesis first studies the physical process of generating generic, full-frame raw data from a SPAD sensor. This "pseudo-realistic" data generation process was necessary due to the lack of available SPAD raw data and the inaccessibility of a SPAD sensor. Besides providing a clear understanding of the impact of the physical parameters on the ToF histograms, this study allows to identify statistical models, not clearly defined by the scientific community, that fit the TCSPC SPAD-based data measurements. In fact, the phase-type (PH) distribution mixture model is the statistical models

that accurately describes the distribution of raw data for general SPAD operating modes. However, in some SPAD operating modes, the PH model can be approximated by a mixture model of Gaussian-uniform distributions or sparse model as considered in some SoA works. These model approximations, at the first glance, appears easier to process than the standard PH distribution mixture model which will be exploited to simplify the computational process for the consideration of material constraints. Therefore, this manuscript will address several SPAD operating modes, both SoA and non-SoA, leading to various related TCSPC statistical model description and approximation.

As a result of this study, a customized Bayesian inference approach has been first developed dealing with the non-convex optimization problems of PH distributions mixture model function. The approach consists of a customized Genetic Expectation-Maximization (c-GEM) algorithm, designed on the basis of an appropriate combination of mathematical tools.

Thereupon, as the initial idea was to completely rethink the canonical TCSPC data acquisition process due to the large amount of data required, a designed hardware implementation was proposed based on the sequential computation and encoding of statistical parameters, namely, 2 stage expectation maximization (EM) algorithm. However, due to the non-convex optimization function of the phase-type distribution mixture model, the 2-stage expectation maximization (EM) algorithm was limited to the SPAD sensor operating mode which provides TCSPC data in the shape of Gaussian-uniform distributions. Under similar SPAD configurations, this solution for online peak detection with its register transfer level (RTL) hardware implementation demonstrates a reduction of the pixel pitch while reaching a superior depth reconstruction accuracy compared to previous works. In addition, compared to SoA works, CO-EM output parameters additionally allow for a direct estimation of the albedo and the Signal to Background Ratio (SBR).

The second alternative representation to the canonical TCSPC data is based on the use of the compressive sensing (CS) paradigm in the temporal domain in order to reduce the data throughput and pixel pitch of SPAD sensors. Unlike, SoA works which by studying the CS technique in the spatial domain only reduces the data rate [14], our approach consists in a pixel-wise CS in the time domain using an in-pixel Cellular Automaton. This was first developed for a specific mode of SPAD operation yielding temporally sparse TCSPC data, in combination with custom regression analysis reconstruction. And it does not suffer from loss of information in the event of a convergence problem as the CO-EM does. The proposed acquisition method allows to reach a reconstruction performance similar to most recent SPAD architecture tricks with an estimated pitch reduction of more than 25% and extends the number of peaks to be detected.

The last chapter of this thesis presents two deep learning models to fully exploit the potential of CS by reconstructing both depth and intensity maps. For this purpose, a first deep learning model based on low photon counts SPAD data has been developed to introduce the deep learning reconstruction algorithm combined with the CS TCSPC data through the use of a learned compression pattern layer, which operates a learned projection into a smaller latent representation space. Subsequently, the embedding layer is replaced by the presented earlier hardware implementation CS scheme, allowing to relax the hardware constraints on the SPAD sensor. The CS scheme firstly reduces the number of memory samples and the number of bits per sample, thus reducing the pixel pitch by using a shuffled cellular automaton (SCA). Secondly, it allows to get out of the restricted mode of operation at low photon flux and thus improving multimodal reconstructions. On the other hand, the second model developed aims to get rid of the sparsity constraint and to provide efficient multimodal and super-resolution reconstructions, thus extending the use of the CS to all the operating modes of SPAD sensors.

Résumé

Depuis l'invention des premiers capteurs de comptage de photons, plusieurs avancées technologiques ont été réalisées dans le domaine des systèmes d'imagerie à comptage de photons. Ces avancées ont été particulièrement marquées dans les années 2000 par l'invention des capteurs à avalanche à photon unique (SPAD pour Single Photon Avalanche Diode). En plus d'être intéressant en termes de coûts, ce nouveau capteur présente de bien meilleures performances en termes de bruit et d'efficacité de détection photonique. Ces premières avancées ont mené la communauté scientifique à améliorer d'autres caractéristiques telles que celle de l'alignement RVB-P (rouge, vert, bleu, profondeur), la résolution temporelle, la dynamique des mesures et l'intégration de traitement au niveau du pixel afin d'étendre l'utilisation des capteurs SPAD à un large éventail d'applications émergentes. Parmis les nouvelles applications visées, l'imagerie 3D à travers l'utilisation de capteur SPAD à temps de vol direct (D-ToF, par opposition à temps de vol indirect (I-ToF)) va être développée dans ces travaux.

Les principales limitations de ce type de capteur pour l'imagerie 3D sont d'une part la résolution spatiale, car la tendance actuelle pousse à l'utilisation des capteurs SPAD nécessitant toujours une plus haute résolution spatiale [1] et temporelle. Or, ceci conduit inévitablement à un flux de données de sortie qui dépassent la capacité des technologies de transfert de données existantes et nécessitent une trop grande surface mémoire au niveau du pixel [4]. D'autre part, l'imagerie 3D implique des modes de fonctionnements bruités (e.g., conditions de fonctionnements extérieures) par rapport à des applications antérieures telle que l'imagerie médicale, rendant les mesures moins fiables [2, 3].

Par conséquent, par rapport aux technologies de détection 3D concurrentes (imagerie stéréoscopique par exemple), l'attractivité du capteur SPAD dans ce domaine applicatif reste limitée.

Pour améliorer la résolution spatiale, la communauté scientifique s'accorde à dire que le format des données de comptage de photon corrélé dans le temps (TCSPC pour Time Correlated Single Photon Counting) est contraignant. C'est pourquoi plusieurs travaux de l'état de l'art tentent d'éviter l'utilisation de ce format de données [5], ou alors tentent de mettre en œuvre des astuces d'architecture de conception [6–8]. Cependant, la plupart des astuces d'architecture proposées par l'état de l'art ont été réalisées au détriment d'autres caractéristiques du SPAD telles que la fréquence d'acquisition, la robustesse de l'éclairage de fond et la résolution temporelle [7, 9].
D'autre part, pour traiter avec précision et fiabilité les données du capteur SPAD certains travaux de l'état de l'art introduisent des approches algorithmiques d'inférence bayésienne [10, 11] et d'apprentissage profond [12, 13]. Les inconvénients de ces travaux sont qu'ils ne portent que sur le mode de fonctionnement à faible flux photonique, et ne tiennent pas compte du phénomène physique, appelé en anglais « Pileup effect », qui correspond à l'empilement de plusieurs évènements dans le temps.

Une présentation contextuelle du vaste monde de l'imagerie 3D ainsi que des travaux de l'état de l'art permettra d'introduire ces travaux de thèse.
Du fait de l'inaccessibilité à des données SPAD brut ou un capteur SPAD, l'étude physique du

processus de génération des données a été rendue nécessaire, conduisant au dévelopement d'un outil de simulation générant des données "pseudo-réalistes". Le modèle de simulation développé rassemble de multiples caractéristiques physiques (albédo de la scène, profondeur des objets, efficacité photonique du capteur SPAD, puissance laser etc.) impactant la nature des données statistiques, donnant lieu à des modèles descriptifs plus ou moins simple à traiter du point de vue algorithmique.
Par suite, deux analyses statistiques permettront d'envisager des reconstructions de profondeur et de luminance.

La première analyse est une approche Bayésienne traitant du problème d'optimisation non-convexe des données statistiques du capteurs SPAD. La méthode développée consiste en un algorithme d'espérance-maximisation couplé avec une heuristique "génétique" (GC-EM), conçu sur la base d'une combinaison appropriée d'outils mathématiques.
Dans l'optique d'améliorer le schéma d'acquisition initial type TCSPC, une approximation de l'analyse statistique a été réalisée limitant le capteur SPAD à un mode de fonctionnement asynchrone. Ceci a permis de développer un schéma d'acquisition CO-EM basée sur le calcul et l'encodage séquentiel des paramètres statistiques. Cette solution permet de réduire le pas pixels tout en atteignant une précision de reconstruction en profondeur supérieure à celle des travaux de l'état de l'art. De plus, les paramètres de sortie du CO-EM permettent une estimation directe du rapport signal à bruit, de la profondeur et l'albédo de la scène mesurée.
Une seconde alternative d'amélioration du schéma d'acquisition des données TCSPC, est basée sur l'utilisation d'une technique d'acquisition compressive (CS pour compressive sensing) dans le domaine temporel permettant de réduire le débit ainsi que le pas pixel des capteurs SPAD. Par opposition à l'état de l'art tel que [14] qui en étudiant la technique d'acquisition compressive dans le domaine spatial ne réduit que le débit de données et contraint fortement le système optique. L'approche proposée consiste en un schéma d'acquisition compressive basé sur l'utilisation d'un automate cellulaire (CA pour cellular automaton) proche pixel combiné à une méthode de régression personnalisée pour la reconstruction. Cette méthode présente comme avantage d'être sans perte d'information contrairement au schéma d'acquisition CO-EM pouvant rencontré des problèmes de convergence et ainsi entraîner des pertes d'informations. Cependant, cette solution ne se limite qu'à un mode de fonctionnement du capteur SPAD spécifique respectant les propriétés fondamentales du CS (parcimonie et échantillonnage incohérent).

Le dernier chapitre de cette thèse présente deux modèles d'apprentissage profond permettant d'exploiter pleinement le potentiel du CS en reconstruisant à la fois des cartes de profondeur et d'intensité. Pour cela, un premier modèle d'apprentissage profond basé sur des données SPAD à faible flux photonique a été développé afin d'introduire l'acquisition comprimée des données TCSPC combiné à un algorithme de reconstruction d'apprentissage profond. Ce modèle se compare aux méthodes de l'état de l'art en intégrant une couche opérant une projection apprise dans un espace de représentation latent plus petit. Par suite, cette couche est remplacée par un schéma CS d'implémentation matérielle, permettant de relâcher en pratique les contraintes matérielles sur le capteur SPAD. Le schéma CS réduit dans un premier temps le nombre de point mémoire et le nombre de bits par point, réduisant ainsi le pas pixel. Et dans un deuxième temps, il permet de sortir du mode de fonctionnement restreint à faible flux photonique et ainsi améliorer les reconstructions multimodales. D'un autre coté, le second modèle développé tente de s'affranchir de la contrainte de parcimonie et de fournir des reconstructions multimodales et super-résolues performantes, étendant ainsi l'utilisation du CS à tous les modes de fonctionnement de SPAD.

Table of Contents

"As we transition imaging to widely available high resolution single photon cameras, we are likely in for more surprises. Light is made up of photons. Light detection is a Poisson process. Light and light intensity are derived quantities that are based on ensemble averages over a large number of photons. It is reasonable to assume that detection and processing methods that are based on the classical concept of flux are sub-optimal. The full potential of single photon capture and processing is therefore not yet known. I am hoping for more positive surprises."

Image Sensors World blog post
by Prof. Andreas Velten
from University of Wisconsin-Madison (August 24, 2022)

1 Introduction

This chapter introduces single-photon avalanche diode (SPAD) image sensors to the vast world of 3D imaging. It reviews the entire 3D imaging market, involving stereoscopic disparity map, Depth from focus, structured light, time of flight, interferometry technologies. Pros and cons of these several 3D optical imaging technologies are provided, with a particular focus on time-of-flight (ToF) imaging sensors based on photon counting systems. A brief introduction to the history of single photon counting (SPC) image sensors from the early 1900s to nowadays is presented with a description of the main characteristics of SPC sensors. Then, an invitation to the 3D imaging field is presented, with a growing interest in SPAD sensors. This one has already and will in the near future revolutionize the use of SPC sensors. However, some SPAD limitations are emerging. Therefore, we will review some interesting proposals of different implementations of the SPAD sensor developed by the scientific community to overcome some of the remaining limitations of SPAD. In conclusion, this chapter will provide the outline of this manuscript briefly detailing the concepts covered in each chapter.

Contents

1.1 ToF techniques in the 3D imaging and sensing ecosystem

1.1.1 An invitation to 3D imaging

Nowadays, 3D imaging with highly accurate depth measurements is a current requirement for many industrial and consumer applications. To address this demand, many market players offer a remarkable variety of 3D imaging sensors based on various sensing and processing techniques, subdivided into two groups *i.e.*, passive and active optical sensing techniques, as shown in Fig. 1.1.

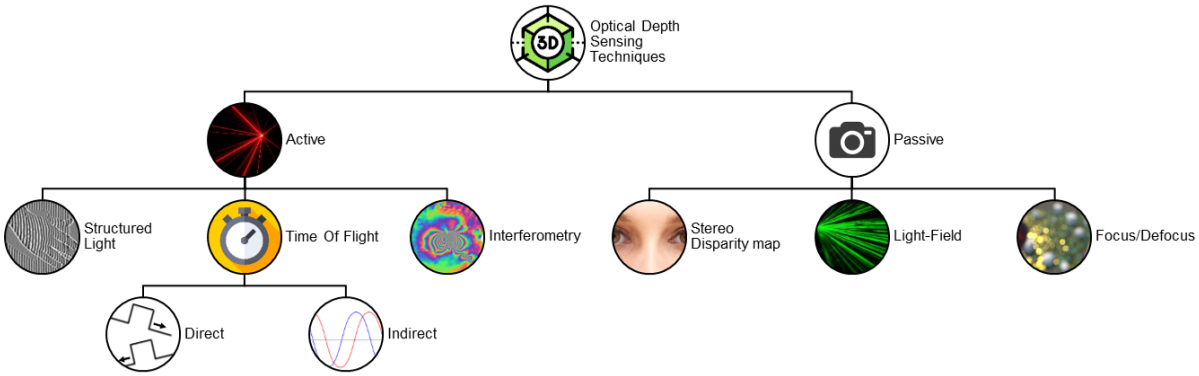


Figure 1.1: Optical Depth Sensing Techniques

Among the passive optical depth sensing techniques, numerous techniques exist with particular specifications and key features. For example, the stereo disparity map mimics the human eyes by capturing images of the same object from various cameras separated by a known and calibrated distance. Note that the stereo disparity map techniques can also be used with a single moving camera instead, knowing the transformation in 3D space between two or more successive image captures (*e.g.*, using accelerometers and/or SLAM techniques [15]). It mimics the human eyes by analyzing all captured images to interpret the displacements between key features (*i.e.*, shapes and 2D positions). Known as a feature-matching problem, the study of the disparity map between the images finally allows to obtain the depth point projection. Disregarding nonrigid effects such as specularities, and transparencies, the stereo disparity map still has to deal with occlusions as written in Tab. 1.1 due to depth discontinuities and the lack of texture in images. Always among passive optical depth sensing techniques, "depth from focus" (DfF) [16] estimates depth points of a scene from a set of two or more images from the same point of view obtained by changing the focal setting or the image plane axial position of the camera. In the same vein, "depth from defocus" (DfdF) [17] consists in measuring the amount of blurring which depends solely on the distance to the surface of exact focus and the characteristics of the lens system on a single image. By measuring the blurring at a given point in the image, it is possible to compute the distance to the corresponding point in the scene. However, DfdF suffers from ambiguity in depth estimation due to the camera depth of field, and needs a scene model and an explicit calibration between blur level and depth value as written in Tab. 1.1. Based on focus techniques, phase detection auto-focusing (PDAF) [18] is also a camera with photodiodes that are physically masked such that light from only one side of the lens reaches the photodiode. This allows providing two images on a single image sensor with phase difference (clearly define on [18]) which determine the focus point. PDAF is a great system for capturing fast objects but is a complex process and needs to be aligned properly cf., Tab. 1.1. Hence, its use only for the estimation of the focal plane. Otherwise, there also exists the light field (LF) which consists in measuring the intensity of light and also precising the direction of each ray from the scene given by the five-dimensional plenoptic function [19]. This technique allows to measure fast moving objects that exceed depth from focus

capabilities [20]. While LF technique is restricted in terms of spatial resolution (*i.e.*, trade-off between spatial and temporal resolution) and is also subject to optical pixel vignetting effects cf., Tab. 1.1.

On the other hand, active optical depth sensing techniques are generally used for accurate depth estimation at the expense of an illumination device. A wide variety of active optical depth sensing techniques were investigated differing in their spectral band or operating principle. For example, structured light (SL) [21] consists of projecting a known infrared pattern –generally irregular grids– onto a scene and analyzing its deformation to calculate the 3D surface information of the scene. Among active imaging methods there also exists, interferometry extracting depth information from the interference of superimposed waves or ToF imaging. Optical interferometry (OI) [22] provides sensitive metrology capabilities but is usually used in a controlled system environment.

Considering this vast 3D imaging ecosystem (see [23, 24] for more details), this manuscript focused on the preferred choices for measurements at longer ranges over the last decades, namely ToF imaging techniques, using light pulses or continuous waves. Indeed, the use of ToF imaging techniques has been generalized for many applications in recent years [25] and seems to have many advantages for 3D imaging, such as its high depth accuracy, its wide dynamic range measurements, its low cost, its robustness in low light conditions, its integration in CMOS technology and processing capabilities. This family of active optical depth sensors, also known as Light Detection And Ranging (LiDAR), consists in estimating depth by illuminating an object with a laser beam and then detecting the reflected light with a photodetector.

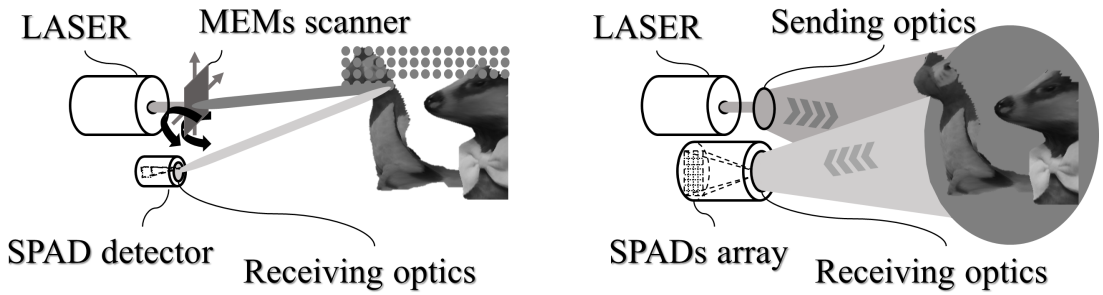


Figure 1.2: Scanner camera (left) and Flash camera (right)

Two antagonist setups can be developed: a scanner or flash camera as illustrated in Fig. 1.2. Raster scanning LiDAR is combined with a scanning system, which is generally composed of a rotating mirror. This allows when using a single axis rotating mirror and employing several vertical laser beams to provide a 360° point-cloud. Alternatively, there exists flash ToF LiDAR by employing an optical system that diverges the light of a single emitter such that the entire scene of interest is illuminated and the reflected light goes to a two-dimensional array of photodetectors, namely a ToF depth camera. The main difference between a flash LiDAR camera and a scanning LiDAR device is that there is no need of a mechanical scanning mechanism [26], *e.g.*, a rotating mirror [27] or a beam steering [28]. A significant advantage to do without scanning is that each photodetector in the direct time-of-flight (D-ToF) 3D flash cameras is exposed to the optical signal for a long period with time measurement consistency. In doing so, the various noise effects are reduced in two ways, by eliminating mechanical misalignments and non-linearity effects in order to acquire a large number of laser cycles. Beyond the fact that two possible setups have been developed, there are above all three operating principles of ToF sensors, indirect continuous-wave time-of-flight (I-ToF), Frequency-Modulated Continuous Wave (FMCW) and D-ToF.

1.1.2 Principle description of Frequency-Modulated Continuous Wave, Indirect and Direct Time-of-Flight LiDAR

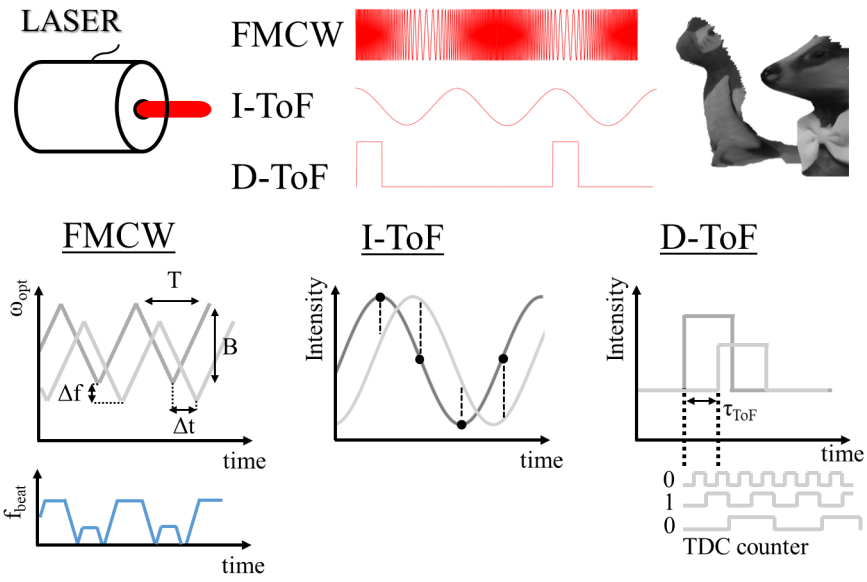


Figure 1.3: Operating principle of a FMCW, I-ToF and D-ToF lidars

FMCW sensors An emerging technology, namely FMCW LiDAR for frequency-modulated continuous wave LiDAR [29] can measure distance by repeatedly linearly chirping frequency of the continuous laser beam as illustrated in Fig. 1.3. Once the measured signal is returned to the receiver, a coherent detector mixes the measured signal with the local oscillator and measures the frequency shift, Δf. These frequency differences allow estimating the distance d from the prior knowledge of the chirp speeds, *i.e.*, $d = \frac{c|\Delta t|}{2} = \frac{c|\Delta f|}{2\frac{d(f)}{d(t)}}$. Where c refers to the speed of light, Δt to the measured time shift, Δf to the measured frequency shift and $\frac{d(f)}{d(t)}$ to the frequency shift per unit of time. Further processing extracts the Doppler shift to estimate the object's velocity relatively to the sensor. The main disadvantages of this technology concern the alignment of the corresponding RGB-D images and its processing cost as written in Tab. 1.1.

I-ToF sensors Alternatively, the continuous-wave I-ToF consists in emitting an amplitude-modulated light signal onto a scene and then receiving it. Once the signal is received, a phase-shift due to the round trip of the light signal can be measured in order to estimate depth. Each pixel cell independently performs demodulation of the received signal by digitally counting incoming photons integrated during specific time gating windows, given by reference clock(s) synchronized with the emitted signal. This operation process, namely time-gated single photon counting (TGSPC), can be performed sequentially or simultaneously with multiple reference phase-shifted clocks in case of multiple counters. In doing so, it can measure its phase delay as well as its amplitude and offset (background illumination). Note, that the received signal is attenuated because of the object's albedo, distance and background illumination. I-ToF Multi-tap is one of the I-ToF implementations using amplitude modulation and multi-tap pixels [30]. This technology is powerful and cost-effective but suffers from a limited temporal resolution.

D-ToF sensors Otherwise, pulsed-light D-ToF sensors are composed of a laser source emitter firing out ultra-short pulses of light that once reflected onto an object can be detected by the combined

receiver which is an array of photodiodes with time-to-digital converters (TDCs). The D-ToF measurement thus consists in measuring the round trip time of the laser light pulse, τ_{ToF} and deduces the measured distance d, *i.e.*, $d = \frac{c\tau_{ToF}}{2}$ (cf., Fig. 1.3). Different technologies can exploit D-ToF imaging with pros and cons such as D-ToF MCT-APD for direct time-of-flight mercury-cadmium-telluride avalanche photodiode [31] and D-ToF SPAD for Single Photon Avalanche Diode [32]. Because of the physical constraints imposed by the measurement technique and the early stage development of advanced IC technology, D-ToF sensors have been highly limited up to now in terms of spatial resolution, power consumption and processing cost cf., Tab. 1.1.

	SL	D-ToF MCT-APD	D-ToF SPAD	I-ToF Multi-tap	Stereo	PDAF	DfF	DfdF	FMCW	OI
Active imaging	yes	yes	yes	yes	no	no	no	no	yes	yes
Spatial Res.	+	+	−	+	+	+	−	−	n/a	− − −
Max range	−	+ + +	+ +	+	−	−	− −	− −	+	−
Depth Res.	+ +	+	+ +	+	−	−	− −	− −	+ +	+ + +
FPS	+	+ +	+ +	+	+	+	− −	+	−	+
Optics	− −	+	+ +	+	+ +	+ +	+ +	+ +	+ +	− −
System size	−	−	+ +	+	− −	+ + +	+	+	+ + +	− −
BOM (price)	+ +	− −	+	+ +	−	+ +	+ +	+ + +	−	− −
RGB-D alignment	+	−	+ +	+ +	−	+	+	+	− −	− −
Processing	−	+ +	+	+	−	−	−	− −	+	− −
System pwr.	−	− −	−	−	+	+ +	+	+	−	−
Applications	smartphone/ metrology	military/ spatial	smartphone/ robotics	smartphones/ tiny robotics		moving object/ focus/			automotive robotics	biological semiconductor
Advantages	spatial resolution	dynamic range	temporal resolution	cost-effective technology	Optical	acquisition frequency			speed measurements	depth accuracy
Disadvantages	occlusions/ specularity	noise level	spatial resolution	power consumption	depth of field	depth resolution			reliability	vignetting

Tableau 1.1: Qualitative, subjective comparison of light-based depth sensing systems.

1.2 Evolution of single photon photodetectors

ToF has been a research topic for several years, during which researchers have developed different types of single photon ToF sensors for a wide range of applications, based on various technologies or physical mechanisms. Here, it is detailed the evolution and the state of the art of the main existing single photon detectors and their advantages and disadvantages.

1.2.1 Photomultipliers: PMTs

The Photomultiplier tube (PMT) was invented in the early 1900s and was until 1980, the only detector of very low intensity photon flux on the market [33]. The device consisted of a source of primary photoelectrons (photocathode), an electron focusing (optics or magnets) and a multistage of secondary electron emitters (dynodes) with a certain coefficient (cf., Fig. 1.4). Each subsequent dynode is supplied with higher electrical potential in comparison with preceding dynode from high voltage power supply. In doing so, the photoelectrons produced by the illumination of the photocathode are accelerated to the first dynode by tearing off more numerous but less energetic secondary electrons from its surface, also accelerated to the next emitter by tearing off more numerous secondary electrons again and so on until they reach the anode. The arrival of these electrons on the anode creates a brief current pulse, which is the detection of a photon.

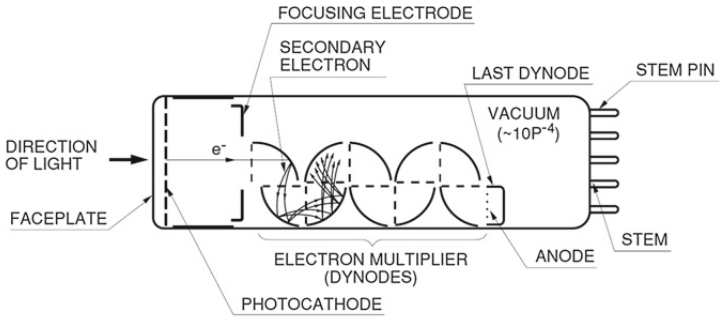


Figure 1.4: Typical photomultiplier tube (PMT) from [34]

The main characteristics of PMT are manifold. Firstly, PMT has a very large sensing area. Secondly, it also has a very high gain in creating a huge number of electrons per incident photon ($\sim 10^6$) with a very low counting rate in the dark thanks to its low noise properties. Finally, its ultra-fast transit time (*i.e.*, the time between the illumination and the moment of the pulse on the anode) reflects an excellent timing resolution. However, the photomultiplier suffers from several drawbacks, mainly a high supply voltage that goes from a few hundred volts to several thousand volts. In addition, the PMT is a heavy detector and very sensitive to the magnetic field.

1.2.2 Electron Multiplying Charged Coupled Devices: EMCCDs

The Electron Multiplying Charged Coupled Device (EMCCD) was launched on the market in the early 1990s by e2v [35]. These sensors were an important innovation in imaging that effectively reduced the relative noise of the readout circuit, by the amplification of the electrons captured from the signal before it passed to the reading phase. This is performed thanks to the use of two additional reading registers compared to the conventional CCD technology, a conventional register and the register dedicated to the multiplication of electrons (cf., Fig. 1.5). High voltages are applied to the additional register so that the electrons acquire sufficient energy for additional electrons to be released by the ionization impact process. This gain step allows multiplication of the signal without adding extra noise with factors of the order of 10^3.

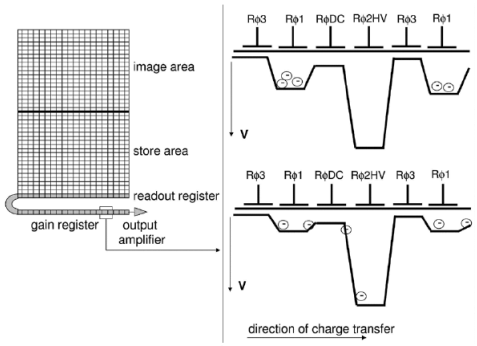


Figure 1.5: Typical electron-multiplying charge-coupled device (EMCCD) from [36]

The main advantages of EMCCDs sensors are its high quantum efficiency and accurate sensitivity to single photon detection. However, the main disadvantage is the high power consumption especially for low intensity levels.

1.2.3 Photon Avalanche diodes: APDs

An avalanche photodiode (APD) [37] is a highly sensitive semiconductor photodiode detector invented in the 90's based on semiconductor photodetectors. The functioning principle to detect photons in low photon flux consists of an internal gain of a junction by applying a strong bias voltage in reverse cf., Fig. 1.7. In general, the APDs consist in structures based on a $p-n$ or $p-i-n$ junctions (cf., Fig. 1.6). In linear mode *i.e.*, with a reverse bias voltage high enough but slightly below the breakdown voltage, the photodiode can be used to obtain a linear response as a function of incident photon flux intensity. When a photon is absorbed in its depletion zone, it creates a pair of electron hole (photo-generated carriers) that are accelerated and then create other electron-hole pairs by ionizing impact which will in turn form others (In a way, this mechanism can be regarded as the semiconductor analog of PMT).

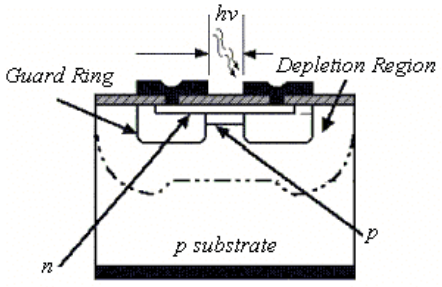


Figure 1.6: Typical avalanche photodiode (APD) from [38]

APDs have the main advantages of being wavelength selective photodetectors which is very interesting for example for fiber optic transmission systems [39]. However, this photo-detector APD has two disadvantages regarding the temporal jitter (temporal resolution), because the APD requires a few photon arrivals for its avalanche detection leading to uncertainty on the arrival time of the photon, considering that the arrival of photons follows a Poisson distribution. The second major drawback comes from the low signal to noise ratio of the photodiode itself.

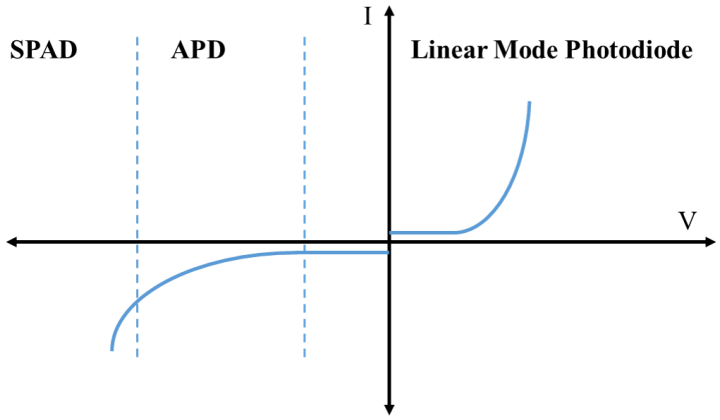
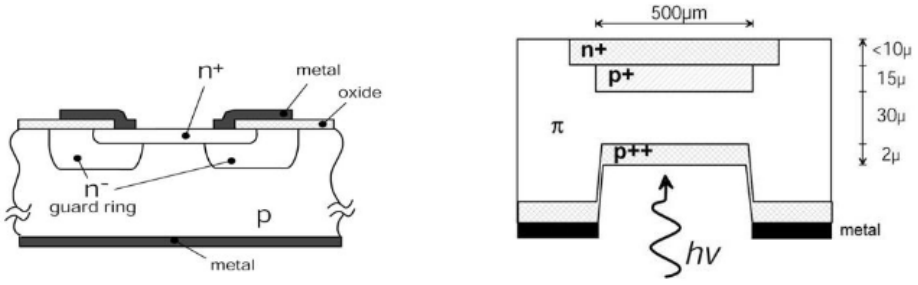


Figure 1.7: Current-voltage relationship of a p-n diode with respect to applied bias voltage.

1.2.4 Single Photon Avalanche diodes: SPADs

Single Photon Avalanche diode (SPAD) is a specific fully depleted photodiode that is reverse-biased as an APD but beyond the breakdown voltage in a steady state namely Geiger mode cf., Fig. 1.7. In doing so, a single photon absorbed in the photosensitive region is enough to trigger diode avalanche by ionizing impact in the crystal lattice. This single free carrier absorbed in the photosensitive region is accelerated by the electric field and leads to a series of impact ionization processes, resulting in a self-sustained avalanche and a consequent high current through the device. This current is stopped and the voltage across the diode is returned to its steady state, *i.e.*, at a voltage above the breakdown thanks to a resistive circuit namely the quenching circuit. This key operation requires appropriate electronics capable of detecting the increase in avalanche current, generating a pulse at the output in synchronization with the change in current, stopping the avalanche by lowering the bias below the breakdown voltage, returning the bias above the breakdown voltage to make it operational again. The quenching circuit influences many characterization parameters of the SPAD, like its Jitter [40], Afterpulsing [41], Fill factor [42], power consumption [43], optical crosstalk [44] detailed in Chap. 2. Generally speaking, this SPAD operating mode allows a very high photon detection efficiency justifying its used in applications where single-photon sensitivity, along with high timing resolution, is required.



(a) Typical reachthrough single photon avalanche diode (SPAD) from [45]

(b) Typical planar single photon avalanche diode (SPAD) from [46]

Figure 1.8: Typical single photon avalanche diode (SPAD) implementations

On the description side of the design implementation, two different implementations of SPADs exist, reachthrough (cf., Fig. 1.8a) or planar SPAD (cf., Fig. 1.8b). Reachthrough structures maximize the absorption probability and show a quantum efficiency of 30% at 500 nm and 70% at 700

nm, thanks to the large absorption region substrate. In the first structure –namely reachthrough SPAD– proposed by McIntyre [46], a customized technology was used with an ultra-low doped p-silicon substrate specially developed to generate the reachthrough avalanche photodiodes. However, this thick depletion region induces a Jitter higher than 300 ps and requires a bias voltage between 300 and 400 V in combination with specialized circuits to dissipate the excess power, making it very expensive [40].

The main difference between the planar structure and the reachthrough structure is the reduction of the surface of the absorbing zone and the use of a guard ring. Planar SPAD topology allows to obtain better performances, such as jitter reduction due to small junction diameters, reduction of dead time due to small junction capacities, reduced power consumption due to narrower depletion zone and thus lower bias voltages and reduced detection probability. Furthermore, the CMOS process allows the on-chip integration of the photodiode readout circuits, but unfortunately implies a low fill factor because of the combination of on-chip integration of the photodiode and its circuitry.

1.3 Mass adoption of SPAD D-ToF sensors

As SPAD technology has matured, a range of biophotonic applications requiring nanosecond timing resolution have been explored, such as fluorescence lifetime imaging microscopy (FLIM), time-resolved Raman spectroscopy, and positron emission tomography (PET). Indeed, despite a high dark count rate [47], SPAD sensors are of choice due to their temporal resolution (picosecond time resolution), single-photon sensitivity, on-chip histogram generation process and, especially for FLIM technology. Nevertheless, for most of these biophotonics applications, the current technology development trend is to increase sensitivity especially in the red/ near-infrared (NIR) regions [48, 49], increase timing resolution [50, 51] and data compression while obtaining compact systems.

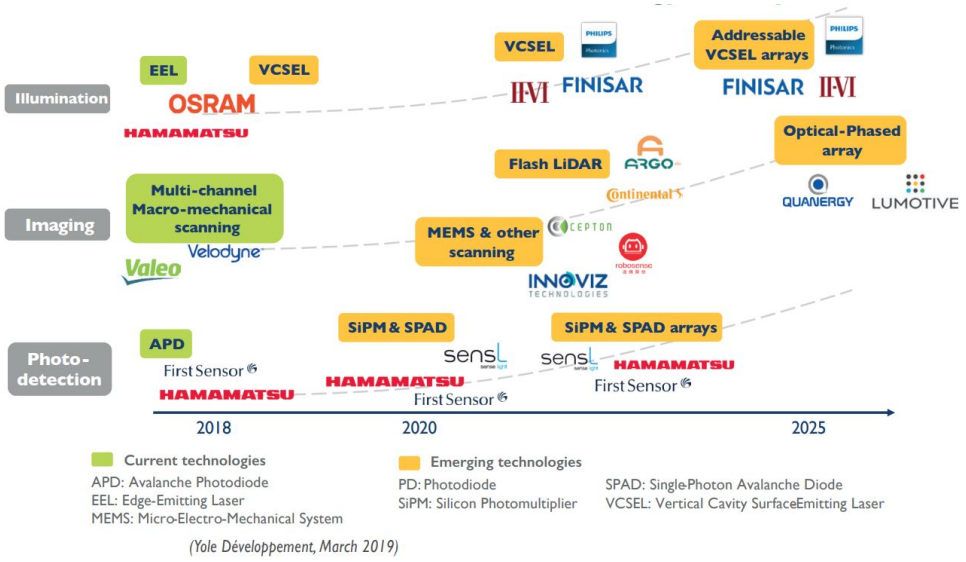


Figure 1.9: LiDAR for automotive technology roadmap [52]

Similarly, global technology companies are actively investing into research and development for the next generation of optical 3D imaging sensors that will make the autonomous driving experience truly possible. These sensors are for indoor and outdoor detection applications in addition

to sensing devices such as automotive cameras and millimeter wave radar. One current investment trend in 3D optical imaging is focused on the SPAD-based ToF sensor cf., Fig. 1.9, which is expected to play a key role in future technologies for high temporal resolution. In addition to its value for high-precision detection, the SPAD-based ToF sensor is also an asset for road condition recognition (*i.e.*, the location and shape of objects such as cars and pedestrians).

Figure 1.10: Evolution of front and rear 3D sensing camera modules in smartphones [53]

Finally, one last but not least, smartphones account for 70% of the total demand of image sensors, which is the largest [53]. Smartphones market is still expected to grow by over 25% [53] because of the emergence of 3D imaging and detection technologies. Indeed, since the early 2010s, some 3D optical image sensing techniques have been used in smartphone features such as AutoFocus2 [54] which is based on autofocusing technique. But, the introduction of the Lenovo PHAB 2 Pro in 2016 and the iPhone X in September 2017 changed the way of 3D sensing and set the technology and use-case standard for 3D sensing in the consumer space. Consequently, the main smartphone manufacturers, such as Samsung, LG, Huawei, OPPO, and Vivo, are engaged in a technical race to come up with the best 3D image sensors whatever the technology used. It is worth mentioning that SPAD-based ToF is less complex than structured light sensors from a software perspective. In addition, the optical system of the SPAD-based ToF sensor simply consists of a Vertical Cavity Surface Emitting Laser (VCSEL) with a diffuser placed on top, whereas the SL requires a complex optical or electrical pattern generator [55]. Regardless of its cost advantage, the SPAD-based ToF sensor system is much easier to calibrate and is more robust to outdoor light conditions. All these key features for all the previously mentioned applications, completely changed the future role of SPAD-based TOF sensing modules, leading to the diversification of its use cases from photography to augmented reality.

1.4 Description of SPAD D-ToF Topologies

The widespread use of SPAD in many applications involves different SPAD architectures for various 3D imaging features. For example, Fig. 1.11 illustrates one possible sensor topology composed of 256×256 SPAD integrated into a 3D-stacked 90nm. The sensor bottom tier is composed of a 64×64 matrix of 4×4 SPADs at 9.18µm SPAD pixel pitch sharing electronic circuitry for a total of 36.72µm grouped pixel pitch. The memory/processing/control electronics circuitry is composed of 16×14 bit counter array with 8 I/O operating at 100 MHz. [6] is one sensor topology among others. And several variants of SPAD imagers with respect to image sensor

format, acquisition scheme, time correlator and processing block architectures, time correlator techniques and post-processing techniques will be detailed in the following subsections.

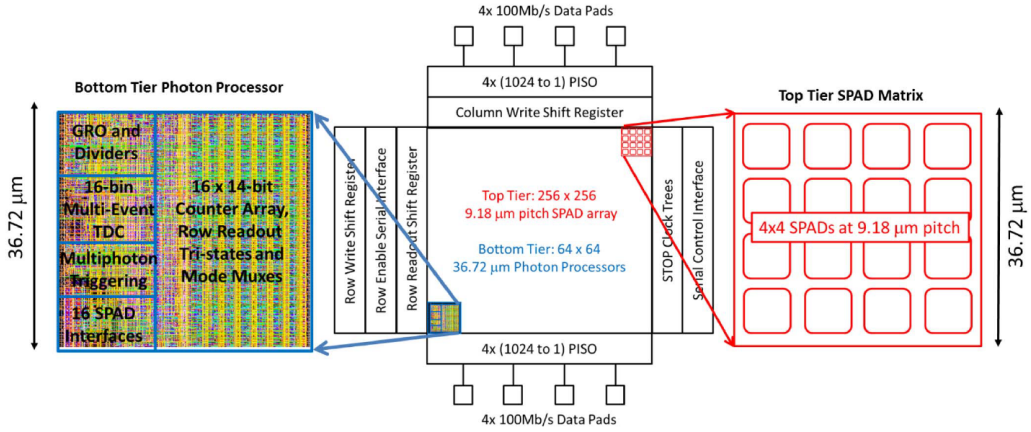


Figure 1.11: Sensor architecture showing the pixel layout on the bottom tier (left) and the SPAD layout on the top tier (right) from [6]

1.4.1 D-ToF image sensor formats

First, SPAD D-ToF image sensors can vary in the number of pixels and their arrangement in the focal plane, by the additional use of a digital micromirror device, or even by the number of processing readout channels. Actually, three different SPAD D-ToF sensor arrangements are clearly identified, single-point sensors, line sensors and array sensors.

First, let's introduce a line sensor which is only composed of a single row of pixels. This image sensor format is commonly used in raster scan LiDAR in 3D imaging applications. The functioning principle is based on the use of a scanning system *e.g.*, rotating mirror on which the receiver and the emitter are mounted or a beam steering, in order to scan one-point at a time horizontally. Consequently, "column pixels" sensor format generally shares time correlation, processing and readout blocks.

The other two image sensor formats are more typically used in LiDAR flash cameras. Single-point sensors are mostly composed of one photodetector [56]. It can also be composed of more photodetectors [57] but with a single readout stream of photon events. In this case, each digital output channel is combined with a tree XOR in order to correctly filter photon events. By doing this, even if the detector samples are independent, they are all sampled and processed as coming from one pixel. Sometimes, the use of a digital micromirror device (DMD) is introduced for the spatial modulation in time of the photodetectors activation [58] allowing a possible future spatial demodulation and thus a zero or almost zero spatial resolution loss.

Otherwise, the development of array sensor formats is composed of columns and rows of independent pixels. This image sensor format commonly used for LiDAR flash cameras provides consistent spatial resolution by using an optical lens in front of the emitter to illuminate the entire scene with a single image readout instead of a scanning system. Some works implement a cluster of photodetectors in a single pixel, namely macro pixel allowing to improve the photon count rate and thus diminishing dark count rate effect [9]. This further allows to optimize pixel fill factor by mutualizing pixel electronics that is to say processing and memory units.

1.4.2 Time correlator variants

Once a photon arrival generates a charge carrier, this voltage drop is correlated in time with the help of a time correlator in the typical case of D-ToF. There exists various timing circuitry with high temporal resolution such as time to amplitude converter (TAC) [59] and time to digital converters (TDC) [60].

Time-to-Amplitude Converter

A TAC [59] consists of a capacity that charges during the time interval from the arrival of the reflected light signal to the synchronization signal of the emitter source. At the end of the integration, the amplitude of the sampled voltage is equivalent to the measured time. This voltage can be then read out in its analogue form, but is usually sampled by an analogue-to-digital converter (ADC) at the pixel level or on-chip to avoid adding more noise to the measurement during the readout process. Although its performance decreases when achieving higher sample rate due to mismatches (*i.e.*, offset and gain) and non-linearity leading to sample-time errors [61]. The advantage of TACs over their digital counterparts is their area. However, the gap is being gradually closed by increasingly smaller CMOS digital technology nodes, which encourage the adoption of the TDC as the circuit block of choice for photon timing applications.

Time-to-Digital Converter

TDC consists in being first triggered to start timing by the periodic reference synchronization signal from the emitter and triggered a second time to stop timing by the occurrence of a photon event. However, the timing convention that inverse start and stop triggers, is rather used to improve the power consumption and data bandwidth performance of the system (*i.e.*, only trigger on a valid photon event). The measured time difference is directly mapped to a binary encoding (*e.g.*, one-hot encoding). TDCs are commonly used in D-ToF cameras at the expense of TACs since TDCs provide a shorter conversion time and are less sensitive to temperature or voltage variations. Four main topologies of TDCs integrated into ToF image sensor architectures have been identified, Flash TDC [60], Vernier Delay line TDC [62], Pulse-Shrinking Delay Line TDC [63] and Gated ring Oscillator TDC [64]. Flash TDC [60] architecture is the most popular due to its simplicity, low latency and reliability. The resolution of the flash TDC is equal to the delay (τ_i) and the length of the delay line (N) defined "commonly" the required dynamic range.

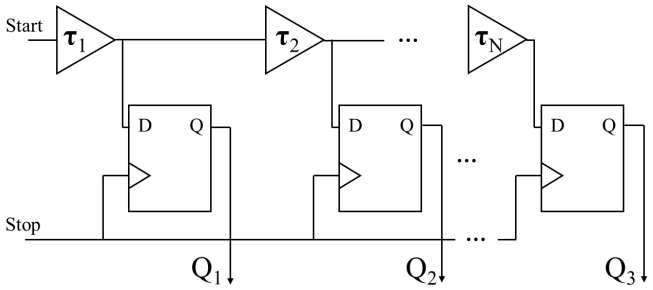


Figure 1.12: Flash Time to Digital Converter illustration

Because the resolution of the classical delay line TDC is limited by the cell delay (τ_i). The second example, namely the Vernier Delay Line TDC [62] enhances the TDC resolution beyond the cell delay in introducing two delay lines, start and stop lines and using the gate delay as a time unit. A start signal is propagated through the first delay line with τ_1 delays and a stop signal propagates through the second delay line with τ_2 delays where $\tau_1 < \tau_2$, with a flip-flop at each corresponding pair of cell delays where the stop signal is clocking the flip-flop at each stage.

The measurement is completed when STOP overtakes the START signal in steps of resolution ($\Delta t_{delay} = \tau_1 - \tau_2$, for more details see [65]).

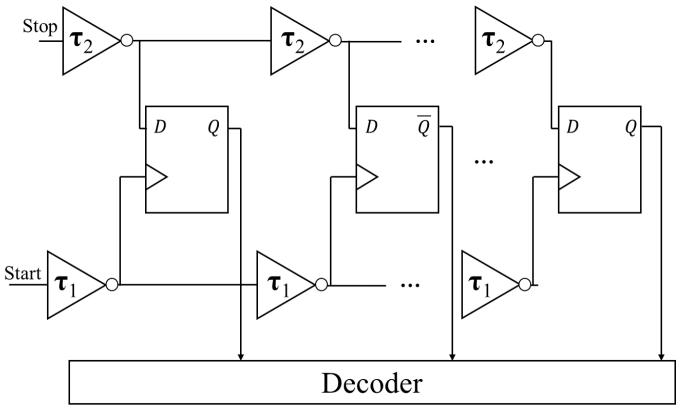


Figure 1.13: Vernier Time to Digital Converter illustration

However, the Vernier delay line TDC requires more silicon area due to the doubling of the number of delay lines compared to the Flash TDC. The simplest way, in terms of silicon area, to measure time as a Vernier TDC delay line is to use a shrunk delay loop with a counter. This method can be implemented in different ways but the concept named Pulse-Shrinking Delay Line TDC [63], simply consists in having all buffers in the delay line (τ_1 in Fig. 1.14) except one (τ_2 in Fig. 1.14) with equal rising and falling times. The shrinkage will operate at each loop iteration until the pulse vanishes. The number of cycles taken to vanish the signal is proportional to the time measurement, Gated ring Oscillator TDC [64].

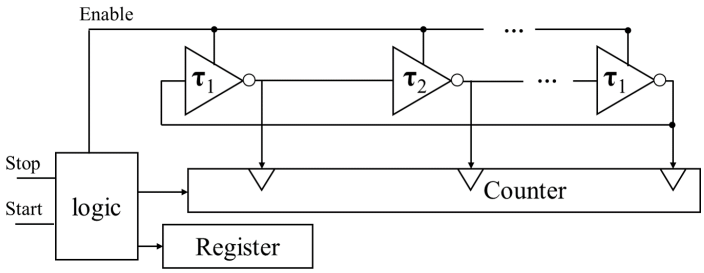
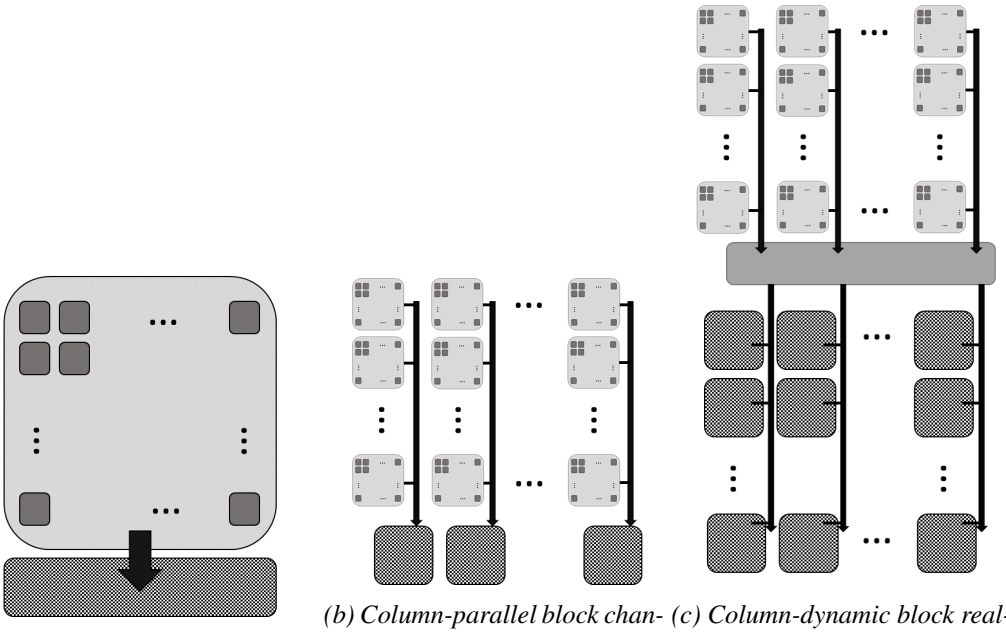


Figure 1.14: Gated Ring Oscillator Time to Digital Converter illustration

1.4.3 Timing unit architecture

There is a wide variety of possible time correlator (as presented above) and processing block architectures (presented in the following sub section), single block, column-parallel block, column-dynamic block reallocation and per-pixel bloc (see Fig. 1.15). All these architectures have their pros and cons depending on the image sensor format. For example, single block is naturally the most relevant architecture allowing higher active pixel area as illustrated in Fig. 1.15a. However, this architecture requires the use of a TDC with a high conversion rate in order to timestamp every event provided by the SPAD array. On the contrary, a column-parallel architecture provides a shared column unit. Therefore, rows of pixels are sequentially processed (cf., Fig. 1.15b). The limitation of this architecture is the same as for single block architecture because a single row may be processed at a time, which might impose long delays to process the whole array. The column-dynamic block reallocation (cf., Fig. 1.15c) is more relevant in a way that several processing blocks are available for each column and are dynamically assigned to column pixels. Depending

on the image sensor format, a per-pixel unit block may be interesting since all pixels get their own processing units and no more photon events are ignored because of the absence of available processing units. Before the recent development of advanced 3D stacked technologies, the processing unit was generally placed in the peripheral circuitry of the active pixel area therefore reducing the fill factor cf., Fig. 1.15d. Thanks to the development of advanced 3D stacked technologies, the integration of processing units is now possible below the active pixel area in order to maximize the fill factor so as the processing unit performance as illustrated in Fig. 1.15e. Some other works customized the per-pixel unit block architecture. For example, [66] replaces the TDC-per-SPAD canonical architecture with a shared TDC, thus optimizing the use of the pixel area in spite of inducing a spatial resolution reduction as illustrated in Fig. 1.15.



(a) Single block *(b) Column-parallel block channel* *(c) Column-dynamic block reallocation*

(d) Per-pixel bloc 2D CMOS *(e) Per-pixel bloc 3D-IC*

Figure 1.15: Architecture of temporal correlation or processing channels in an image sensor.

1.4.4 Memory unit/ IO pads architecture

There exists two main groups of SPAD architectures, one based on time gating single photon counting (TGSPC) [67, 68] approach cf., Tab. 1.3 and the other based on time correlated single photon counting (TCSPC) [6, 69] approach cf., Tab. 1.2. For TGSPC approach, there is almost no processing, memory and readout electronics in the pixel, since no time correlator or even on-chip memory are used. Although this approach reduces complexity and pixel pitch, it is limited by the long measurement time and is hardly applicable to 3D imaging.
Otherwise, for the TCSPC approach, after the conversion of the photon carrier detection into a

timestamp measurement (through the use of a time correlator as presented above), this time stamp value or time of flight (ToF) needs to be stored for later statistical analysis. For example, it simply consists in finding the time when the most photons hit the SPAD sensor because of the synchronized laser pulsed-light for depth estimation. To do this, the ToF sequences must be stored, but these are too large to be stored at the pixel level. That is why, ToF values are stored in the corresponding time bins of a histogram. However, the histogram memory blocks also remain a challenge due to the amount of the local storage required for the ToF histograms. To illustrate this point, let us take the example of a 10-bit coded ToF value, the one-hot encoded histogram would be composed of 1024 bins providing measurements with a resolution of 1cm for a depth range of 10m. In order to acquire a relevant distribution of ToF values, each bin needs its own counter with a sufficient bit-width (*e.g.*, 10bit). The memory required by each pixel is then 10kb. Furthermore, the crucial issue of the sensor data throughput arises. Let's take the example of a VGA pixel matrix (640×480), which already requires more than 3Gb per frame. For a frame rate of 30fps, this would imply a 92Gb/s chip data throughput. So, even assuming high-speed chip I/O pads capable of 300Mb/s, the TCSPC solution would require more than 300 pads.

Besides technology trends which allow the removal of the memory/processing/control electronics from the sensor's focal plane, we review in the literature several unconventional system designs to overcome SPAD limitations. For example, to optimize the processing of SPAD data two methods consist in dividing the histogram construction into subparts if the speed of acquisition is not a concern, namely the Partitioned inter-frame Histogram and the Folded/Shifted inter-frame Histogram as illustrated in Fig. 1.16 with the description of an architecture with only 16 counters.

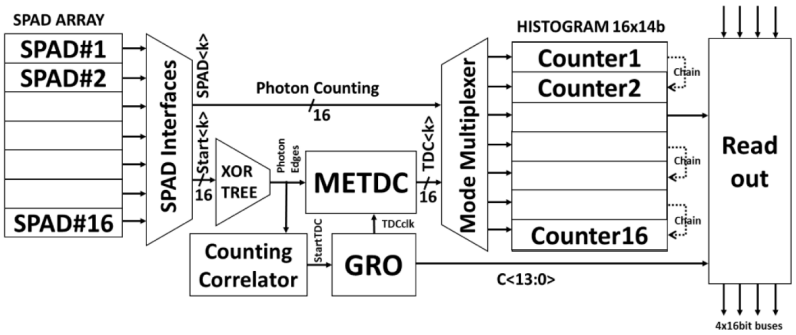


Figure 1.16: Pixel sharing architecture from [6]

Partitioned inter-frame Histogram (PiFH)

This implementation is straightforward, since it sequentially builds complete histograms in a partitioned manner and then requires only comparing every highest peak location of each partial histogram in order to find the peak location (see Fig. 1.17). In taking the example of a 10-bit ToF pixel counter with a 8-bit coded ToF value for the partial histogram, we therefore obtain $2^{10-8} = 4$ partitions. In doing this, it replaces the build of the histogram of $2^{10} = 1024$ bins by 4 sequential partial histograms of $2^8 = 256$ bins. Physical memory is thus $4\times$ reduced in comparison to the complete histogramming by using the same memory footprint per partial histogram and building it sequentially in time. This approach unfortunately reduces the acquisition frequency of the sensor since it requires more photons to be acquired for all sub-histograms.

Folded/Shifted inter-frame Histogram (FiFH/SiFH)

This approach is based on building two or more histograms with various binning in time for all the sequential histograms. Let's take the example of a complete histogram of n bins. The time binning

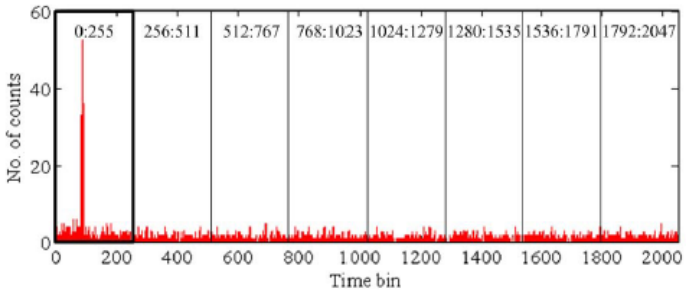


Figure 1.17: Partition histogramming from [70]

depends on the number of steps of histograms built (s) that is to say equal to $\sqrt[s]{n}$. For a folded inter-frame histogram in typically two steps, each sub histogram will thus be composed of $2^{\frac{log2(n)}{2}}$ bins. The first histogram will have a binning of $\sqrt[2]{n}$ while the second histogram will have no binning but a dynamic range equivalent to the bin size of the first histogram. In other words, the first histogram corresponds to the most significant bits (MSB) and the second the least significant bits (LSB). In comparison to the complete histogramming method, the memory footprint is decreased by a factor of $2^{\frac{log2(n)}{2}}$ (*e.g.*, for 10 bits ToF pixel value with a 5 bits coded ToF value for the folded histogram). One of the major drawbacks of this approach is related to the noise floor of the most significant bit histogram and of the least significant bit that are larger than the complete histogram noise floor. Regarding the least signifiant bit histograms, this results from the noise folding, *i.e.*, the cumulative sum of all the bins modulo $\sqrt[s]{n}$ that are equal (the smaller the number of bins of representation, the higher the folding order). To address this issue, [71] proposes "Shifted inter-frame Histogram" (SiFH) to filter the second histogram using the estimated result from the first, to increase the Signal to Background Ratio.

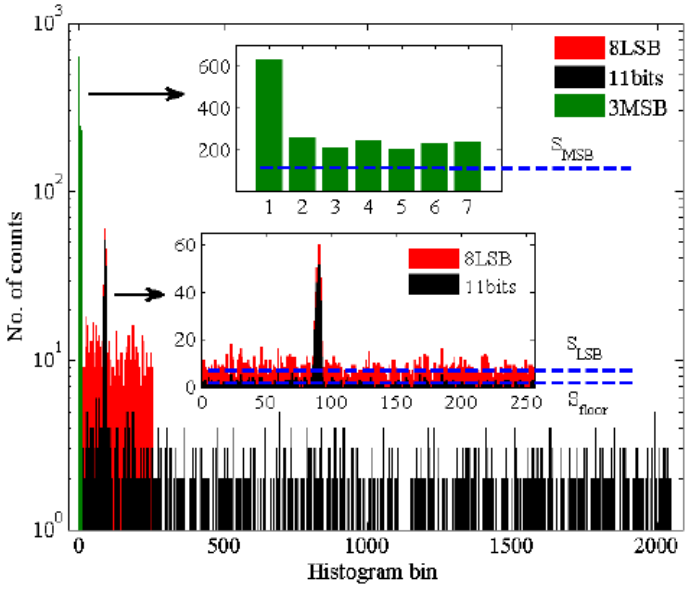


Figure 1.18: Folded inter frame histogramming from [70]

Fig. 1.18 illustrates another operating mode, where the MSB histogram is 3-bit histograms and detects the peak with a boarder error around 255 bins (complete histograms of 2^{11} bins). The second histogram (*i.e.*, LSB histogram) is thus built from the 8 less siginficant bits, *i.e.*, a histogram size of 2^8 bins.

1.4.5 SPAD activation time dependent

In conventional ToF cameras, the pulse laser and SPAD sensor are temporally synchronized. However, due to the peculiar histogram formation process, the SPAD cameras cannot operate reliably under ambient light (*e.g.*, sunlight in outdoor conditions). Indeed, the early arriving ambient photons prevent the SPAD from measuring the signal (laser) photons that may arrive at a later time bin of the histogram. This nonlinear distortion, the pileup effect, distorts the histogram measurements towards earlier time bins, as shown in Fig. 1.19. Reliably locating the laser pulse is consequently challenging and usually results in large depth errors. Therefore, an asynchronous acquisition method has been proposed by [72]. This consists in decoupling the SPAD on/off times from the laser cycles by allowing the SPAD acquisition windows to have arbitrary start times with respect to the laser pulses. As for the SPAD cycle, it is defined as the duration between the SPAD activation and the photon detection (or following SPAD activation time). SPAD listening windows misalignment is operated with various shifts in order to spread out the noise effect while mitigating the pile-up effect [73]. This method typically enables a reduction of the digital dynamic range related to histogram bin counters and thus lowers pixel memory requirements. Nevertheless, since the TDC is triggered to start timing by the periodic reference asynchronization, this method implies a higher power consumption. In addition, this can only be used for specific SPAD operating systems working with a single photon-per-pulse detection.

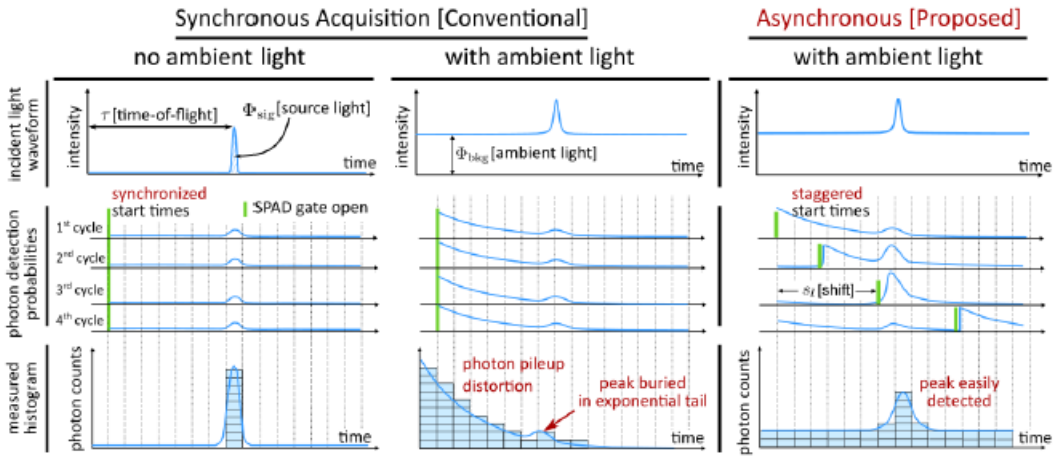


Figure 1.19: Asynchronous acquisition scheme from [72]

1.5 Limitations of SPAD D-ToF

Since the invention of SPAD sensors in 2003 [32], most of their key characteristics have been particularly improved, namely RGB-D alignment, timing resolution, dynamic range and pixel-level processing. For instance, one of the first SPAD D-ToF sensor arrays [74] was simply multiplexed to an external TDC IC. This was limiting in terms of acquisition time because of the sequential pixel measurements. Moreover, the SPAD array design was not scalable to a high resolution sensor, which is necessary for the 3D imaging application. Various works qualitatively compares depth map sensing technologies with respect to each system feature and typically highlights that spatial resolution and power consumption remain one of the main issues related to SPAD D-ToF sensors [75]. Since then, the current trend is to design larger SPAD arrays [76] to address the mass industry requiring high-performance sensing solutions (*e.g.*, for advanced driver-

assistance systems [77]) and to be competitive against alternative depth sensing approaches (*e.g.*, stereoscopic disparity maps (Stereo) [78]). The first step forward after was to integrate the timing circuits on the same silicon die as the SPAD sensors, which allows for parallel measurements for a higher degree of scalability and thus shorter acquisition time [79]. These sensor-level improvements successfully extended the use of SPADs to a wide range of emerging applications such as medical imaging [80], astronomical imaging [81], lidars [82] and light communication [31].

Although offering a competitive depth accuracy of a few ps and slightly higher spatial resolution, D-ToF still faces considerable hardware constraints which limit higher acquisition frame rate and higher image formats (*i.e.*, number of pixels) with a reduced pixel pitch. To illustrate this point, Tabs. 1.3 and 1.2 shows the performance metrics for the recent SPAD ToF imaging sensor arrays. It highlights the trade-off between low complexity to achieve high fill factor with a reduced pixel pitch and a high level of parallelism to achieve low acquisition times. Several actors like STmicroelectronics (ST), École Polytechnique Fédérale de Lausanne (EPFL), Fondazione Bruno Kessler (FBK), Adaps Photonics Company (APC), Xidian University, Ulsan University, Sony, Canon and Panasonic have made many efforts to go beyond the limits of standard SPAD development, creating large-format SPAD cameras combined with time gating (cf., Tab. 1.3) or time-to-digital converters (TDCs) on chip (cf., Tab. 1.2). Actually, to the best of our knowledge, we have reported three ways to improve the fill factor of a smart SPAD ToF pixel: 3D-stacking, microlensing and reducing the per-pixel complexity. Microlensing is a technique used to improve the effective fill factor of SPAD arrays by redirecting photons away from dead zone (*e.g.*, electronic circuitry) within the pixel and toward the active area of the SPAD. Let's take the example of [83] which increased the fill factor from 13% to 40% with microlensing (cf. Tab. 1.2). However, this approach does not result in a compact system since it does not reduce the pixel pitch either.

Alternatively, the use of advanced IC technology, including 3D-stacking (3D-IC) [25] firstly enabled to increase overall performances, reducing power consumption at a smaller footprint than traditional 2D-IC [6, 84, 85]. For example 3D-stacking allows the integration of backside-illuminated SPADs on the top tier, and control/processing/readout electronics on the bottom tier [6, 84, 86, 87]. The 3D-stacking further enables the implementation of two different process nodes for the SPAD photodiodes and electronics circuitry optimizing SPAD process (as proposed by [6] allowing more complex per-pixel circuitry).

Even if 3D stacking seems very relevant, first of all in terms of power consumption, the gain in pixel pitch and thus the maximum size of the SPAD array is not yet sufficient. The key for a larger image sensor remains the pixel pitch related to the amount of functionality per pixel. Therefore, several researchers in the scientific community have implemented hardware implementation tricks to reduce the electronic circuits per pixel, such as TDC sharing [6, 69, 87] or partial histogramming [6, 85, 88, 89]. For example, [6] proposed five different operating modes based on an histogramming TCSPC-based approach but in a block of 4×4 SPADs connected to an advanced functionality pixel circuitry. However, in larger SPAD arrays, massively parallel data processing must be addressed, requiring a large number of memories for pixel-level histogramming [85, 87]. Because the memory area is especially challenging, a coarse-to-fine historamming approach was implemented by [6,85,89]. [88] even proposed a dichotomic processing methods which consists in sequentially shrunk by half the temporal resolution of analysis with the used of Up/Down counters (UDC) at the expense of image acquisition frequency.

Another processing methods have been further investigated, such as time gating approach [86], requiring less in-pixel processing, memory and readout electronics. This approach is typically used to reduce dark noise, as well as to reduce the pixel pitch and the amount of memory required per pixel with the 1b pixel counter. However, this approach is limited by the extended measurement time due to the large number of counts required for each gate window shift and by

the generation of the gated window cf., Tab. 1.3.

Actually, to the best of our knowledge, even though TDC-based approach enables precise photon arrival time stamping. The TCSPC-based approach is not suited for scaling because of large circuitry requirements, since it practically consists in detecting incident photons and storing all the measured arrival times in a large histogram in order to estimate the statistical distribution over a large number of pulse periods. This operating mode unfortunately involves a large amount of raw data associated with each depth map. **From this premise, this manuscript focuses on alleviating the TCSPC imaging mode by proposing alternatives of canonical acquisition systems.** This technical challenge is further complicated by the fact that, depending on the incident photon flux, the collected histogram may include a majority of photon arrival times due to the ambient light and not laser pulses, making the extraction of information more difficult as the ratio of photons coming from the active light source and those from Background Illumination (BI) could be very low (lower than 1%) [70].

	[6]	[87]	[84]	[83]	[88]	[89]	[85]
University	ST	FBK	Sony	ST	Ulsan	FBK	APC
Techno.	90nm 3D-BSI	40nm 3D-BSI	90nm 3D-BSI	40m CMOS	110nm CMOS	110nm CMOS	65nm 3D
Time architecture	GRO	TWC+TDC	TDC	Gated RO	TDC+UDC	RO	TDC
Array size	64x64	320x240	189x600	192×128	80x60	64x64	240x160
Pixel Pitch (µm)	36.72	12.5	10	18.4x9.2	75	48	16
Fill Factor (%)	51	21	27.4	42	10.4	12.9	49.7
Range (m)	50	4.4/2.2	300	40	45	8.2/1200	30
Temporal Res. (ps)	35/560	200/300	1000	33	500	250-10000	97.65
Sensor Spec.	Shared TDC			microlensing	In-Pixel Histogram		Shared Mem.
Number of SPAD per pixel	16	16	9/16	1	6	4	64
Number of TDC	4096	4800	384	24 576	4800	4096	600
TDC depth	14-b	5-b	12-b	12-b	7-b/6-b	16-b/13-b	10-b
Memory	Ripple Counters	SRAM	SRAM	Ripple Counters	UDC	Counters	SRAM
On-chip mem. (kb)	896	1800	9108	196	86	65	2450
Architecture	Reconfigurable	Partial Histogram	Full Histogram	Digital Calibration	Partial Histogramming		
Fps	30 - 760	30	20	18.6k	30	25	20
Power consumption (mW)	77.6	<1500	-	140	1500	205.7	90

Tableau 1.2: Survey of major SPAD D-ToF sensor developments.

	[67]	[68]	[90]	[5]	[91]	[76]	[92]
University	ST	ST	Fraunhofer IMS	EPFL	ST	Canon/EPFL	MOTIE South Korea
Method	1-bit I-ToF	1-bit I-ToF	cw-I-ToF	1-bit I-ToF	I-ToF	1-bit scan. win.	Analog I-ToF
Techno.	130nm	40nm CMOS	0.35µm CMOS	180nm CMOS	40 nm	0.18 µm	110 µm
Array size	320x240	256x256	64x32	512x512	128x128	1024x1000	64x64
Pixel pitch (µm)	8	16	150	16.38	40x20	9.4	32
Fill factor (%)	26.8	61	3.14	10.5	13	7	17.3
Pixel counters	1-b	1-b	3x9-b	1-b	2x16-b	1-b	1x analog
Range (m)	0.6	60	20	15	3	1.6	40
Fps	130k	100k	33k	97.7k	500	24k	90
Power consumption (mW)	69	-	50	26.7-700	185	284	20

Tableau 1.3: Survey of major SPAD I-ToF sensor developments.

1.6 Thesis Structure

This section will present all contributions from each chapter as illustrated in Fig. 1.20.

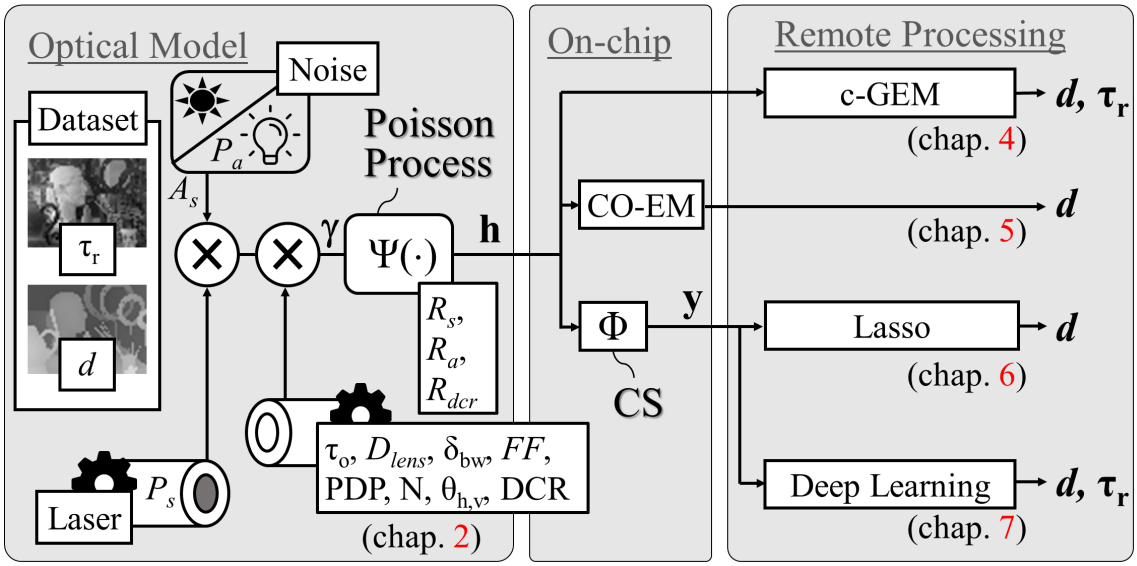


Figure 1.20: Schematic description of our contributions. Where **d** *and* $\boldsymbol{\tau}_r$ *refer to modality reconstructions, which are respectively depth and luminance.*

Chapter 2 This chapter will firstly present system-level characteristics illustrated on the left side of the Fig. 1.20, namely "Optical Model" (*i.e.*, the sensor, the laser and the radiometric characterisations of the active imaging system). Since no SPAD sensors were available during this thesis, a physic-based analytical model has been developed for generating physically-plausible SPAD raw data, **h**, used to evaluate our on-chip/remote processing methods. Secondly, the theoretical study of the statistics distributions of **h** from the time correlated single photon counting (TCSPC) single photon avalanche diode (SPAD) based imaging system is presented. This study gave various interpretations on the nature of the statistical data and the mixture probability distribution functions as a function of the 3D imaging system configurations (*e.g.*, depth range measurements, background illumination, laser pulse width, time resolution) leading to several solutions based on Expectation-Maximization Algorithm in Chap. 4, Chap. 5, Compressive Sensing combined with regularization algorithm in Chap. 6 or Deep Learning Algorithm in Chap. 7.

Chapter 3 Before presenting all on-chip and remote processing solutions in order to overcome SPAD limitations, which are, as you know background noise robustness and TCSPC data format memory requirement, this chapter presents a quick overview of the compressive sensing (CS), expectation maximization (EM) algorithm, deep learning (DL) background and image super-resolution algorithm. It starts with the presentation of the Expectation-Maximization algorithm with its variants, giving basic mathematical tools for understanding Chap. 4 and Chap. 5. The second section of this chapter presents the key theoretical concepts of CS, then presents the main problem statements of the CS recovery methods and ends with state-of-the-art (SoA) implemented algorithms in order to give all the keys to understand the solutions proposed in Chap. 6 and Chap. 7. The third section of this chapter presents the deep learning field for image reconstruction, which appears to be a very powerful algorithmic tool. This section will briefly introduce the mathematical background behind the success of deep learning for computer vision tasks and will finish with the presentation of the main deep learning topology, the Convolutional Neural Networks (CNN)

(used in Chap. 7). Finally, the last section presents various techniques that enable to virtually increase the resolution of an imaging system, namely image super-resolution, which inspired the proposed solution reported in Chap. 7.3.

Chapter 4 This chapter presents a remote processing method (cf., Fig. 1.20) which consists in estimating the depth map, $\mathbf{d}$, and luminance map, $\boldsymbol{\tau}_r$, from the estimation of the 4 latent parameters of each truncated-Shifted Erlang ($\mathscr{E}$) distribution of a mixture model, only from an observed draw of the process distribution in the shape of a time-of-arrival histogram, $\mathbf{h}$. In contrast to SoA approaches, the proposed solution aims to explore the high photon count synchronous SPAD operating mode which cope with the short range 3D imaging (*i.e.*, few meters under high background illumination). To address this highly non-convex problem, a customized nested Expectation-Maximization algorithm (c-EM) has been designed based on a proper combination of Maximum Likelihood Estimation, Moments Matching, Parametric Imputation and support estimation via Variable Neighborhood Search. Furthermore, a Genetic search with a statistical test has been added on top in order to handle the order selection in mixture model and to ease parameters estimation without being stuck into a local minimum. This complex processing method does not allow for its pixel level integration, so we propose to continue to the next chapter with other system level configurations enabling on-chip processing.

Chapter 5 Aiming at replacing TCSPC system at pixel level, the proposed solution in this chapter only considers the high photon counts asynchronous SPAD operating mode resulting in Gaussian-Uniform statistical SPAD data model approximation, commonly considered in SoA works. The proposed algorithm is performed by a 2-stage Expectation Maximization (EM) algorithm for online peak detection with its Register Transfer Level hardware implementation (cf., on-chip CO-EM block in Fig. 1.20). This online estimator consists in sequentially computing the statistical parameters that directly encode the latent spatial distance $\mathbf{d}$ and albedo $\boldsymbol{\tau}_r$ information of the illuminated scene. This proposed solution allows for a reduction of the pixel pitch by around 40 % while improving depth reconstruction accuracy in comparison to SoA acquisition methods.

Chapter 6 One major drawback of the efficient statistical compression method proposed in Chap. 5 is related to the inability to remotely correct depth misestimation because of SPAD raw data information loss since statistical parameters estimated don't fit to $\mathbf{h}$. This chapter consequently explores another way to effectively overcome the TCSPC data format problem, in the high photon counts asynchronous SPAD operating mode with a sparse SPAD data model (*i.e.*, short light pulse more or less equivalent to the histogram bins of duration). The second proposed TCSPC alternative is based on pixel-wise CS with its implementation using an in-pixel Cellular Automaton (cf., on-chip $\boldsymbol{\Phi}$ and Bi-objective histogram reconstruction blocks in Fig. 1.20). It demonstrates its effectiveness in terms of probability of the correct histogram peak detection while extending capabilities to multi-peak detection. From the hardware point of view, examples of digital logic syntheses are also reported, namely for silicon area reduction and comparison with existing designs. According to all the detailed results, the proposed approach paves the way to alternative acquisition schemes reducing the dimensionality of the data without limiting the extraction of information. Indeed, the point is that this approach does not limit the remote interpretation of the collected data to a simple peak detector but allows for understanding more complex underlying information, for example using deep learning tools as proposed in the next chapter.

Chapter 7 The last chapter of this thesis presents two deep learning models to fully exploit the potential of CS by reconstructing both depth and intensity maps in all SPAD operating modes (cf., Multimodal DL block in Fig. 1.20). A first deep learning model designed to process low photon counts SPAD data (long range imaging system model which neglects noise photon counts pileup

effect and the hardware limitations of the SPAD sensor imaging system) has been developed to introduce multimodal reconstructions. From the topology point of view, a specific compression pattern layer operates a learned projection into a smaller latent representation space. Subsequently, the learned compression pattern layer is replaced by a data agnostic hardware-compliant implementation CS scheme, allowing to relax the hardware constraints on the SPAD sensor. The CS scheme firstly reduces the number of memory points and the number of bits per point, thus reducing the pixel pitch by using a cellular automaton (CA) as presented in Chap. 6. This finally allows to get out of the restricted mode of operation at low photon counts and thus improving multimodal reconstructions. On the other hand, the second model developed aims to get rid of the sparsity constraint and to provide super-resolution reconstructions, thus extending the use of the CS to all the operating modes of SPAD (*e.g.*, non sparse SPAD raw data, SPAD pixels sharing electronic circuitry).

2

A Physics-based analytical Model of TCSPC data

In active 3D imaging based on a Time Correlated Single Photon Counting (TCSPC) system, photons "of interest" measured with Single Photon Avalanche Diodes (SPADs) are often mixed with high background photon noise. This low Signal-to-Background ratio (SBR) makes the reconstruction of luminance and depth maps difficult since the depth information is then deduced by analyzing the statistical distributions of the photon arrival times. This chapter starts with the introduction of the receiver and transmitter properties, such as the distance range measurements, the probability detection efficiency, the sensor dead time, the crosstalk. As a second step, the radiometric characterization of the active imaging system, i.e., physical phenomena related to the environment of use, is provided. This chapter then presents the physic-based analytical model for D-ToF sensor. This model aims to simulate pseudo realistic TCSPC SPAD raw data by taking into account physical phenomena related to the SPAD/LASER characteristics and to the scene during the light pulse propagation and reflection. Finally, we determine the density functions that describe the statistical distribution of the photon arrival times. This analysis will allow us to further understand the nature of the statistical data according to the operating mode of the SPAD D-ToF sensor and to generate synthetic SPAD raw data that will later be used to evaluate our acquisition/reconstruction methods.

Contents

2.1 Characterization of single photon ToF sensor visions

The importance of accurate modeling of SPAD sensor data is threefold. First, it gives us the possibility to generate pseudo-realistic SPAD data. Second, it allows us to understand the interactions between all the three device components of the imaging system: the transceiver, the receiver and the environment. Third, a system-level understanding of the ToF vision system helps to determine the advantage of one topology over another depending on the intended application. Before introducing the TCSPC-based SPAD simulation model, a brief list of system-level analysis properties is provided. The correlation between these properties unfortunately leads to trade-offs in SPAD performance and various modeling of the SPAD raw data distribution.

2.1.1 Properties of SPAD sensor

- **Distance range** is usually expressed as a distance or time value and is preliminary defined by the timing electronics. However, the measurable distance range essentially depends on the transmitter power and on the optical specifications because of the signal-to-noise ratio impact.
- **Spatial resolution** describes the number of pixels in the 2D dimension of the sensor focal plane.
- **Frame rate** is the rate at which a depth image is provided by the depth camera. The value is expressed in frames per second (fps). This parameter defines, for a pre-determined pulse period time, the number of pulse laser acquisition period per frame and thus determines the quantity of photons contained in the statistical representation of photon arrival times.
- **Quantum efficiency (QE)** is the percent of photons effectively converted to charge carriers. It is thus defined as the ratio between the number of charge carriers detected and the incident photon flux. The QE properties stands for the SPAD sensitivity which is heavily dependent on the wavelength of the incident light because of physical properties of semiconductors.
- **Fill Factor (FF)** is the ratio between the photodiode active surface of the pixel and the total surface of the pixel. A photon quantity loss can be observed when timing and peripheral electronics are integrated in-pixel. But it can be bypassed, by an appropriate use of microlens on top of each pixel to redirect the light back to the photodetector.

$$FF = 100 \cdot S_{active} / S_{pixel} \tag{2.1}$$

- **Photon Detection Efficiency (PDE)** defines the probability that an incident photon absorbed by the photodetector triggers an avalanche as a function of the ratio of the size of the photosensitive region to the total surface of the photodetector. PDE depends on the bias of the avalanche photodiode and is wavelength-dependent, since the PDE is derived by the QE previously defined (TP stands for trigerring probability) [44].

$$PDE = QE \cdot TP \cdot FF \tag{2.2}$$

- **Dead Time (DT)** is the time following an avalanche while the SPAD is unable to detect another photon. This dead time [93, 94] refers to the time that the SPAD is maintained with a polarization below the avalanche voltage of the junction. It is thus mainly related to the quenching time [45]. This parameter causes specific distortions of the histogram distributions cf., sec. 2.3.

- **Dark count rate (DCR)** stands for thermally-generated carriers that trigger the avalanche process. Indeed, it is still possible to observe output pulses when the SPAD is completely into the dark. The resulting average number of counts per second is called dark count rate (DCR) and defines the self-detector noise. This parameter is a function of temperature and bias voltage via Shockley Read Hall theory [95]. From a statistical point of view, this process of generating false detections follows a uniform distribution and with a high DCR value can completely corrupt the photon arrival time distributions.

- **Afterpulsing** stands for another noise effect triggering the avalanche process right after photon-generated carriers. Indeed, carriers photogenerated during an avalanche that were captured in traps in the multiplication region can be randomly released later and can potentially re-trigger an avalanche correlated in time with the original pulse. The afterpulsing effect [41] is correlated to the avalanche history, *i.e.*, the higher the avalanche rate, the greater the afterpulsing effect (non-Markovian property). Note that afterpulsing is a parameter strongly related to the dead time of the photodiode because the carriers caught in the deep traps have a higher probability of being released before the photodiode is biased above the breakdown voltage again [46].

- **Photon timing jitter** corresponds to the time uncertainty to determine the photon arrival time. The measurement uncertainty of a photon's arrival is caused by the statistical variation in the detection mechanism within the SPAD. The detection mechanism uncertainty can be due to a depth of absorption of photons, a diffusion time to the active p-n junction, the build up statistics of the avalanche, the jitter of the detection and timing circuitry. For more details please refer to [40] which presents an analytic modeling method for SPAD timing jitter prediction.

- **Cross Talk** is a phenomenon by which a signal detected on one pixel creates an undesired effect on another neighbor pixel. There exists two kinds of cross talk. First, the Optical cross talk stands for released photons during the avalanche due to electroluminescence that cause avalanches and thus generate spurious pulses in the neighboring area. Secondly, an electrical crosstalk is the fact that an absorbed photons beyond the p-n junction deep in the substrate generates carriers diffusing in the neighboring pixels.

2.1.2 Radiometric characterisation

- **Solid angle** is commonly noticed ω and is the area of the projection of the object onto the centered unit sphere. The solid angle is typically used to estimate the radiance of a light source by dividing the irradiance by the solid angle value. The smaller the solid angle of divergence the higher the brightness of the laser, which increases the signal-to-noise ratio, but reduces the spatial illumination.

- **Inverse-square law** states that a given physical quantity is inversely proportional to the square of the distance traveled by the light source. Indeed, if we consider an isotropic point

source, then by definition, this source will radiate in all directions with the same intensity. Consequently, a sensor located at a distance d from the source will capture a fraction of the energy I which represents the ratio between the surface of the sensor s and the imaginary sphere whose center is the source and whose radius is d, *i.e.*, $I = \frac{s}{4\pi d^2}$.

- **Pulse durations of ultrafast lasers** can be interpreted as a plane wave with a Gaussian envelope that localizes the pulse in space and time. Generally, the laser pulse is composed of a narrow bandwidth which oscillates and travels independently, inducing group delay dispersion (GDD) across the operational bandwidth, called chromatic dispersion. This pulse broadening artifacts of an optical component can simply result from the dispersion of its parts, from interference effects or from geometric effects [96] and impacts the physical quantity measurement accuracy.
- **Backscatter of waves in physical space**:
  - **Specular**: the reflection of the beam by the object is said to be specular when the reflecting surface behaves like a mirror. The angle of incidence is identical to the angle of reflection and the incident and reflected rays are contained in planes perpendicular to the reflecting surface thanks to the Snell-Descartes laws. Specular reflections can be used in reflecting telescopes, but for LiDARs near visible light wavelengths, specular reflection is rare, except for reflection from interface with liquid surfaces.
  - **Diffuse**: the reflection is called diffuse when the light ray is absorbed by the surface of the object and a fraction of the ray is retransmitted in all directions of space with an equal amplitude. When the object is perfectly diffusing, it is said to be Lambertian.

    $$L_R = \frac{\rho}{\pi} \times I \tag{2.3}$$

    where ρ is the albedo of the surface, I is the illumination at the point W/m^2, and L_R is the reflected luminance.

Most of the issues in radiometry can be understood using geometrical optics or wave notions, but in this thesis, we also consider the particle behaviour of light based on the quantum theory [97]. The quantum theory of light was proposed by Einstein [98], which states that light travels in bundles of energy, and each bundle is known as a photon. Each photon carries a quantity of energy equal to the product of the frequency of vibration of that photon and Planck's constant cf., Eq. 2.4.

$$E = h\nu \tag{2.4}$$

The study of light in this thesis according to statistics focuses on the supposedly Poissonian regimes statistical distributions [99].

2.2 Pseudo-realistic SPAD - TCSPC simulation model

Once having introduced the properties of the system-level of a TCSPC-based SPAD vision system, the following sections will detail the developed end-to-end simulation model deeply inspired by the work presented in [9]. Firstly, let us consider only direct illumination and neglect the 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

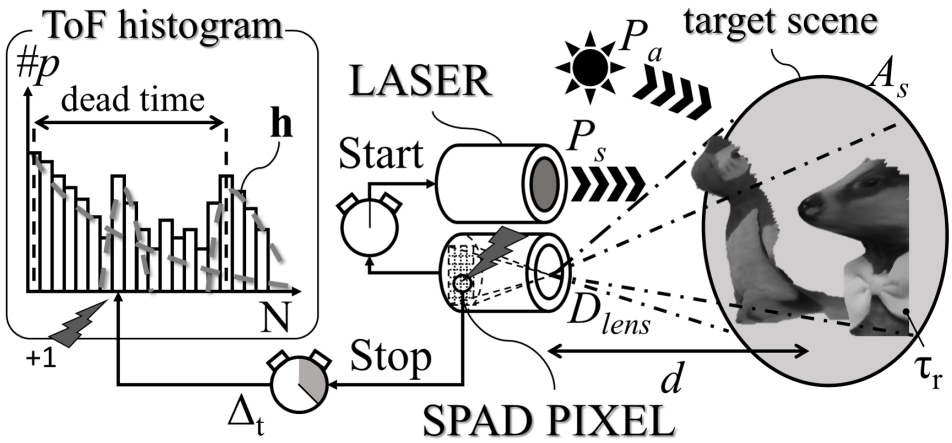


Figure 2.1: SPAD Operation System Overview: direct time-of-flight measurement (D-ToF) of a light pulse reflected by a target using a TCSPC system.

reflection of two light sources, which are the laser source (with power P_s [W]), and the ambient light (with intensity P_a [W/m^2]). For the sake of relevance, several configurations are considered regarding the average ambient light, P_a. For an indoor scene, ambient light illumination is of the order of 5 to 35klux ($P_a < 350\text{W/m}^2$) while being between 50 and 100klux ($P_a < 1050\text{W/m}^2$) for an outdoor scene. Besides, we denote P_{AS} the solar radiation integrated over the laser-illuminated surface,

$$P_{AS} = P_a \cdot A_s = P_a \cdot 4d^2 \cdot \tan\left(\frac{\theta_H}{2}\right) \cdot \tan\left(\frac{\theta_V}{2}\right), \tag{2.5}$$

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 as,

$$P_s^* = P_s \cdot \tau_r \cdot \frac{2}{\pi}\left(\frac{D_{lens}}{2d}\right)^2 \cdot \tau_o \cdot \delta_{bw} \frac{FF \cdot PDP}{\text{H} \cdot \text{W}}. \tag{2.6}$$

And the pixel-wise light power due to ambient light,

$$P_a^* = P_{AS} \cdot \tau_r \cdot \frac{2}{\pi}\left(\frac{D_{lens}}{2d}\right)^2 \cdot \tau_o \cdot \delta_{bw} \frac{FF \cdot PDP}{\text{H} \cdot \text{W}}. \tag{2.7}$$

D_{lens} and τ_o are the optical parameters determining the diameter of the lens and the efficiency of the optics. The integral of the light power knowing the efficiency of the bandpass filter along all wavelengths is approximated by δ_{bw} (where the background illumination (BI) power density is supposed uniform). Note that the laser pulse power reflected to the SPAD is subject to an inverse square law (*i.e.*, inversely proportional to the square of the distance d) whereas BI is independent to this (simplification of d while expanding P_{AS} in Eq. 2.7). 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 [100]. Despite the invariance of the scattering intensity related to the solid angle of the reflecting object (specular reflectance [101] 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. 2.1. For our simulation framework –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,

$$R = \frac{\lambda P^*}{hc}, \tag{2.8}$$

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

For each pixel-wise photon flux, inter-photon waiting times $\{x[i]\}_{i\in\mathbb{N}}$ respect the memoryless property, *i.e.*, the time between two successive photons is exponentially distributed and independent to an arbitrary starting time $\omega \geq 0$ [102],

$$\mathbb{P}(X > \omega + x | X > \omega) = \mathbb{P}(X > x), \tag{2.9}$$

$$\mathbb{P}(X > x) = \exp(-Rx). \tag{2.10}$$

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 performed a Monte Carlo sampling strategy using an exponential random number generator with a rate parameter equal to the photon flux R. Let us denote $\{t[i]\}_{i\in\mathbb{N}}$ the theoretical photon arrival time sequences for a unique light source,

$$t[i] = t[i-1] + x[i]. \tag{2.11}$$

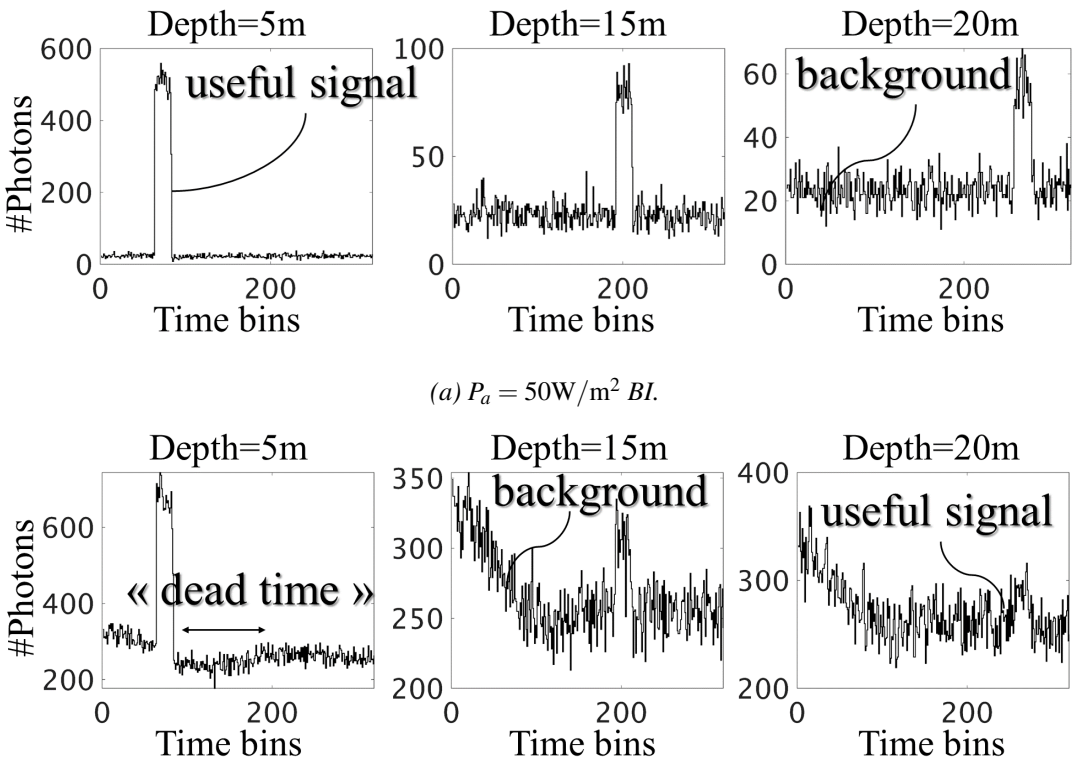
$\{t_a^f[i]\}_{i\in\mathbb{N}}$, $\{t_s^f[i]\}_{i\in\mathbb{N}}$ and $\{t_{dcr}^f[i]\}_{i\in\mathbb{N}}$ are thus representing the sequences of the three main contributors: BI, laser signal and dark count rate (DCR) at inter-frame f. Note that one specificity of SPADs is that they are reverse-biased at a higher voltage than the breakdown voltage (*i.e.*, Geiger Mode [103]), thus a single charge carrier can trigger a self-sustaining avalanche. This operating mode implies a said "Dead Time" (D_t) [104], 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 actually performed by an iterative process sequentially keeping the first arrival time and then removing photons arrivals from the three sequences that are between $t_{SPAD}[i]$ and $t_{SPAD}[i] + D_t$ (see Alg. 1). Consequently, all arrival times in every inter-frame are at least one D_t apart. Note that the number of inter-frames is equal to the ratio between the sensor frame rate and the laser repetition rate, $\lfloor \frac{fps}{Fe} \rfloor$.

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

for $f = 1 : \lfloor \frac{fps}{Fe} \rfloor$ **do**
  while $min(\mathbf{t}_a^f, \mathbf{t}_s^f, \mathbf{t}_{dcr}^f) < \frac{1}{Fe}$ **do**
    $t_{SPAD}[i] \leftarrow min(\mathbf{t}_a^f, \mathbf{t}_s^f, \mathbf{t}_{dcr}^f)$
    $\mathbf{t}_a^f \leftarrow \mathbf{t}_a^f \setminus \{\forall j, t_a^f[j] \leq t_{SPAD}[i] + D_t\}$
    $\mathbf{t}_s^f \leftarrow \mathbf{t}_s^f \setminus \{\forall j, t_s^f[j] \leq t_{SPAD}[i] + D_t\}$
    $\mathbf{t}_{dcr}^f \leftarrow \mathbf{t}_{dcr}^f \setminus \{\forall j, t_{dcr}^f[j] \leq t_{SPAD}[i] + D_t\}$
    $i{+}{+}$
  end
end
$\mathbf{h} \leftarrow hist(\mathbf{t}_{SPAD})$

In addition, other sensor defaults such as afterpulsing [41], or crosstalk [105] are not considered because of their low influence. Fig. 2.2 illustrates simulated SPAD histograms, $\mathbf{h}$, of $\mathbf{t}_{SPAD}$ under typical scene conditions. Contrary to the SPAD model depicted in [12], our proposed model therefore provides more complex data without neglecting D_t that induces histogram distortions (*e.g.*, pile-up effect and rebounds as illustrated in Fig. 2.2). The histograms generated this way

correlate well with the ones reported in the paper [106], especially for a configuration with a wide laser pulse width and a high BI.



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

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

*Figure 2.2: Per-pixel Time-of-Flight histograms with a low BI (a) and a high BI (b), for three different object distances (*5*m,* 15*m and* 20*m). The laser Pulse Width is of* 5ns *while the bin width is of* 260ps *and the SPAD* D_t *equals to* 27ns*.*

2.3 Photons statistics

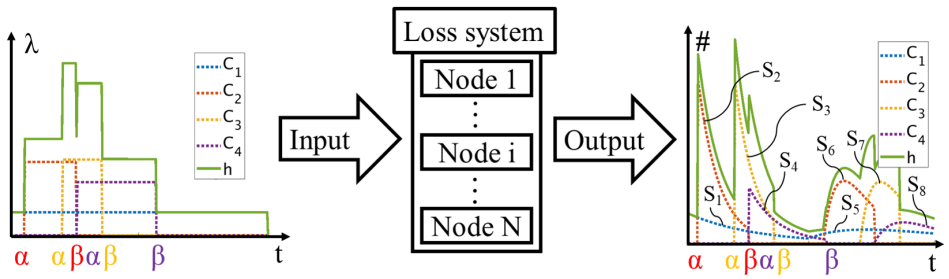


Figure 2.3: Loss system overview.

The main goal in a photon-efficient reconstruction algorithm is to distinguish the different light sources by the corresponding statistical distributions they imply in the photon arrival times

process. This section aims at properly modeling the unknown "stored photons" process. For a deep system-level analysis, the TCSPC SPAD-based imaging system is considered as a queueing loss model [107]. Where photon arrivals are the customers, generated from the Poisson process property of photons emitted by a light source [108]; SPAD sensors refer to the servers with a constant service process time "Dead Time"; and "stored photons" (in a TOF histogram) refer to the departure process. Due to the multiplicity of light sources, as well as their intermittence, the photon arrival times process will instead be considered as being constructed from a Piecewise Constant Poisson Process (PCPP) composed of C Poisson intensity levels (*i.e.*, C intermittent light sources as illustrated in Fig. 2.3). Furthermore, the "stored photons" can be interpreted in two ways. That is to say, when the "stored photons" come from several SPAD sensors set up in parallel periodically as illustrated by nodes in Fig. 2.3. Otherwise, when the "stored photons" come from a single SPAD sensor –in the considered case where TCSPC SPAD-based imaging system is a periodic photon counting sequence measurement (cf., Fig. 4.3)–.

Mathematically speaking, we seek to prove that a PCPP of C Poisson intensity levels can be decomposed into C Poisson sub-processes with various finite time supports [109] (*i.e.*, the independence and identical distribution, *i.i.d.*, of enumerated arrivals $M(x)$ over a period of time $x > 0$ as written in Eq. 2.12). The joint Probability Mass Function (PMF) Poisson counting sub-processes conditioned on the number of arrivals of the Poisson process $M(x) = \mathscr{M}$ can be expressed in two ways as depicted in Eq. 2.12. It can be either noted as a multinomial distribution or using the conditional Bayes Rule notation. Note that each Poisson counting sub-process indices c, $M_c(x) = \mathscr{M}_c$, has its own probability denoted $\pi_c = \frac{\mathscr{M}_c}{\mathscr{M}}$.

$$\begin{aligned}
&\mathbb{P}\left(M_1(x) = \mathscr{M}_1, ..., M_C(x) = \mathscr{M}_C | M(x) = \mathscr{M}\right) \\
&= \frac{\mathscr{M}!}{\prod_{c=1}^{C} \mathscr{M}_c!} \prod_{c=1}^{C} \pi_c^{\mathscr{M}_c} \\
&= \frac{\mathbb{P}\left(M_1(x) = \mathscr{M}_1, ..., M_C(x) = \mathscr{M}_C\right)}{\mathbb{P}\left(M(x) = \mathscr{M}\right)}
\end{aligned} \tag{2.12}$$

Then, combining these two formulations while moving up the original Poisson process PMF term leads to Eq. 2.13.

$$\mathbb{P}\left(M(x) = \mathscr{M}\right) = \frac{(\lambda x)^{\mathscr{M}} \, \mathrm{e}^{(-\lambda x)}}{\mathscr{M}!} \tag{2.13}$$

It finally shows that the joint PMF $\forall c,\ M_c(x)$ can be defined as the product of all subdivided Poisson processes PMF with corresponding intensity parameter $\lambda_c(x) = \pi_c \lambda(x)$ as detailed in Eq. 2.14.

$$\begin{aligned}
&\mathbb{P}\left(M_1(x) = \mathscr{M}_1, ..., M_C(x) = \mathscr{M}_C\right) \\
&= \frac{(\lambda x)^{\mathscr{M}} \, \mathrm{e}^{-(\lambda x)}}{\mathscr{M}!} \frac{\mathscr{M}!}{\prod_{c=1}^{C} \mathscr{M}_c!} \prod_{c=1}^{C} \pi_c^{\mathscr{M}_c} \\
&= \frac{(\lambda x)^{\sum_{c=1}^{C} \mathscr{M}_c} \, \mathrm{e}^{-\left(\sum_{c=1}^{C} \pi_c \lambda x\right)}}{\mathscr{M}!} \frac{\mathscr{M}!}{\prod_{c=1}^{C} \mathscr{M}_c!} \prod_{c=1}^{C} \pi_c^{\mathscr{M}_c} \\
&= \frac{\prod_{c=1}^{C} (\lambda x)^{\mathscr{M}_c} \, \mathrm{e}^{-(\pi_c \lambda x)}}{\prod_{c=1}^{C} \mathscr{M}_c!} \prod_{c=1}^{C} \pi_c^{\mathscr{M}_c} \\
&= \prod_{c=1}^{C} \mathbb{P}\left(M_c(x) = \mathscr{M}_c\right)
\end{aligned} \tag{2.14}$$

Based on the Poisson process stationary and independent increments property [110] and the above demonstration of superposition property of the independent Poisson (sub)processes [111, 112], we can state that the "stored photons" process distribution is, just as the photon arrival times process (input of the loss system illustrated in Fig. 2.3), a piecewise uniform temporal distribution of the event. Unfortunately, the servers of the equivalent queueing loss system involve Poisson process time dependence due to a constant service process –*i.e.*, the SPAD "Dead Time" denoted τ–. This so-called "Dead Time" actually corresponds to the time duration following an avalanche during which the SPAD device is unable to detect another photon. Therefore, the "stored photons" must be considered using the generalized problem of the Poisson process characterization in terms of a time-correlated counting process with the phase-type distribution (PH) which results from the convolution of the probability densities of the events [113] (output of the loss system illustrated in Fig. 2.3). For example, the probability distribution of the sum of two or more independent random variables can be viewed as the convolution of their individual distributions. The Erlang-k Probability Distribution Function (PDF) is the one that properly models the sum of k independent random exponential inter-arrival times *i.e.*, $Z_k = X_1 + X_2 + ... + X_k$ in the typical case of a system of one or more inter-related Poisson processes occurring in phases [114], as written in Eq. 2.15.

$$\begin{aligned}\mathbb{P}(Z_k = z_k) &= \mathbb{P}\left[\sum_{j=1}^{k} X_j\right] \\ &= \int_0^{z_k}\int_0^{x_1} ... \int_0^{x_{k-1}} \lambda\, \mathrm{e}^{-\lambda(z_k - x_1)} \\ &\quad ...\ \lambda\, \mathrm{e}^{-\lambda(x_{k-1}-x_k)}\ \lambda\, \mathrm{e}^{-\lambda(x_k)}\, dx_k\, ...\, dx_1 \\ &= \frac{\lambda^k z_k^{k-1}\, \mathrm{e}^{-\lambda z_k}}{(k-1)\,!}\end{aligned} \tag{2.15}$$

Without photon detection loss, the time distribution of events with this notation still converges to a uniform distribution by summing the Erlang-k phases from $k = 0$ to $k \to \infty$. However, because of the service time τ, the queuing loss system considered here implies that events occurring during the service time (*i.e.*, generated by a previous event z_k *i.e.*, $\forall\ k \in \{1, ..., N\}$, $[z_k; z_k + \tau]$) are less likely to be observed by the system due to its congestion. Since then, "stored photons" shows a time shift τ between two sequential photon arrival times, z_k and z_{k+1}. The Erlang-k PDFs are consequently shifted by $(k-1)\tau$ revising the Eq. 2.15 to the Eq. 2.16 and implying a non-uniform time distribution of events if summing the Erlang-k phases from $k = 0$ to $k \to \infty$.

$$\begin{aligned}\mathbb{P}(Z_k = z_k) &= \mathbb{P}\left[\sum_{j=1}^{k} \left((k-1)\tau + X_j\right)\right] \\ &= \int_0^{z_k-(k-1)\tau} \int_\tau^{x_1+\tau} ... \int_{(k-1)\tau}^{x_{k-1}+\tau} \lambda\, \mathrm{e}^{-\lambda(z_k - x_1 - (k-1)\tau)} \\ &\quad ...\ \lambda\, \mathrm{e}^{-\lambda(x_{k-1}-x_k+\tau)}\ \lambda\, \mathrm{e}^{-\lambda(x_k-(k-1)\tau)}\, dx_k\, ...\, dx_1 \\ &= \frac{\lambda^k (z_k - \alpha)^{k-1}\, \mathrm{e}^{-\lambda(z_k-\alpha)}}{(k-1)\,!},\ \text{where } \alpha = (k-1)\tau\end{aligned} \tag{2.16}$$

An additional parameter β refers to the ending time of the server measurements or of the input Poisson process. Indeed, the superposition of Poisson sub-processes does not occur in phase since the laser pulse light source admits a time dependent starting and ending points respectively denoted α and β, which yields to an additional Truncated-Shifted property of the Erlang-k ($\mathscr{E}$).

2.4 Conclusion

This chapter presents an end-to-end simulation model in order to generate more physically-plausible SPAD data and to better control the SPAD parameters having an impact on the shape of the measured histogram. The resulting measured photon statistics of such an imaging system has been clearly modeled as a Truncated Shifted Hyper Erlang Distribution $\mathscr{E}$. This allows us to carefully select the detector and transmitter parameters to provide specific shape of histograms (cf. Fig. 2.4). Such as the choice of the laser pulse width in terms of histogram bins that match with the application requirements and the combined reconstruction method (cf., Tab. 2.1). The asynchronous acquisition method from [72] has also been carefully exploited in this thesis allowing to mitigate the pile-up effect and to consider the uniform distribution of photons coming from background illumination with the Gaussian distribution of photons coming from laser pulse illumination. Once the "stored photons" process distribution is properly defined, the remainder of this manuscript presents various solutions for the estimation of intensity-depth maps.

	Parameters	Chap. 4	Chap. 5	Chap. 6	Chap. 7.2	Chap. 7.3
Laser	Max range (m)	25	2 - 20	50	3	25
	Repetition rate, Fe (MHz)	12	150 - 15	6	12	12
	Wavelength, λ (nm)	780	671	671	780	780
	Pulse Width, PW (ns)	5	0.1	0.1	1	5
Environnent	Background light, P_a (W/m^2)	50 - 1050	50 - 1050	50 - 1050	50 - 1050	50 - 1050
	Object reflectivity, τ_r	1/256 - 1	1/256 - 1	1/256 - 1	1/256 - 1	1/256 - 1
SPAD	Photon Detection Probability, PDP	0.4	0.4	0.4	0.4	0.4
	Fill Factor, FF	0.7	0.7	0.7	0.7	0.7
	Dead Time, D_t (ns)	27	10	10	27	27
	TDC dynamic range, N	1-320	1-1024	1-1024	1-256	1-320
	TDC Resolution, $(N \times Fe)^{-1}$ (ps)	260	10	260	78	6 - 60
	Synchronous	Yes	No	No	No	yes
	Photon counts	$10^3 - 10^4$	10^4	$10^3 - 10^4$	$10^3 - 10^4$	100 - 10^3
Related TCSPC distribution modeling		Mixture of Truncated - Shifted Erlang distributions	Mixture of Gaussian - Uniform distributions	Sparse	Sparse/ Mixture of Gaussian - Uniform distributions	Mixture of Truncated - Shifted Erlang distributions

Tableau 2.1: This table reports all simulation parameters in this manuscript.

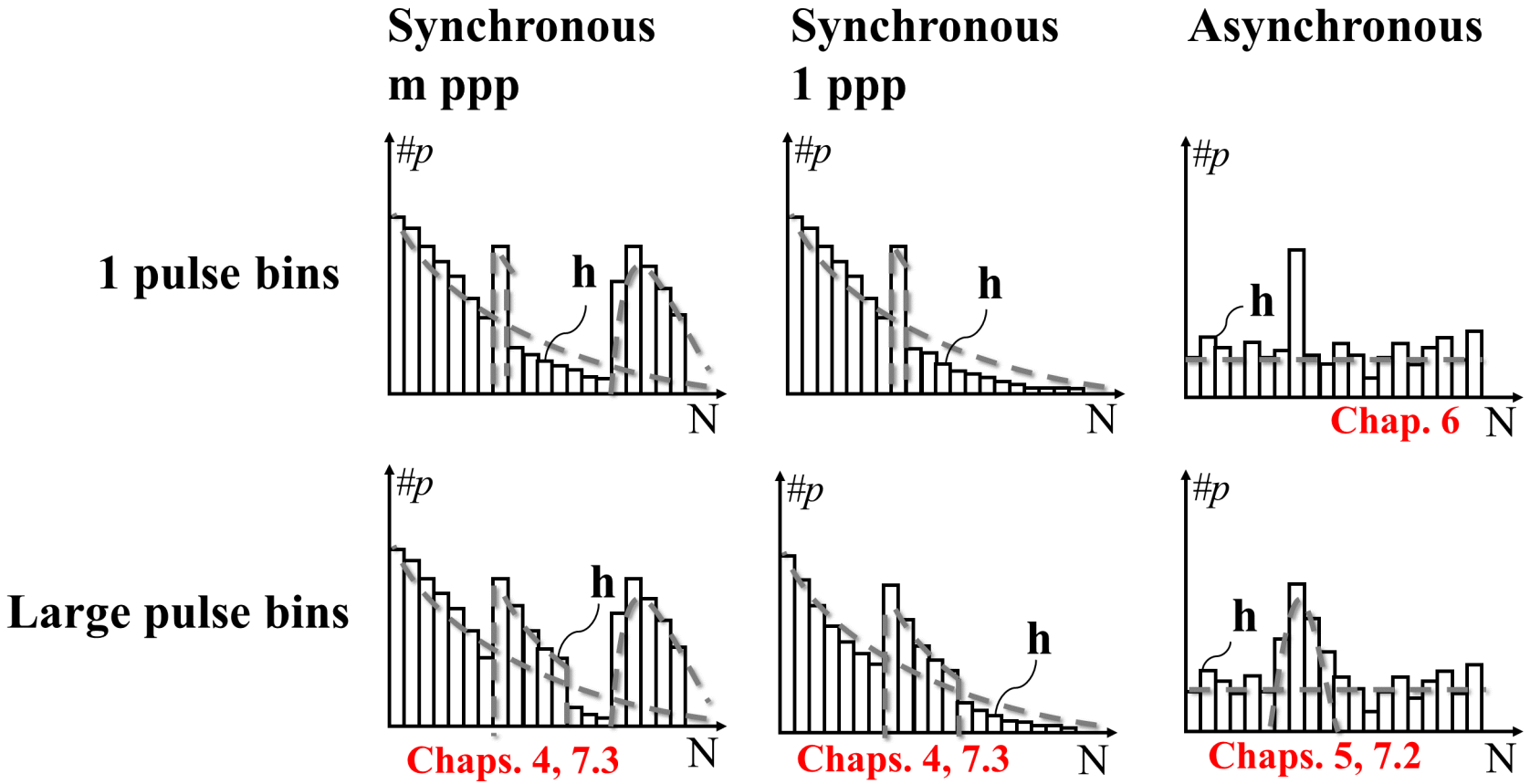


Figure 2.4: TCSPC distribution modeling assumptions throughout this manuscript.

3

Signal processing background

This chapter provides an overview of the signal processing core techniques involved in the work of this thesis. The chapter is divided into four sections. The first section deals with the Expectation-Maximization algorithmic approaches with a short introduction to clarify the definition of this method. This section will also present extensions, such as its stochastic approximation in order to reduce the computational cost of the E-step. The additional use of imputation method to obtain an unbiased convergence and the online version to reduce the EM algorithmic complexity will be also introduced. The second section discusses the theoretical background of compressive sensing (CS) and its related reconstruction algorithm. Only the basics will be presented for a better understanding of the underlying tools used in the following chapters. The third part of this chapter presents the deep learning field for image reconstruction, which appears to be a very powerful algorithmic tool. This section will briefly introduce the mathematical background behind the success of deep learning and will finish with the presentation of the main deep learning topology, the convolutional neural networks (CNN). Finally, the last section presents various remote processing techniques that increase the resolution of an imaging system.

Contents

3.1 Expectation-Maximization algorithmic background

3.1.1 General Statement of the EM Algorithm

Most of the real word data observations are the function of underlying parameters. For example, the phenotype as observed data is a function of the underlying unobserved gene pattern (the genotype) [115]. In economics, quality of life is a latent variable that is inferred from observed variables such as wealth, employment, environment, physical and mental health. Let us therefore consider the observed data, x, as a function of a statistical distribution model, $f(x;\theta)$, composed of one or more distribution with latent parameters. In this regard, the EM algorithm is a partially non-Bayesian maximum likelihood method. Its purpose is to find the (local) maximum likelihood parameters, θ, of this statistical distribution model, $f(x;\theta)$, namely mixture of distribution. But it cannot be solved directly since we do not know sample responsibility ($\Delta_{i,k}$ in Eq. 3.2). In fact, the idea behind the EM algorithm is that, even though we do not know x, knowledge of the underlying distribution $f(x;\theta)$ can be used to determine an estimate for the parameters θ. This is done by first estimating the underlying data, then using these data to update our estimate of the parameter. This is repeated until convergence. A simple EM algorithm task would be the maximum likelihood estimation (MLE) of the parameter, θ, of a signal composed of only one distribution, like the mean, μ, and the standard deviation, σ, of a Gaussian distribution only using the sample and under statistical priors. In order to generalize the problem, the mixture distribution is, mathematically speaking, defined as a weighted sum of K distributions:

$$f(x;\Theta) = \sum_{k=1}^{K} w_k g_k(x;\theta_k) \tag{3.1}$$

where g_k denotes the density function, w_k the value of the associated weight and θ_k its parameters. To introduce the method, let first consider we have the prior knowledge about the distribution from which a sample with size n, *i.e.*, $x_1,....x_n$, has been independently randomly drawn without prior knowledge of the distribution parameters. Because, the random variables are considered independent and identically distributed (i.i.d.) *i.e.*, $x_i \overset{iid}{\sim} f_X(x_i;\theta)$ and considering the Bayes Rule, the likelihood function can be expressed as written in Eq. 3.2.

$$\begin{aligned} L(\Theta|x_1,...,x_N) &= \frac{f_X(x_1,...x_N|\Theta)w(\Theta)}{f_X(x_1,...,x_N)} \\ &= f(x_1,...x_N;\Theta) \\ &= \Pi_{i=1}^{N} f(x_i;\Theta) \\ &= \Pi_{i=1}^{N} \sum_{k=1}^{K} \Delta_{i,k} w_k g_k(x_i;\theta_k) \end{aligned} \tag{3.2}$$

For convenience, in most cases, the log-likelihood is used instead of the likelihood as written in Eq. 3.3 where $\Delta_{i,k}$ is equal to zero $\forall\ i \in \mathbb{R}^N$ except for only one component (*i.e.*, for one value of k, $\Delta_{i,k}$ is equal to one).

$$\begin{aligned} l(\Theta|x_1,...,x_N) = \log(L(\Theta|x_1,...,x_N)) &= \sum_{i=1}^{n} \log f(x_i,\Theta) \\ &= \sum_{i=1}^{N} \log \sum_{k=1}^{K} \Delta_{i,k} w_k g_k(x_i;\theta_k) \\ &= \sum_{i=1}^{N} \sum_{k=1}^{K} \Delta_{i,k} \log w_k g_k(x_i;\theta_k) \end{aligned} \tag{3.3}$$

As mentioned above, the data can not be fully observable and the MLE approach can thus not be directly applied hence the use of the EM algorithm proposed by Arthur Dempster [116]. Therefore, the "objective function" noted $Q(\Theta|x_1,...,x_N)$ of the EM algorithm is restated (cf., Eq. 3.4) with sample responsibility $\Delta_{i,k}$.

$$Q(\Theta|x_1,...,x_N) = \sum_{i=1}^{N}\sum_{k=1}^{K} \mathbb{E}[\Delta_{i,k}|X,\Theta] \log w_k g_k(x_i;\theta_k) \tag{3.4}$$

The EM algorithm is an iterative algorithm with two major steps: an Expectation step estimating sample responsibility ($\Delta_{i,k}$), followed by a Maximization step aiming at finding the best parameters maximizing the objective function with the updated value of the responsibility. According to the Bayes rule [117], we can then define the E step as follows:

$$\mathbb{E}[\Delta_i|X,\Theta] = \frac{\hat{w}_k g_k(x_i;\hat{\theta}_k)}{\sum_{k=1}^{K} \hat{w}_k g_k(x_i;\hat{\theta}_k)} \tag{3.5}$$

The M step then aims to find parameter Θ which maximizes $Q(\Theta|x_1,...,x_N)$ cf., Eq. 3.6.

$$\hat{\Theta} = \arg\max_{\Theta} Q(\Theta|x_1,...,x_N) \tag{3.6}$$

To do so, the ML estimate of the parameter is obtained by taking the derivative of $Q(\Theta|x_1,...,x_N)$ with respect to Θ, equal to zero, and solving it for Θ.

$$\nabla_{\Theta} Q(\hat{\Theta}|x_1,...,x_N) = 0 \tag{3.7}$$

where,

$$\nabla_{\Theta} = \begin{bmatrix} \dfrac{\partial}{\partial\theta_1} \\ \dfrac{\partial}{\partial\theta_2} \\ \vdots \\ \dfrac{\partial}{\partial\theta_r} \end{bmatrix} \tag{3.8}$$

This yields to the new parameter estimate Θ_{k+1} where k refers to the algorithm iteration index. These two steps (Expectation and Maximization step) are repeated until convergence.

3.1.2 Online EM Algorithms variants

In the case of large datasets, the standard EM algorithm is very expensive in time and memory computing due to the need to have the entire dataset available at each iteration of the algorithm. The scientific community has advanced the online variants of the EM algorithm making possible the estimation of the latent distribution model parameters without the need to store all the data. The online estimation means that the data must be run through only once, making the iteration index value identical to the observation index i and strongly requiring that observations data are independent and identically distributed (i.i.d.) for its proper functioning. For a clear understanding of the variants, let firstly introduce the Lange work late 90's [118], who proposed to replace the M step by a Newton update based on some regularity condition, *i.e.*, θ_k musts to be twice differentiable. This Newton update modification results in the following recursion of the maximization step:

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \gamma_{k+1}[J(Y;\hat{\theta}_k)]^{-1} \sum_{i=1}^{N} \mathbb{E}_{\hat{\theta}_k}[\nabla_{\theta} \log f(x_i;\hat{\theta}_k)|Y_i] \tag{3.9}$$

where γ_{k+1} refers to the step size update and $J(Y;\hat{\theta}_k)$ refers to the negative of the second derivative of the log-likelihood function *i.e.*, the Hessian matrix. Note that if $\gamma_{k+1}=1$, Eq. 3.9 is locally equivalent to the EM convergence algorithm. In fact, γ_{k+1} is rather selected for linear searches to ensure that the likelihood is actually increased at each iteration. Moreover, to avoid numerical problems because $J(Y;\hat{\theta}_k)$ is not necessarily a positive definite matrix, a well known modification of the Newton recursion consists in replacing $J(Y;\hat{\theta}_k)$ by the Fisher Information Matrix associated to a complete observation cf., Eq. 3.10.

$$I(\hat{\theta}_k) \overset{def}{=} -\mathbb{E}_{\hat{\theta}_k}[\nabla^2_{\theta} \log f(X;\hat{\theta}_k)] \tag{3.10}$$

Based on Eq. 3.9, the first online parameter estimation process was finally proposed by Titterington (1984) [119] with the use of stochastic approximation version of the modified gradient recursion:

$$\hat{\theta}_{i+1} = \hat{\theta}_i + \gamma_{i+1} I^{-1}(\hat{\theta}_i) \nabla_{\theta} \log g(x_{i+1};\hat{\theta}_i) \tag{3.11}$$

where γ_{i+1} is a decreasing sequence of positive step sizes. The use of the online EM algorithm can be recommended in the case of large sample sizes with the possibility of batch training [119].

3.1.3 Stochastic & Classified EM algorithm variants

A second variant of the EM algorithm is the Classified EM (CEM) algorithm corresponding to an iterative clustering algorithm that computes the maximum likelihood estimates of θ_k and simultaneously finds the classification. This algorithm has been designed to overcome one of the fundamental limitations of the standard EM algorithm, which is its slow convergence. This extension shows a close similarity with the K-means algorithm since, the EM algorithm makes a "soft" assignment based on the posterior probabilities, the K-means algorithm performs a "hard" assignment of data points to clusters (as a CEM). In fact, the CEM can be regarded as a canonical EM algorithm with a classification step between the E-step and the M-step using a maximum a posteriori principle [120]. By doing this, we obtain a hard assignment of data points to clusters, just as in the K-means, so that the responsibility in Eq. 3.5 is equal to 1 or 0. Each data point is thereby assigned to the Gaussian having the closest mean. The canonical EM re-estimation Eq. 3.7 thus reduces to simpler operation, such as in the particular context of Gaussian mixtures the mean parameter is simply equal to the mean of all of the data points and not a weighted sum of the data points. From this, we can derive an online CEM algorithm by applying the Robbins-Monro procedure [121] to the problem of finding roots of the regression function 3.12 given by setting its derivative with respect to the mean parameter to zero.

$$J = \sum_{i=1}^{N} \sum_{k=1}^{K} r_{i,k} \|x_i - \mu_k\|^2 \tag{3.12}$$

This leads to a second online update function for each data point x_i:

$$\mu_k^{new} = \mu_k^{old} + \eta_i \left(x_i - \mu_k^{old} \right) \tag{3.13}$$

where η is the learning rate parameter, which can decrease monotonically, logarithmically and even stochastically as more as data points are considered.

On the other hand, the Stochastic version of the EM algorithm (SEM) has been proposed by Celeux and Diebolt late 80's [122] to address EM limitations related to its strong dependence on

the initialization and its convergence to saddle-points. The SEM algorithm incorporates a stochastic step between the E-step and the M-step of the EM algorithm, which consists in randomly assigning to each x_i a posterior probability. This refers, in a way, to the principle of random imputation since it generates a complete sample by drawing it randomly from the posterior distribution.

3.1.4 Implementations of the EM algorithm

For the design of high-performance, low-power EM algorithm circuits and systems, approximate computing has emerged as a new paradigm. One of the main challenges of approximate computing is to ensure quality of approximate computations [123]. A variety of approximation strategies has been identified as relevant in the literature, such as precision scaling [124], voltage scaling [125], operations approximation [126, 127], memoization [128], loop perforation [129], skipping operation [130], but only some of them will be presented as follows:

- **Precision scaling** , as its name suggests, refers to the accurate selection of the data word length (*i.e.*, input or intermediate values) in order to reduce the memory and computing requirements.
- **Operation approximation** is based on hardware-friendly operation like logarithmic approximation [131] leading to piecewise linear approximation. For example, a floating-point division can be approximately computed by a subtraction operation. Which can be then refined with an error compensation lookup table.
- **Memoization** is an optimization technique which consists in storing the results of computationally intensive functions for multiple inputs and returning the cached result for the corresponding input. The interest of this technique can be even more important at the cost of a possible approximation of the calculations if the results are reused several times for a similar combination of functions and inputs. The input/output matching matrix, called a lookup table, is interesting for its speed, but depending on how many input/output matching are stored, it can require a huge amount of memory.

After presenting the basics of the EM algorithm on which two of our contributions are based in Chap. 4 and Chap. 5, we present in the next section the theoretical background of compressive sensing (CS).

3.2 Compressive Sensing Background

3.2.1 Theory of compressed data acquisition

In the world of the Internet of Things, recent technology devices face an increasing amount of information coming from various type of sensors. These high quantity of information need to be primarily sensed, transmitted and processed, requiring power, data processing, large bandwidths for data communications and large storage memory systems. Traditional sensing/sampling methods used were based on the Nyquist-Shannon sampling theory [132] which establishes that, for a continuous-time signal of finite bandwidth (*i.e.*, zero outside of a finite region of frequencies with the possibility to use an antialiasing low-pass filter to "bandlimit" the signal not naturally bandlimited) the sampling rate must be at least twice as large as the bandwidth. For example, in the field of data conversion, standard analog-to-digital converter (ADC) [133] are sampling signal inputs at or above the Nyquist sampling rate. This signal processing pipeline (*i.e.*, detection, transmission, processing) encounters an unavoidable bottleneck for storage and detection due to

the high dimensionality of the data. Therefore, the Compressive Sensing (CS) emerged as a powerful hardware-friendly framework for sensing and compressing data simultaneously and completely rethought the signal processing pipeline. The CS theory [134] states that sparse signals (cf., $\mathbf{x} \in \mathbb{R}^n$ in Eq. 3.14) can be identically recovered from a small set of compressed measurements collected (cf., $\mathbf{y} \in \mathbb{R}^m$ in Eq. 3.14, such that $m << n$) based on non-adaptative linear projections, $\mathbf{\Phi} \in \mathbb{R}^{m \times n}$ as written in Eq. 3.14.

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} \tag{3.14}$$

This signal-independent dimensionality reduction raises important questions about the possibility of existing methods to approximate $\mathbf{x}$ from compression measurements $\mathbf{y}$. Admittedly, the process of recovering $\mathbf{x}$ from $\mathbf{y}$ is ill-posed in general because $\mathbf{\Phi}$ becomes an overcomplete dictionary instead of a basis. In fact, thanks to the CS, signal recovery is possible for a wider class of signal models. To make this possible, CS theory takes advantage of three fundamental premises, namely the sparsity of the signal, the incoherent sampling, and the restricted isometry property (RIP).

Sparsity CS relies on the small impact of small coefficients in a "sparse domain", after sparsification of the signal thanks to a transform operation (*e.g.*, Discrete Fourier Transform (DFT) [135], Discrete Cosine Transform (DCT) [136] or Discrete Wavelet Transform (DWT) [137]), and thus the possibility to remove them without major loss of information. In the strict sense of the word, a vector $\mathbf{s}$ is said k-sparse in an orthonormal basis $\Psi \in \mathbb{R}^{n \times n}$ if $\|\mathbf{s}\|_0 = k << n$ (*i.e.*, all except k entries of $\mathbf{s}$ are equal to 0). Mathematically speaking, $\mathbf{s}$ is a sparse vector expanded in an orthonormal basis, $\mathbf{\Psi}$, as follows:

$$\mathbf{x} = \mathbf{\Psi}\mathbf{s} \tag{3.15}$$

In the literature, $\mathbf{s}$ is commonly considered "almost sparse" (in other words, "compressible") in the sense that high magnitudes of $\mathbf{s}$ are few compared to others. That is to say, when considering $\mathbf{s}_k$ resulting from the inverse projection in an orthonormal basis, $\mathbf{\Psi}$, of $\mathbf{x}$ only with the k largest values of $\mathbf{s}$ and the rest set to 0. We can assume that a large fraction of the coefficients of $\mathbf{s}$ can thus be removed, and that $\mathbf{s}$ is compressible if $\|\mathbf{s} - \mathbf{s}_k\|_2 \approx 0$ with $\|\mathbf{s}_k\|_0 = k << n$.

Incoherence The second CS fundamental is the coherence of the $\mathbf{\Psi}$ and $\mathbf{\Phi}$ matrices. For the proper functioning of CS, a low coherence is requested. The coherence between the sensing matrix $\mathbf{\Phi}$ and the "sparsifying matrix" $\mathbf{\Psi}$ is described in Eq. 3.16 and measures the largest cross-correlation between any two elements of $\mathbf{\Phi}$ and $\mathbf{\Psi}$.

$$\mu(\mathbf{\Phi}, \mathbf{\Psi}) = \sqrt{n} \cdot \max_{1 \leq k,j \leq n} \|\mathbf{\Phi}_k, \mathbf{\Psi}_j\| \tag{3.16}$$

To comply with this property, random matrices are commonly used since they are largely incoherent with any fixed basis $\mathbf{\Psi}$ [138]. Indeed, by selecting the linear mapping $\mathbf{\Phi}$ randomly, we can asset with high probability that the coherence between $\mathbf{\Phi}$ and $\mathbf{\Psi}$ is about $\sqrt{2\log(n)}$ [134]. For more details and explanation of incoherence property please refer to [139].

Restricted Isometry Property Another key notion to evaluate the robustness of a sensing matrix and ensure the success of signal recovery in the context of CS has been introduced by [140], namely Restricted Isometry Property (RIP).

$$(1 - \delta_S)\|\mathbf{s}\|_2^2 \leq \|\mathbf{\Phi}\mathbf{s}\|_2^2 \leq (1 + \delta_S)\|\mathbf{s}\|_2^2 \tag{3.17}$$

The matrix $\mathbf{\Phi}$ obeys the RIP of order K (*i.e.*, for all K-sparse vectors $\mathbf{s}$) if there exists a $\delta_K \in [0,1[$ for the Eq. 3.17. The RIP is all the more respected if δ_k is close to 0 which is desired. The RIP plays a major role to deal with noisy measurements and guarantee a stable embedding [134]. Even though some fundamentals premises allow to facilitate signal recovery *e.g.*, by neglecting the impact of noise on the quality recovery, the CS based systems can be limited by the processing complexity related to the signal recovery compared to the classical approach (cf., Subsec. 6.1.3) as well as the generation of the sensing matrix at the sensor level (cf., Subsec. 3.2.3).

3.2.2 Recovery Algorithms

In the context of CS, sparsity is exploited as a prior knowledge in order to overcome the ill-posed problem when $k < m << n$ and recover the original signal $\mathbf{x}$ from the compressed signal $\mathbf{y}$. The optimization problem consists in finding the sparsest signal $\hat{\mathbf{s}}$ in the sparsity basis $\mathbf{\Psi}$ as written in Eq. 3.18:

$$\mathbf{x}_r = \arg\min_{\mathbf{x}} \|\mathbf{\Psi}^T \mathbf{x}\|_0 \;\; w.r.t. \; \mathbf{y} = \mathbf{\Phi}\mathbf{x} \tag{3.18}$$

Several greedy methods solve this optimization problem in a step-by-step fashion such as Orthogonal Matching Pursuit (OMP) [141]. These signal recovery methods are advantageously combined with a threshold variant for a custom selection of the sparsity level of the estimated signal cf., Eq. 3.19.

$$\mathbf{x}_r = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2 \;\; w.r.t. \; \|\mathbf{\Psi}^T \mathbf{x}\|_0 \leq \hat{K} \tag{3.19}$$

Even though greedy algorithms are simple to implement and highly consistent with the theoretical prerequisites for sparse signal recovery, the drawback of this optimization problem is that it is NP-hard. An alternative solution is to approximate the l_0 minimization problem by an other mathematical optimization problem of the form:

$$\mathbf{x}_r = \arg\min_{\mathbf{x}} \|\mathbf{\Psi}^T \mathbf{x}\|_q \;\; w.r.t. \; \mathbf{y} = \mathbf{\Phi}\mathbf{x} \tag{3.20}$$

since $\|\mathbf{x}\|_q$ approaches $\|\mathbf{x}\|_0$ as $q \to 0$. If $q = 1$, this relaxation allows to solve a $l1$ minimization problem which is more easily manageable, namely using the Basis Pursuit (BP) algorithm [142]. However, the current limitation of this approach depends on a certain condition expressed by the sparsity level. If we assume that the coefficient sequence $\mathbf{s}$ of $\mathbf{x}$ in the basis $\mathbf{\Psi}$ is S-sparse, then the recovery via $l1$ minimization is in a proven way exact. Otherwise, considering the noisy data (where z is an error term):

$$\mathbf{y} = \mathbf{\Phi}\mathbf{\Psi}\mathbf{x} + \mathbf{z}, \tag{3.21}$$

it's preferred to use the $l1$ minimization with relaxed constraints for reconstruction as written in Eq. 3.22. This regression analysis method – namely least absolute shrinkage and selection operator (lasso) – was firstly proposed by [143] and consists in adding a penalized least squares method with an $l1$ penalty on the regression coefficients.

$$\mathbf{x}_r = \arg\min_{\mathbf{x}} \|\mathbf{\Psi}^T \mathbf{x}\|_1 \;\; w.r.t. \; \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2 \leq \varepsilon \tag{3.22}$$

This mathematical optimization problem can be equivalently expressed by using its augmented Lagrangian [144] form:

$$\mathbf{x}_r = \arg\min_{\mathbf{x}} \|\mathbf{\Psi}^T x\|_1 + \lambda \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2. \tag{3.23}$$

where λ is an inner regularization parameter controlling the importance of the fidelity term, in other words the trade-off between the sparsity level and the noise level. The least absolute

shrinkage and selection operator (lasso) has shown success in many applications and through a variety of developed algorithms from convex analysis and optimization theory, such as coordinate descent [145], subgradient methods [146], least-angle regression (LARS) [147] and proximal gradient methods [148]. However, the lasso has encountered limitations in some scenarios, hence the recent development of extensions. These extensions or alternative methods can be represented generally by two cost functions. One corresponds to the fidelity term $FT(\mathbf{x},\mathbf{y})$ and the second to the regularization operator $Op(\mathbf{x})$ as written in Eq. 3.24 where the $Op(\mathbf{x})$ function can describe ridge, spatial collaboration or many other possible regularization operators as presented in this survey of $l1$ regression [149].

$$\mathbf{x}_r = \arg\min_{\mathbf{x}} \; Op(\mathbf{x}) + FT(\mathbf{x},\mathbf{y}). \tag{3.24}$$

For example, alternatively to the lasso, the Dantzig selector [150] also aims at recovering the true sparsity pattern but under various fidelity term. Indeed, the main difference is that the Dantzig selector regularizes the sum of absolute coefficients $\|\mathbf{\Psi}^T x\|_1$ with $\|\mathbf{\Phi}^T(\mathbf{y}-\mathbf{\Phi}\mathbf{x})\|_\infty$ whereas the lasso fidelity term regularization $FT(\mathbf{x},\mathbf{y})$ is the residual sum of squares cf., Eq. 3.23. The Dantzig selector shows that it is more robust against measurement errors, especially quantization errors, thereby making it well suited for some practical applications.

In addition, note that the lasso gives a sparse solution with the use of one $l1$ normalization penalty term [151] (the solution saturates). To avoid solution saturation in the presence of predictors correlation, the Elastic-Net [152] can include more predictors into the model by relaxing the regularization operator $Op(\mathbf{x})$ with a ridge penalty term as written in Eq. 3.25 added to the standard lasso optimization problem.

$$\mathbf{x}_r = \arg\min_{\mathbf{x}} \|\mathbf{\Psi}^T\mathbf{x}\|_1 + \lambda_1\|\mathbf{y}-\mathbf{\Phi}\mathbf{x}\|_2^2 + \lambda_2\|\mathbf{\Psi}^T\mathbf{x}\|_2^2. \tag{3.25}$$

By going even further, for particular optimization problems namely grouping effect, which consists in including all the highly correlated variables in groups as presented in [152], custom algorithms have been developed as surveyed in [149] such as group lasso [153] or composite absolute penalties [154].

3.2.3 Implementations of compressive sensing for imaging

CS implementation is a non negligible point of study since its interest is to limit the acquisition of high dimensional data. The concept from a high level perspective of CS methods is to be able to reconstruct image from randomized measurements, which reduces the dimensionality of the data. This is true even though the reconstruction process involves time-consuming iterative optimization. Therefore, hardware implementations of CS for imaging typically consist of randomly activating pixels within a group and their weighted sum within that same group. Several ways to implement CS have been explored for a wide variety of imaging sensor, *e.g.*, hyperspectral imaging [155], ToF imaging [156], Magnetic Resonance Imaging [157], as well as Multispectral imaging [158]. Among all these types of imaging sensors, we identify two main classes for the randomization of pixel measurements, the optical and on-chip implementations.

CS optical implementations

CS optical implementation approaches can perform basically the CS linear measurements and operate at the optical level in different fashion allowing to save energy for the photo-transducer and the analog-to-digital conversion level.

One implementation, so called Single Pixel Camera (SPC), is composed of a single photodiode combined with a Digital Micromirror Device (DMD) to sequentially capture CS measurements, *i.e.*, image capture with random spatial selection. This DMD is electronically controlled to transmit or not the incident light towards the photodiode for each micromirror. One of the major drawback related to the use of this approach is the need of a bulky DMD that can be a limitation for embedded systems [159, 160] and involving optical issues.
Otherwise, almost equivalent to SPC, coded aperture differs from its composition of a 2D photodiode array which allows to parallelize the CS over each pixel [161–163], *e.g.*, for temporal acquisition/compression.

CS on-chip implementations

Due to significant advances in complementary metal oxide semiconductor (CMOS) technology, the implementation of CS on the focal plane of CMOS sensors is preferred in terms of cost, robustness and stability. For some applications such as embedded applications, the electrical implementation of CS also avoids the use of a CS optical system that results in a large camera size due to the optical design. Recall that, the disadvantage of the canonical approach of first sampling according to the Nyquist-Shannon theorem, converting to a digital representation using an analog-to-digital converter (ADC), and then compressing using a compression standard is that all of the digital values must be stored and then summed together with pseudo-random weights to perform the CS measurement. Thereupon in order to reduce memory usage and instead of detecting the entire image and then removing redundant information during the compression step; the other chosen solution is to collect only the necessary amount of information by performing random projections. These random projections can be applied to the pixels in two different ways in terms of ASIC design: on-focal plane implementations and end-of-column implementations.

On-focal plane CMOS implementations may involve the use of in-pixel memories or digital circuits for the random generator as detailed in [164–166]. These implementations are composed of a finite state machine, linear feedback shift registers (or any Pseudo Random Generator, *e.g.*, based on a Cellular Automaton) and analog processing unit, allowing either the in-pixel real-time convolution or the application of random coefficients over the entire array. For example, [167] presents one of possible on-focal plane implementations based on the projection of the image onto a separable 2-D basis and the measurement of the corresponding expansion coefficients. This implementation consists in using a combination of focal-plane processing performed directly in the pixels (reducing its intrinsic performances), and an analog vector–matrix multiplier to perform computation before the analog-to-digital conversion occurs.

In case of layout restrictions, having individual access to each pixel is impractical for a large array sensor. To tackle this issue, [168] introduced the first end-of-column implementation using a standard pixel architecture, rolling shutter readout and a block matrix transform method. The CS measurement is performed by a compressed sensing multiplexer over 16×16 pixel blocks at the analog to digital conversion level. Furthermore, a random scrambling is performed at the input of each 1^{st} order incremental $\Sigma\Delta$ ADC allowing a random pixels per-column parallel selection, summation and conversion (*i.e.*, concept of CS-$\Sigma\Delta$ presented in [169]). Thereupon, [170] demonstrated the interest of using a high order $\Sigma\Delta$ ADCs instead of a 1^{st} order due to the need of a high oversampling ratio to ensure a proper conversion. Otherwise, [171] presents a novel current-based CS thanks to the Kirchhoff law [172] image sensor without acting on the classical pixel design, as well as the overall sensor architecture. The main change is that HDR is enabled thanks to a specific time diagrams of the control signals and a nondestructive column-based readout. Furthermore, an advanced extension has been proposed by [173] which consists in a compact CMOS

Image Sensor (CIS) architecture enabling embedded object recognition facilitated by a dedicated end-of-column Compressive Sensing (CS). However, the CS approach generally involves in most of the use cases unwanted artifacts due to the reduced dimensionality of the support and the parallelization of the measurements. The following section provides the mathematical foundation of the algorithmic approach, namely the deep learning algorithm as a method for efficient signal processing reconstruction.

3.3 Deep learning Background

Deep learning belongs to the broad field of machine learning, which is "the science and engineering of creating machines that can improve its performance for targeted applications by accumulating knowledge through experience", according to John MacCarthy [174]. Machine Learning (ML) seems to offer state-of-the-art accuracy and even better accuracy, where traditional signal processing methods have shown some limitations. For this reason, the signal processing community has been interested over time in integrating these new methods into the solution of a wide variety of signal processing problems. This has led to the fast development of ML which offers new possibilities to develop predictive models for reconstruction [175], classification [176] and segmentation [177] for computer vision applications.

3.3.1 Deep neural networks (DNNs)

Deep Neural networks takes its form or functionality from the human brain analysis. In fact, the brain is composed of approximately 86 billion neurons themselves composed of axons, acting as the output component of the signal, and dendrites, acting as the input component of the signal. The connection between the neurons through the axon and the dendrite called synapse are about 10^{14}, leading to a huge system with complex interaction networks. From this neuron's computation inspiration, deep neural networks (DNNs) are composed of layers that are themselves composed of activation functions, $f()$ often referring in a way to neurons, and learned weights, $w_{i,j}$ somehow referring to the synapses. The mathematical formulation of one layer is written in Eq. 3.26. Layer cascading involves weighted sums from the inputs or one or more "hidden layers", whereupon it is commonly accepted that we can speak of a DNN when it has more than three layers.

$$y_i = f\left(\sum_{i=1}^{N} w_{i,j} \times x_i + b\right) \tag{3.26}$$

Actually, DNNs are able to learn high-level nonlinear features with high degree of abstraction because of the similarity of its functioning with that of the human brain. Let's take the example of visual data processing. The DNN takes as input the pixels of an image, on which, through each layer, it performs an extraction analysis of various features. Then, the combination of the features between all the layers allows the detection of an object in case the DNN is trained for that task. Indeed, just like the human brain a learning phase is necessary, this process optimizes features extraction using networks parameters (*i.e.*, weights $(w_{i,j})$ and biases b) for various tasks following a certain training paradigm (*i.e.*, learning strategy):

- one called supervised learning when all the training samples are labeled [178],
- another is called unsupervised learning, *i.e.*, without labellisation of the training samples [179],

- a third one, namely semi-supervised learning, consists of a mixture of supervised and unsupervised [180].

Furthermore, there exists various optimization algorithms for optimizing network parameters [181, 182]. The gradient descent process [183] is one of the basic optimization processes which consists in updating the weights with the use of a weighted gradient of the loss relative to each weight. In other words, this consists in calculating the partial derivative of the loss function, noticed L in Eq. 3.27, with respect to each weight $(w_{i,j})$, cf. Eq. 3.28. The computation of the partial derivatives of the gradient is performed through a process namely backpropagation which is iteratively performed backward from the last layer by the chain rule [184] (with α called the learning rate).

$$\mathbb{L}(W) = \frac{1}{N}\sum_{i}^{N} L(f(\sum_{j=1}^{N} w_{i,j} \times x_i + b), y_i) \tag{3.27}$$

$$w_{i,j}^{t+1} = w_{i,j}^{t} - \alpha \left(\frac{\partial \mathbb{L}(W)}{\partial w_{i,j}} \right) \tag{3.28}$$

The gradient descent process is the basic optimization process that works well for convex functions but is only efficient for relatively small datasets. Several other optimization algorithms, each with its own advantages and disadvantages have been further implemented. For example, stochastic gradient [185] allows faster computation at the expense of slower convergence. Alternatively, the Root Mean Square Propagation (RMSprop) [186] is an adaptive gradient descent process, where the square of the gradients are accumulated for maintaining a moving average of the squared gradients of each weight. The second RMSprop insight is that it divides the gradient by the root of this average. Adam optimization [187] is a stochastic gradient descent method based on adaptive moments estimation of first-order and second-order. Its advantages are numerous, like computationally efficient, little memory requirements, appropriate for problems with very noisy and/or sparse gradients. Even though research has been conducted to define the optimization method that outperforms others for a specific task [183], the scientific community admits that there is no general method for all problems.

These optimization processes indicate weight update changes, and is repeated iteratively in order to reduce the overall loss function that can take several forms (*e.g.*, Mean Squared Error (MSE) [188], $l2$ normalization [189, 190], Kullback Leibler (KL) Divergence [191], Hinge loss [192]). The choice of the loss function depends on the problem task. For example, the MSE loss function is commonly used for regression and consists in computing the mean of squares errors between labels and predictions. It implies that large errors between ground truths (or labels) and predictions will be significantly more penalized than small errors. Instead, hinge loss functions are often used for training classifiers because it includes a penalty according to a margin (as SVM do [193]) from the classification boundaries. Finally, the KL loss function is suited for probability distribution prediction (or Generative Adversarial Network (GAN)) since it estimates the difference between two probability distributions (the observed and the targeted).

Note that the training, evaluation and test steps are therefore based on the used of three dataset x_{train}, x_{val} and x_{test} for which there is jointly an associated ground truth dataset or labels y_{train}, y_{val} and y_{test}. The train datasets (x_{train}, y_{train}) are used to train the model, that is to say optimize weights and biases of the DNN. While, the validation datasets (x_{val}, y_{val}) are used to track overfitting issues during the training stage. This frequent evaluation thus allows to fine-tune the model hyperparameters. Once the model is completely trained, the test datasets (x_{test}, y_{test}) are used to provide a performance estimation of a final model fit, on data that were not taken into account

during the training stage.

Focusing on computer vision application, Deep Neural Network (DNN) architectures are split into two groups, which differ in form: feedforward and recurrent. This section presents only the prominent feedforward DNN type of layer *i.e.*, the convolutional layer. The origin of the CNN architecture comes from the "Neocognitron" introduced by Fukushima [194] in early 80's based on the hierarchical receptive field model of the visual cortex of Hubel and Wiesel [195]. The CNNs have found their final name in the work of Waibel et al. [196] with weights shared among temporal receptive fields and back propagation training for phoneme recognition and in the work of LeCun et al. [197] on a practical CNN architecture for optical character recognition (OCR). CNNs consist in a shared-weight structure of the convolutional filters that translate along input features maps (ifmaps) resulting in each point of the weighted sum of all input channels by the related weight value of the filter, *i.e.*, the output feature map (ofmap) cf., Fig. 3.1.

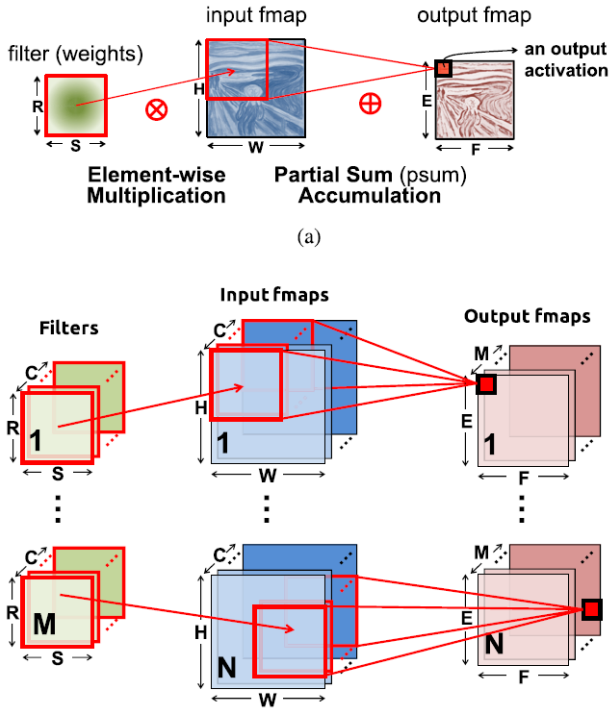


Figure 3.1: Dimensionality of convolutions. (a) 2-D convolution in traditional image processing. (b) High dimensional convolutions in CNNs from [198].

Convolutions are generally followed by a nonlinear function applying a element wise activation to feature maps in order to introduce nonlinearity into the CNN. Various conventional nonlinear functions exist such as sigmoid, hyperbolic tangent, rectified linear unit (ReLU) as well as leaky ReLU and exponential LU, as reported in [199]. Optional layers can be also applied then, like pooling layers and normalization layers. Pooling layers consist in reducing spatial resolution to each channel separately by replacing small neighborhoods in a feature map. This makes the neural network robust and invariant to local distortions while reducing its dimensionality. On the other hand, normalization layers control the feature maps data distribution, by normalizing the distributions of the input data, *e.g.*, so that their sum is equal to one or they have a zero mean and a unit standard deviation. CNNs were originally designed to be used in image recognition and processing. They differ from each other in many ways, as for the list of the most famous CNN architectures, AlexNet [200], VGGNet [201], and ResNet [202].

3D Convolutional Neural Networks (3D CNNs) can be considered as the extension of 2D CNNs for image recognition to the convolutional neural networks for spatio-temporal image recognition *e.g.*, video or 3D imaging application. The third dimension in the 3D CNNs is generally the temporal dimension, and allows the extraction of spatio-temporal features through the use of a 3D kernel moving along the three dimensions for a better capture of dependencies. These 3D CNNs provide competitive results on offline processing video for gesture recognition tasks where 2D images are stacked in a temporal dimension [203, 204], as well as for online video processing even if restricted with the requirement to process a spatio-temporal block at a time [205]. Another use case where 3D CNNs work well and that we mainly consider in this thesis is 3D imaging like medical imaging [206] or LiDAR imaging [12], involving raw spatio-temporal data such as point clouds. In this context, DNNs algorithms have been explored for multimodal reconstructions from SPAD raw data in Chap. 7. In addition, another class of techniques, presented in the next section, was considered in the typical use of SPAD raw data processing to improve their spatial resolution.

3.4 Image Super Resolution

Image Super Resolution (SR) methods consist in recovering high resolution (HR) images from low resolution (LR) images. This process of recovering is highly relevant to enhance the resolution of images in computer vision for a wide variety of applications [207, 208]. In the context of the TCSPC SPAD image sensor, SR techniques would overcome the hardware limitations in terms of spatial resolution of the image sensor, by performing either spatial separation of sources from a group of physical pixels or virtual spatial separation of sources. In this section, we have identified some traditional algorithms based on interpolation and deep learning based oversampling methods.

3.4.1 Interpolation-based upsampling

Image interpolation-based upsampling stands for three main interpolation-based methods, namely the nearest-neighbor interpolation [209], the bilinear interpolation [210] and the bicubic interpolation [211]. The nearest-neighbor interpolation is the most naive approach and consists in selecting the only value of the nearest pixel for every pixel position to be interpolated. However, this approach unfortunately produces poor SR results. Alternatively, bilinear interpolation consists in performing a quadratic interpolation *i.e.*, linear interpolation on the two axes of the image, functions of the two nearest neighbors in each direction. Finally, bicubic interpolation [212] consists similarly to the Bilinear interpolation performing cubic interpolation on each of the two axes functions of the four nearest neighbors in each direction. This approach, if process speed is not an issue, provides smoother surface results with fewer artifacts. One of the main drawbacks of these interpolation-based upsampling methods is that they can amplify noise measurements or lead to blurred results, as the image scaling is only based on a single image, hence the use of DL-based upsampling methods instead.

3.4.2 Deep Learning-based SR

SoA works have proposed a variety of super resolution deep learning-based models, trained with both LR images and corresponding HR images. The differences between all these works are nu-

merous, based on various frameworks, upsampling methods, network design. Without focusing on the network design (*i.e.*, upsampling methods position in the topology), an overview of the different DL-based upsampling layers and other DL-based methods for performing SR is provided.

Transposed Convolution Layer [213], also called a deconvolution layer [214], increases the resolution of the image by first expanding the image with inserted zeros and then performing the convolution. For example, with a kernel size of 3×3, a stride equal to 1 and padding equal to 1, the input is upsampled by a factor of 2. This operation is the opposite transformation of a normal convolution *i.e.*, aims to predict possible input from the output feature maps. However, because of zero inserted and overlapping pixels, this layer can create a checkerboard-like pattern.

Sub-pixel Layer [215] instead consists in generating several channels with standard convolution and then reshaping such that the number of channels are reduced and the image size increased. For example, let's take the example of the 2D input image size of $h \times w \times c$. The convolution layers provide an output feature map of $h \times w \times s^2c$, yielding to a possible scaling of the image output size of $sh \times sw \times c$ after reorganisation of the tensor data. This approach has the main drawback of having artifacts on the edges of objects in the images since some pixels share the same receptive fields.

DL-based Fusion method is a common way to improve the resolution and quality of temporal (or depth) information when being associated to high resolution color images. The fusion method can be performed in various ways. For example, [216] proposed to fuse multiple slightly moved depth measurements into one depth map with the use of a bilateral regularization. Otherwise, [217] also proposed to fuse the intensity information of an additional sensor with the depth value of ToF. The fusion is performed by an iterative refinement process of bilateral filtering of a depth volume and an RGB image. Although these methods provide high quality depth results, their performance depends on a complex calibration process and the use of additional sensors.

3.5 Conclusion

Throughout this chapter is provided a brief description of the four fields of signal processing methods that are Compressive Sensing, Expectation Maximization algorithms, Deep Learning and Image Super Resolution methods. All these methods of signal processing presented allow to improve the performances of reconstructions of the measured physical quantities, either by improving the spatial resolution, or the temporal resolution. In the context of TCSPC SPAD-based 3D imaging, some of these methods were employed in different ways, as will be further detailed in the SoA subsection of each chapter.

4

Time distribution modeling for piecewise-constant Poisson process rate

In active 3D imaging based on a Time Correlated Single Photon Counting (TCSPC) system, photons "of interest" measured with Single Photon Avalanche Diodes (SPADs) are often mixed with high background noise. This low Signal-to-Background ratio (SBR) makes the reconstruction of albedo and depth maps difficult. State-of-the-art (SoA) works relying on Bayesian approaches [10, 11] or Deep Learning (DL) [12, 218] usually study the restrictive case of low-photon counts mode of operation. On the contrary, this chapter aims to consider the high photon counts operating mode, where the SPAD "Dead Time" and laser pulse are spread over a large number of bins of the Time-of-Flight (ToF) histogram (cf. Fig. 4.1) relaxing system-level constraints such as ultra short laser pulse at shallow depths and photon detection probability of SPAD (resulting in pile-up effect). A new method is then proposed to estimate pixel parameters from such a ToF histogram in which the photon arrival times is assumed to follow a truncated-Shifted Erlang ($\mathscr{E}$) distribution. The underlying algorithmic task presented in this chapter consists in estimating 4 latent parameters of each $\mathscr{E}$ distribution of a mixture model, only from an observed draw of the process distribution in the shape of a ToF histogram. To solve the highly non-convex nature of this problem, a customized nested Expectation-Maximization algorithm (c-EM) has been designed based on a proper combination of Maximum Likelihood Estimation, Moments Matching, Parametric Imputation and support estimation via Variable Neighborhood Search. The proposed framework was successfully evaluated with synthetically generated data leading to accurate depth reconstructions under 0.03 m *RMSE and albedo reconstructions over 15dB PSNRs.*

Contents

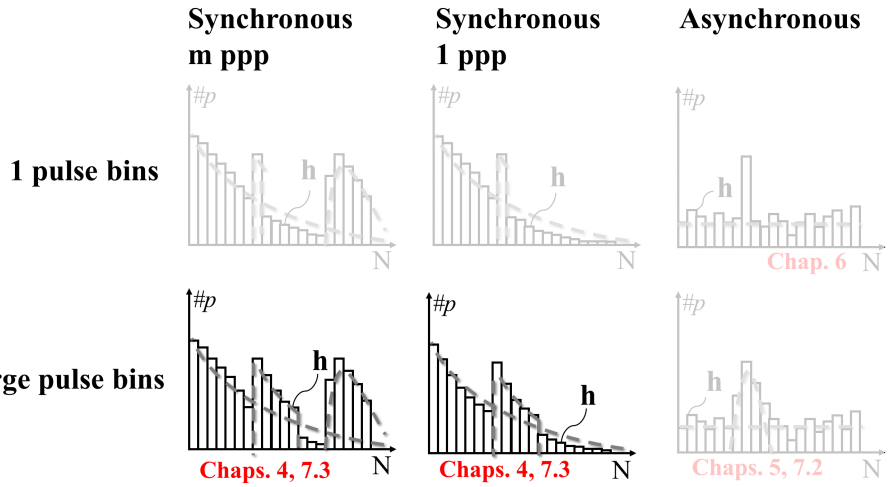


Figure 4.1: ***TCSPC distribution modeling assumptions throughout this chapter.***

4.1 Related works

4.1.1 Non-stationary Poisson process parameters inference

Non-stationary Poisson process (NSPP) well describes most of service systems subject to time varying arrival rates, and has been intensively studied in the litterature. For example, nonparametric solutions have been developed for inferring NSPP parameters such as Auto Regressive Integrated Moving Average (ARIMA) [219], and Neural Networks (NN) [220]. However, these solutions are suitable for linear and non-stationary signals without high Poisson rates discontinuities. That is why, [221] proposes to introduce the Taylor's fluctuation scaling effect into the State Space Model with the Particle Filter to handle high discontinuities issues. Otherwise, for the same purpose, [222] aims at constructing a table of the inverse cumulative arrival rate function in order to decompose the NSPP into constant Poisson sub-processes. The disadvantage of these methods is that they are unattractive when systems involve high dimensional data with harsh memory constraints (*e.g.*, 3D imaging system when using a TCSPC system [223]). To relax the memory constraints, [224] proposes statistical analysis instead of analyzing the sequence leading to more complex analysis.

4.1.2 State-of-the-art on statistical tools involved

In this section, we report an Expectation-Maximization (EM) algorithm ideally suited to problems of this sort *i.e.*, for Time Correlated Counting Process, and also when direct access to the data is only possible in its discrete distribution format or when some data are missing. Some additional mathematical tools are also presented like stochastic and genetic-based optimization methods, and order of the mixture model estimation methods.

Expectation-Maximization algorithm

The Expectation-Maximization algorithm [116] is partially non-Bayesian algorithm for estimating the parameters of mixture of distributions. This method has been first introduced by [225] for estimating the parameters of time distribution of events generated by a Poisson process (*i.e.*, Phase-type (PH) distribution of fixed order). Although many improvements have been developed during the last few years, in order to fit more phases [226] and to overcome the truncated data issue [227]. Aside from [228], which is a Bayesian approach to quantify non-stationary Poisson process inputs uncertainty using Gaussian process. The inference of the time distribution of a queuing system [226] outputs with a PCPP input via an appropriate PH distributions has not yet been fully addressed due to the high complexity of the model.

Stochastic optimization extensions

Local convergence can be encountered using a standard EM algorithm to infer Phase-type distribution (PH) parameters. To find a global optimum, greedy algorithms may be used with random restart initialization of several local search methods (*e.g.*, hill climbing algorithms [229]). Otherwise, [230] reports EM extensions to overcome local convergence without excessively increasing the computational complexity of the algorithm. For example, the Stochastic Expectation Maximization (SEM) method, relies on the stochastic imputation principle. In case of erratic behavior of SEM, a variant has been developed: Simulated Annealing (SA) Expectation Maximization [231, 232]. This SA variant consists in weighting the update values of the EM and SEM methods with a slow decrease in the probability of accepting worse solutions as the solution space is explored. However, this may not always be sufficient for highly nonlinear objective functions. The proposed method therefore uses additional metaheuristic methods from combinatorial optimization that generate a good solution into a finite set of solutions (*e.g.*, the Genetic Algorithm

(GA) [233,234]). The proposed approach has been inspired by the natural selection process which consists in selecting the best solution among all the possibilities to create a new mutation for the next iteration ("generation").

Order of the mixture model estimation

Besides estimating distribution parameters, an additional purpose is to evaluate the number of components in the sample distribution. The order selection in the mixture model has a direct impact on the convergence of the EM algorithm. Among all state-of-the-art solutions [235], two main techniques have been chosen for minimizing the number of components and finding the best convergence model. First, the information criteria technique, is based on the bias-corrected likelihood by adding a negative term to the likelihood function in order to penalize the model fit (*e.g.*, the Akaike criterion (AIC) [236], the Bayesian Information Criterion (BIC) [231]). Otherwise, the Likelihood ratio test technique consists in starting with a small number of components in the mixture model in order to converge to the adequate number of components using likelihood ratio test function, like a chi-square test.

Incomplete data analysis

Finally, queuing systems can suffer from incomplete data of three types as surveyed in [237]. Focusing on the Missing Not At Random (MNAR) incompleteness type which disrupts EM algorithm convergence, [238] proposes an algorithm for multivariate time series model estimation in combining a SEM algorithm with a Markov Chain Monte Carlo. As of now, the most common technique in combination with an EM algorithm, is the imputation method as presented in [239–241] which consists in substituting incomplete values using various imputation methods. Classic imputations refer to cold-deck [242] (*i.e.*, static imputation) and hot-deck [243] (*i.e.*, dynamic imputation) where the unobserved values are individually substituted making the algorithm computationally intensive. Regression imputation [244] will be not considered in this work due to non-parametric censored data. The chosen imputation method refers to the mean substitution [245], avoiding any bias estimation of the EM algorithm for density function tending to infinity.

4.2 c-EM for a Mixture of truncated-Shifted Erlangs

The objective, through the use of a mixture of $\mathscr{E}$, is to estimate the number of PCPP intensity levels and finds their boundary locations with the use of the corresponding time index of each variation, α and β. Considering that the PCPP is composed of C Poisson intensity levels being the superposition of C independent Poisson arrival sub-processes [109], the number of components of the mixture distribution is equal to S, *i.e.*, the amount C of rate change multiplied by the ratio of measurement period time over the service time (*i.e.*, $S = \sum_{c=1}^{C} Card(\mathbb{G}_c)$ commonly with $Card(\mathbb{G}_c) \leq 2$ in the typical TCSPC use case). This gives the following mixture model $m\mathscr{E}$:

$$m\mathscr{E}(z;\boldsymbol{\Theta}) = \sum_{s=1}^{S} \pi_s \frac{\lambda_s^{k_s}(z-\alpha_s)^{k_s-1}\,\mathrm{e}^{-\lambda_s(z-\alpha_s)}}{(k_s-1)\,!}\mathbf{1}_{z\in[\alpha_s,\beta_s]}, \tag{4.1}$$

where π_s corresponds to the weight of the s^{th} phase and $\boldsymbol{\Theta}$ is the concatenation of S different set of parameters $\boldsymbol{\theta}_s^T = [\lambda_s, k_s, \alpha_s, \beta_s]$. λ_s, k_s, α_s and β_s respectively stand for the intensity parameter, the shape parameter, the starting and the ending times of the s^{th} phase.

Algorithm 2: c-EM algorithm to fit $\mathscr{E}$

Data: $\mathbf{h} \in \mathbb{N}^B$
Result: Estimated parameter $\hat{\boldsymbol{\Theta}}$ from its random initialization
// E-step (section 4.2.1)
for $s \leftarrow 1$ **to** S **do**
  $\hat{\gamma}_{b,s} \leftarrow h_b \frac{\pi_s \mathscr{E}(b;\hat{\boldsymbol{\theta}}_s)}{\sum_{j=1}^{S} \pi_j \mathscr{E}(b;\hat{\boldsymbol{\theta}}_j)}, b \in \{1,2,...,B\}$
end
// M-Step (section 4.2.2)
for $s \leftarrow 1$ **to** S **do**
  $\hat{\lambda}_s \leftarrow \frac{k_s \hat{\pi}_s}{\hat{\mu}_s^{(obs)} + \hat{\mu}_s^{(cens)}}$, $\hat{\pi}_s \leftarrow \frac{\sum_{b=1}^{\hat{\beta}_s} \hat{\gamma}_{b,s}}{\mathscr{E}_c(\hat{\boldsymbol{\theta}}_s)}$

$$\hat{\alpha}_s \leftarrow \begin{cases} \hat{\alpha}_s - \varepsilon_s, \text{ if } D_{KL}(\hat{\mathbf{h}} \| \mathbf{h}) \geq D_{KL}(\hat{\mathbf{h}}_{\varepsilon_s} \| \mathbf{h}) \\ \hat{\mu}_s - \sqrt{k_s \hat{v}_s}, \text{ otherwise.} \end{cases}$$

  $\hat{\beta}_s \leftarrow \hat{\alpha}_s - \frac{W_{-1}\left(\left(\frac{1}{e}\right)^{(k_s-1)}(\rho_s-1)\right)+(k_s-1)}{\hat{\lambda}_s}$
end

4.2.1 Histogram-based expectation (E-step)

In our typical use case, we state that the collected data, $\mathbf{z} \in \mathbb{R}^N$, can be depicted through an approximation of its $\mathscr{E}$ distribution, the measured histogram $\mathbf{h}$. Indeed, all the observation times values are counted onto a series of time-intervals, which results in an histogram of the observations. Although losing the information provided by the observation sequences for the parameters estimation, storing and processing only the histogram makes the inference problem more tractable both in terms of memory requirements and computing resources. Therefore, the time series of size N are sampled through the observations distribution only, *i.e.*, $\mathbf{h} \in \mathbb{N}^B$, with bins indexed $b \in \{1,...,B\}$ and bin width B_w. For notation simplification, α_s and β_s will now define $\alpha_s = \left\lfloor \frac{\alpha_s}{B_w} \right\rfloor$ and $\beta_s = \left\lfloor \frac{\beta_s}{B_w} \right\rfloor$, *i.e.*, bin-quantized shift and censoring times. The truncated-Shifted Erlang PDF $\mathscr{E}$ in Eq. 4.1 is thus revised here:

$$\mathscr{E}(b;\boldsymbol{\theta}_s) = \pi_s \frac{\lambda_s^{k_s}(b-\alpha_s)^{k_s-1}\, \mathrm{e}^{-\lambda_s(b-\alpha_s)}}{(k_s-1)\,!} \mathbf{1}_{b \in [\alpha_s, \beta_s]}. \tag{4.2}$$

Due to its better tractability, the "log-likelihood" optimization trick is preferred (cf. Eq. 4.3).

$$\begin{aligned} l(\boldsymbol{\Theta}|\mathbf{h}) &= \sum_{b=1}^{B} \sum_{i=1}^{h_b} \ln \left(\sum_{s=1}^{S} \delta_{b,i,s} \pi_s \mathscr{E}(b;\boldsymbol{\theta}_s) \right) \\ &= \sum_{b=1}^{B} \sum_{s=1}^{S} \gamma_{b,s} \ln \left(\pi_s \mathscr{E}(b;\boldsymbol{\theta}_s) \right) \end{aligned} \tag{4.3}$$

Following this notation, $\delta_{b,i,s}$ is defined as equal to 1 if and only if the corresponding sample belongs to the s^{th} phase, and 0 otherwise. It advantageously allows to move the summation over s outside the logarithm into the logarithm in order to rewrite the log-likelihood function while introducing $\gamma_{b,s}$. Note that h_b represents the measured bin value of $\mathbf{h}$ at index b and $\gamma_{b,s} = \sum_{i=1}^{h_b} \delta_{b,i,s}$ refers to the total amount of events occurring at time bin b for the phase s. Moreover, $\gamma_{b,s}$ variables are latent, in the sense that they are not directly observed, justifying the use of the EM algorithm rather than regular Maximum Likelihood Estimation (MLE).

4.2.2 Combination of optimization methods (M-step)

Note that, [246,247] already derived an EM type of algorithm for PH distributions under truncated data. However, an EM algorithm is not straightforwardly compatible in the case of $m\mathscr{E}(b;\boldsymbol{\Theta})$ because PH random variables do not have the same sampling space, especially due to the shift parameter $\hat{\alpha}_s$. Therefore, this chapter presents parameter estimation variants in order to overcome this issue.

Erlang weight, $\hat{\pi}_s$

Firstly, to avoid a bias on the weight estimation obtained by simply adding the corresponding estimated values (E-step), as written in Eq. 4.4.

$$\hat{\pi}_s = \frac{\sum_{b=1}^{B} \hat{\gamma}_{b,s}}{\sum_{s=1}^{S}\sum_{b=1}^{B} \hat{\gamma}_{b,s}} \tag{4.4}$$

The truncated time parameter β_s needs to be "removed" from $\hat{\pi}_s$ calculation. To do so, $\hat{\pi}_s$ can be inferred thanks to the Cumulative Distribution Function (CDF) $\mathscr{E}_c\left(\hat{\boldsymbol{\theta}}_s\right)$ as written in Eq. 4.5 which normalizes the estimated values only over the interval of distribution support, defined as $[\hat{\alpha}_s, \hat{\beta}_s]$.

$$\hat{\pi}_s = \frac{\sum_{b=1}^{B} \hat{\gamma}_{b,s}}{\mathscr{E}_c\left(\hat{\boldsymbol{\theta}}_s\right)} = \frac{\sum_{b=1}^{B} \hat{\gamma}_{b,s}}{1 - \sum_{j=0}^{k_s-1} \frac{\mathrm{e}^{-\hat{\lambda}_s(\hat{\beta}_s-\hat{\alpha}_s)}\left(\hat{\lambda}_s(\hat{\beta}_s-\hat{\alpha}_s)\right)^j}{j!}} \tag{4.5}$$

Erlang intensities, $\hat{\lambda}_s$

The standard MLE of the Erlang intensities would be:

$$\hat{\lambda}_s = \frac{k_s\hat{\pi}_s}{\hat{\mu}_s - \hat{\alpha}_s} = \frac{k_s\hat{\pi}_s}{\hat{\mu}_s^{(obs)} + \hat{\mu}_s^{(cens)}} = \frac{k_s\hat{\pi}_s}{\sum_{b=\hat{\alpha}_s+1}^{\infty} \hat{\gamma}_{b,s}(b-\hat{\alpha}_s)}. \tag{4.6}$$

where μ_s refers to the first moment estimation. However, a proxy problem is rather considered here because the zeroing truncation of $\hat{\gamma}_{b,s}$ over the interval $[\![\hat{\beta}_s, +\infty]\!]$ corrupts the estimation of $\hat{\lambda}_s$. Thus, a Parametric Imputation (PI) is used to infer the first moment of the truncated samples, $\hat{\mu}_s^{(cens)}$, in addition to the first moment of observed samples, $\hat{\mu}_s^{(obs)} = \sum_{b=\hat{\alpha}_s+1}^{\hat{\beta}_s} \hat{\gamma}_{b,s}(b-\hat{\alpha}_s)$ (cf. Alg. 2-line 5). Indeed, assuming that $\hat{\lambda}_s$ remains the same before and after censoring, the calculation of $\hat{\mu}_s^{(cens)}$ can be performed using the standard unshifted Erlang distribution integral (cf., Eq. 4.7).

$$\begin{aligned}
\hat{\mu}_s^{(cens)} &= \hat{\pi}_s \int_{\hat{\beta}_s}^{\infty} b\mathscr{E}(b;\hat{\boldsymbol{\theta}}_s)db \\
&= \hat{\pi}_s \int_{\hat{\beta}_s}^{\infty} b\frac{\hat{\lambda}_s^{k_s}(b-\hat{\alpha}_s)^{k_s-1}\,\mathrm{e}^{-\hat{\lambda}_s(b-\hat{\alpha}_s)}}{(k_s-1)\,!}db \\
&= \hat{\pi}_s \frac{k_s}{\hat{\lambda}_s} \int_{\hat{\beta}_s}^{\infty} \frac{\hat{\lambda}_s^{k_s+1}(b-\hat{\alpha}_s)^{k_s}\,\mathrm{e}^{-\hat{\lambda}_s(b-\hat{\alpha}_s)}}{k_s\,!}db \\
&= \hat{\pi}_s \frac{k_s}{\hat{\lambda}_s} \sum_{j=0}^{k_s} \frac{\mathrm{e}^{-\hat{\lambda}_s(\hat{\beta}_s-\hat{\alpha}_s)}\left(\hat{\lambda}_s(\hat{\beta}_s-\hat{\alpha}_s)\right)^j}{j!}
\end{aligned} \tag{4.7}$$

Erlang shift, $\hat{\alpha}_s$

A Moment Method (MM) [248] combined with a custom Variable Neighborhood Search (VNS) [249] is introduced here to properly estimate $\hat{\alpha}_s$ while bypassing the intractability of the standard log-likelihood maximization formulation. First, considering the Shifted-Erlang distribution function $\mathscr{E}$, and defining the latent equivalent variance $\hat{v}_s$ for an untruncated Erlang distribution as a sum of two independent random variables (respectively consisting in observed and truncated data), we then have $\hat{\mu}_s = \hat{\mu}_s^{(obs)} + \hat{\mu}_s^{(cens)} + \hat{\alpha}_s$ as well as $\hat{v}_s = \hat{v}_s^{(obs)} + \hat{v}_s^{(cens)}$ and we can therefore define the following update rule for $\hat{\alpha}_s$,

$$\hat{\alpha}_s = \hat{\mu}_s - \sqrt{k_s \left(\hat{v}_s^{(obs)} + \hat{v}_s^{(cens)} \right)}. \tag{4.8}$$

The empirical estimation of the variance for the observed data is then obtained using the König-Huygens theorem [110]:

$$\hat{v}_s^{(obs)} = \sum_{b=\hat{\alpha}_s+1}^{\hat{\beta}_s} \hat{\gamma}_{b,s} (b - \hat{\alpha}_s)^2 - \left(\hat{\mu}_s^{(obs)} \right)^2. \tag{4.9}$$

On the other hand, $\hat{v}_s^{(cens)}$ is estimated using the standard unshifted Erlang distribution integral.

$$\hat{v}_s^{(cens)} = \eta - \left(\hat{\mu}_s^{(cens)} \right)^2, \tag{4.10}$$

where η is computed using Eq. 4.11:

$$\begin{aligned}
\eta &= \hat{\pi}_s \int_{\hat{\beta}_s}^{\infty} b^2 \mathscr{E}(b; \hat{\boldsymbol{\theta}}_s) db \\
&= \hat{\pi}_s \int_{\hat{\beta}_s}^{\infty} b^2 \frac{\hat{\lambda}_s^{k_s} (b - \hat{\alpha}_s)^{k_s - 1} \, \mathrm{e}^{-\hat{\lambda}_s (b - \hat{\alpha}_s)}}{(k_s - 1)\,!} db \\
&= \hat{\pi}_s \frac{k_s (k_s + 1)}{\hat{\lambda}_s^2} \int_{\hat{\beta}_s}^{\infty} \frac{\hat{\lambda}_s^{k_s+2} (b - \hat{\alpha}_s)^{k_s+1} \, \mathrm{e}^{-\hat{\lambda}_s (b - \hat{\alpha}_s)}}{(k_s + 1)\,!} db \\
&= \hat{\pi}_s \frac{(k_s + 1) k_s}{\hat{\lambda}_s^2} \sum_{j=0}^{k_s+1} \frac{\mathrm{e}^{-\hat{\lambda}_s (\hat{\beta}_s - \hat{\alpha}_s)} \left(\hat{\lambda}_s (\hat{\beta}_s - \hat{\alpha}_s) \right)^j}{j\,!}.
\end{aligned} \tag{4.11}$$

However, if $\hat{\alpha}_s$ overpasses the latent value α_s, bins that are observed but overruled by the c-EM (*i.e.*, $bB_w < \hat{\alpha}_s$) cannot contribute to reduce the value of $\hat{\alpha}_s$. It motivated the introduction of a VNS, randomly drawing a step back variable ε_s that follows an exponential decay rate of variance given by the MM update value (Eq. 4.12). The drawn value giving the lowest distance using the Kullback–Leibler divergence, $D_{KL}()$ [250] is then selected, cf., Alg. 2-line 6.

$$\varepsilon_s \sim \mathrm{Exp}\left(\hat{\mu}_s^{(obs)} + \hat{\mu}_s^{(cens)} - \sqrt{k_s \hat{v}_s} \right) \tag{4.12}$$

Erlang right-truncation, $\hat{\beta}_s$

Instead of using a likelihood function of the truncated data which is not tractable in our case [251], we propose to find the value $\hat{\beta}_s$ for which the numerical calculation of the CDF on the interval $[\hat{\alpha}_s, \hat{\beta}_s]$ is equal to the following normalized analytical formulation:

$$1 - \sum_{j=0}^{k_s-1} \frac{\mathrm{e}^{-\hat{\lambda}_s (\hat{\beta}_s - \hat{\alpha}_s)} \left(\hat{\lambda}_s (\hat{\beta}_s - \hat{\alpha}_s) \right)^j}{j\,!} = \frac{\sum_{b=1}^{B} \hat{\gamma}_{b,s}}{\hat{\pi}_s}. \tag{4.13}$$

Unfortunately, this method fails as soon as $\hat{\beta}_s < \beta_s$ because the observed data belonging to the s^{th} $\mathscr{E}$ will be rejected on both terms, disabling an appropriate update of $\hat{\beta}_s$. In order to enable stretching forwards $\hat{\boldsymbol{\beta}}$, we propose to add a Laplace-windowed version of the residual distribution $(\hat{h}_b - h_b)$ to the normalized numerical calculation term. Therefore, we define:

$$\rho_s = \frac{\sum_{b=1}^{B} \hat{\gamma}_{b,s} \left(1 + (\hat{h}_b - h_b) \frac{1}{\sqrt{(B-\hat{\beta}_s)}} \mathrm{e}^{-\frac{1}{\sqrt{(B-\hat{\beta}_s)}} |b-\hat{\beta}_s|} \right)}{\hat{\pi}_s}. \tag{4.14}$$

To update $\hat{\beta}_s$, we then force the following equality:

$$1 - \sum_{j=0}^{k_s-1} \frac{\mathrm{e}^{-\hat{\lambda}_s(\hat{\beta}_s - \hat{\alpha}_s)} \left(\hat{\lambda}_s(\hat{\beta}_s - \hat{\alpha}_s) \right)^j}{j\,!} = \rho_s \tag{4.15}$$

Solving Eq. 4.15 for $\hat{\beta}_s$ thus leads to the Eq. of Alg. 2-line 6, $\forall k_s \in \{1,2\}$ without bias and without unwanted local convergence of $\hat{\beta}_s$, where W_{-1} refers to the second principal branch of the Lambert function [252].

4.3 Priors in c-EM facilitating global optimization

Even if we consider that the number of phases S is unknown, it is assumed to be the sum of the number of sub-phases of each Poisson sub-processes (*i.e.*, $S = \sum_{c=1}^{C} Card(\mathbb{G}_c)$ with $Card(\mathbb{G}_c) \in \{1,2\}$). From this and based on some other prior knowledges described below, extensions to the c-EM algorithm have been further developed to tackle the highly non-convex mixture parameters estimation in combination with the order selection of this mixture model as written in Alg. 3.

4.3.1 Custom regularization

A custom regularization is introduced in a matter of mutually constraining the parameters optimization of every $\mathscr{E}$, under application-specific priors and in order to ease a faster convergence. Firstly, $\mathscr{E}$ distributions describing the same Poisson sub-process c should have a common scale parameter $\boldsymbol{\lambda}_c$, (*i.e.*, $\forall s \in \mathbb{G}_c,\ \boldsymbol{\lambda}_s = \boldsymbol{\lambda}_c$). Secondly, arrival times measured by one node of the system are necessarily spaced by one service time (known) due to the previous arrivals (cf., $\mathbf{D_T} \in \{\tau\}^{Card(\mathbb{G}_c)}$ in Eq. 4.16). These equality constraints are relaxed by using $l2$ regularization [253] terms $\|\boldsymbol{\Delta}\hat{\boldsymbol{\lambda}}_c^{\top}\|_2^2$ and $\|\boldsymbol{\Delta}\hat{\boldsymbol{\alpha}}_c^{\top} - \mathbf{D_T}\|_2^2$ as reported in Eq. 4.16, with $\boldsymbol{\Delta}$ a differential operator. In addition, a "low-pass filter" term weighted by a scaling parameter ζ is introduced to avoid abrupt changes from one iteration to the next, this for both $\hat{\boldsymbol{\lambda}}_c^{\top}$ and $\hat{\boldsymbol{\alpha}}_c^{\top}$.

$$\begin{aligned} \hat{\boldsymbol{\lambda}}_c^{\top} &= \underset{\hat{\boldsymbol{\lambda}}_c}{\arg\min}\ (\zeta\|\hat{\boldsymbol{\lambda}}_c^{\top} - \hat{\boldsymbol{\lambda}}_c^{\top(prev)}\|_2^2 + \|\boldsymbol{\Delta}\hat{\boldsymbol{\lambda}}_c^{\top}\|_2^2) \\ \hat{\boldsymbol{\alpha}}_c^{\top} &= \underset{\hat{\boldsymbol{\alpha}}_c}{\arg\min}\ (\zeta\|\hat{\boldsymbol{\alpha}}_c^{\top} - \hat{\boldsymbol{\alpha}}_c^{\top(prev)}\|_2^2 + \|\boldsymbol{\Delta}\hat{\boldsymbol{\alpha}}_c^{\top} - \mathbf{D_T}\|_2^2) \end{aligned} \tag{4.16}$$

4.3.2 Order selection of the mixture model

To properly estimate the order of the mixture model, an outer loop aims at identifying the number of mixture groups, embedding the inner loop which estimates each mixture model parameters until a stopping criterion is reached (*i.e.*, the relative difference –between successive iterations–

Algorithm 3: Genetic based c-EM (c-GEM) algorithm to fit a mixture of truncated Shifted Erlang distribution

Data: $\mathbf{h} \in \mathbb{N}^B$
Input: $P, C, \varepsilon_d, \varepsilon_{KL}, \chi^2, \hat{\chi}^2$
Result: "Global" estimated parameter $\hat{\boldsymbol{\Theta}}^{(1)}$

while $\hat{\chi}^2 > \chi^2$ **do**
  for $p \leftarrow 1$ **to** P **do**
    $C \leftarrow C + 1$
    $\hat{\boldsymbol{\Theta}}_C^{(p)} = \left[\hat{\boldsymbol{\lambda}}_C^\top, \hat{\boldsymbol{\pi}}_C^\top, \hat{\boldsymbol{\alpha}}_C^\top, \hat{\boldsymbol{\beta}}_C^\top\right]$
  end
  while $\left(\Delta\boldsymbol{\Theta}^{(1)} > \varepsilon_d\right) \times \left(D_{KL}^{(1)} > \varepsilon_{KL}\right)$ **do**
    for $p \leftarrow 1$ **to** P **do**
      // c-EM (Sec. 4.2):
      $\hat{\boldsymbol{\alpha}}^{(p)}, \hat{\boldsymbol{\beta}}^{(p)}, \hat{\boldsymbol{\pi}}^{(p)}, \hat{\boldsymbol{\lambda}}^{(p)} \leftarrow$ Alg. 2
      // Regularization (Sec. 4.3.1):
      $\hat{\boldsymbol{\lambda}}_c^{(p)} \leftarrow \zeta \hat{\boldsymbol{\lambda}}_c^{(p)} (\boldsymbol{\Delta}^\top \boldsymbol{\Delta} - \zeta \mathbf{I}_d)^{-1}$
      $\hat{\boldsymbol{\alpha}}_c^{(p)} \leftarrow (\boldsymbol{\Delta}^\top \boldsymbol{\Delta} - \zeta \mathbf{I}_d)^{-1} (\boldsymbol{\Delta}^\top \mathbf{D}_\mathbf{T} + \zeta \hat{\boldsymbol{\alpha}}_c^{(p)})$
    end
    // Genetic search (Sec. 4.3.2)
    if *Iterations* == *MaxIt* **then**
      // Selection
      for $p \leftarrow 1$ **to** P **do**
        $\hat{\boldsymbol{\Theta}}^{(p)} \leftarrow Sort\left(\hat{\boldsymbol{\Theta}}\right)$
      end
      // Crossover and mutation
      for $p \leftarrow \lfloor \frac{P}{2} \rfloor + 1$ **to** P **do**
        $\hat{\boldsymbol{\Theta}}^{(p)} \leftarrow \hat{\boldsymbol{\Theta}}^{(P+1-p)}$
        for $s \leftarrow 1$ **to** S **do**
          $\hat{\alpha}_s^{(p)} \overset{iid}{\sim} Supp(max\{\hat{\mathbf{h}}^{(p)} - \mathbf{h}, 0\})$
        end
      end
    end
    // Evaluation
    for $p \leftarrow 1$ **to** P **do**
      $D_{KL}^{(p)}(\hat{\mathbf{h}}^{(p)} \| \mathbf{h}) = \sum_{b=1}^{B} \hat{h}_b^{(p)} log\left(\frac{\hat{h}_b^{(p)}}{h_b}\right)$
    end
  end
  // Order selection (Sec. 4.3.2)
  $\hat{\chi}^2 = \sum_{b=1}^{B} \frac{(\hat{h}_b^{(1)} - h_b)^2}{h_b}$
end

of the parameter vectors values below $\varepsilon_d = 10^{-7}$, and the relative difference of the Kullback-Leibler [250] below $\varepsilon_{KL} = 10^{-7}$ as written in Alg. 3). In practice, the outer loop iteratively adds a new group of G phases in the mixture model depending on the result of a Chi-square test [254] (with a confidence threshold of 0.95). Note that, the number of groups, C, is initialized to one with an initialization of G phases per group.

In order to handle the high non-convexity of the problem, several runs (noticed P in Alg. 3) of the c-EM with various parameters initialization are first evaluated using the Kullback-Leibler divergence function ($D_{KL}()$). Half of the resulting models with the lowest divergence are duplicated and then mutated, as the natural selection step of an evolutionary algorithm [255] (cf., Alg. 3). The mutation consists in a random draw of $\hat{\alpha}_s$, $\forall s \in S$ from the positive residual support function (*i.e.*, $\hat{\alpha}_s \overset{iid}{\sim} Supp(max\{\hat{\mathbf{h}} - \mathbf{h}, 0\})$).

4.4 Experimental results

4.4.1 Synthetic data trace fit, parametric estimation

Our genetic based c-EM (c-GEM) is first evaluated on 100 synthetic $m\mathscr{E}$ distributions with random parameters $\boldsymbol{\Theta}$; and with an order of mixture from 1 (C = 1, G = 1) to 8 (C = 4, G = 2). Typical proper model fits are reported in Fig. 4.2 that even include groups with overlapping distribution supports.

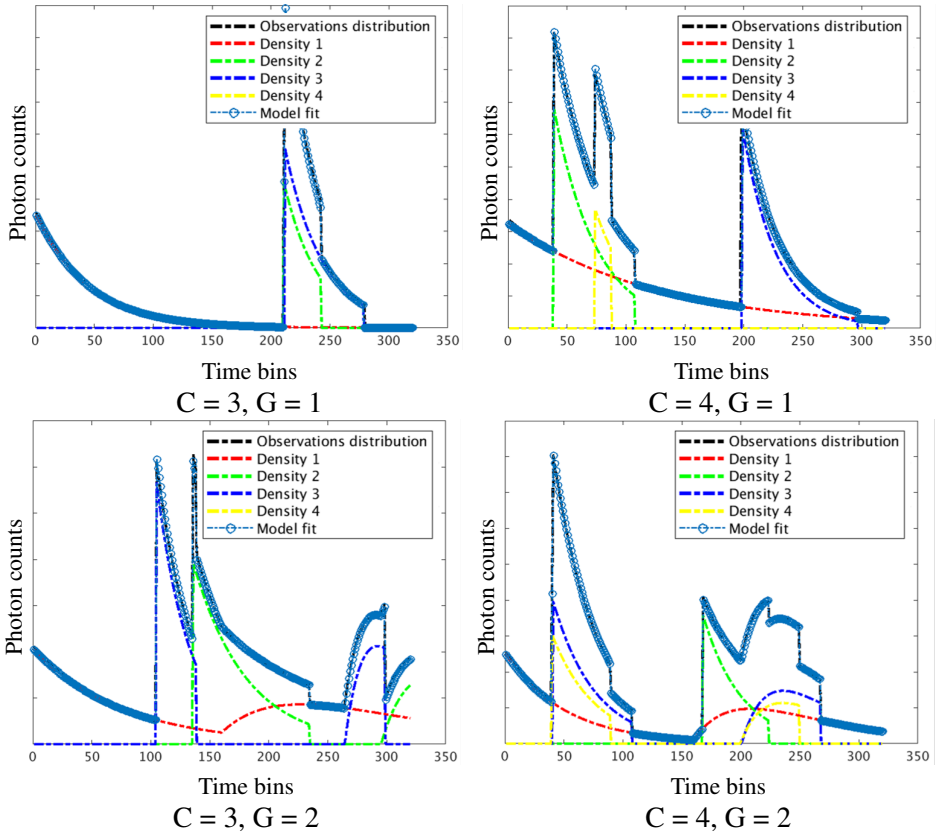


Figure 4.2: c-EM estimations on synthetic histogram data.

Table 4.1 shows the average of parameters relative error obtained from the c-GEM algorithm

	$\mathbf{h}$ (%)	$\boldsymbol{\lambda}$ (%)	$\boldsymbol{\alpha}$ (%)	$\boldsymbol{\pi}$ (%)	$\boldsymbol{\beta}$ (%)
C=1, G=1	0.7	0.5	0.0	0.8	0.0
C=1, G=2	1.2	0.7	0.0	0.7	0.0
C=2, G=1	0.3	2.8	1.0	2.1	0.0
C=2, G=2	0.4	5.9	0.6	4.9	0.2
C=3, G=1	0.1	7.2	1.8	12.0	3.8
C=3, G=2	0.1	17.2	2.4	15.8	2.0
C=4, G=1	0.1	25.1	5.0	26.2	7.7
C=4, G=2	0.1	25.6	5.9	30.6	4.5

Tableau 4.1: Parametric estimation performance of the proposed c-GEM on synthetic data, in terms of the relative error (in %).

on the overall generated histogram database. We can observe that the relative error significantly increases with the number of phases of the model but remains acceptable. The c-GEM may yet unfortunately converge to a local minimum.It typically arises because the $\hat{\alpha}_s$ exploration has not been performed properly, especially when several Poisson processes share very similar parameters values, either it is the starting point, the ending point and/or the intensity parameter.

4.4.2 Real traffic traces, simulated SPAD data

A SPAD TCSPC simulation platform has been derived from the model described in [9,256] to provide "pseudo-realistic" pixel-wise histogram data from depth-luminance scenes of the Middlebury dataset [257] as illustrated in Fig. 4.3. The estimation of the parameters of $m\mathscr{E}$ allows the extraction of multiple physical scene characteristics, such as depth inferred from the $\mathscr{E}$ shift position $\boldsymbol{\alpha}$ and the luminance from the $\mathscr{E}$ intensity parameter, $\boldsymbol{\lambda}$. Figs. 4.4 and 4.5 report the histograms associated to the color-framed pixels in the corresponding reconstructed depth and luminance images. Note that all Poisson sub-processes refer to two illumination sources, the background illumination and the laser as described in the model of [258]. In addition, because of the identification of 4 photon flux variations involved by the laser pulse, the background light distribution is somehow subdivided into 4 parts (*i.e.*, before, during, after the laser pulse, and after the "Dead Time" cf., Figs. 4.4 and 4.5).

The joint reconstructions using our proposed algorithm are shown in Fig. 4.6. Depth-luminance reconstruction accuracy for the considered scene has been respectively estimated to 0.03m root mean square error (RMSE) and 24dB peak signal to noise ratio (PSNR). Tab. 7.4 reports a comparison with respect to prior works of [10] and [11] that respectively exhibit a 0.9m and 1.28m RMSE (depth) and a 16.2dB PSNR and 13.2dB PSNR (luminance) under 0.10 SBR. [10, 11] actually fail due to their inappropriate model neglecting the pile-up effect which has been taken into account for our simulations. Indeed, these techniques were specifically designed for uniform time-correlated Poisson noise, an irrelevant assumption in case of pile-up. More specifically, regarding [10, 11] depth reconstruction under low SBR, the accumulation of photon counts completely hides the background object (due to the inverse square law of photons quantity). Indeed, it unfortunately leads to shorter distance detection, explaining the detection of many foreground objects (*i.e.*, dark pixels in Fig. 4.6). In addition, incorrect modeling of the temporal distribution of the number of photon arrival times from the laser pulse itself (*i.e.*, exponential distribution poorly modeled by a Gaussian distribution) inevitably leads to an offset error in the estimation of the mean as observed in rows 1, 2, 3 and columns 2, 3 of Fig. 4.6, for example for the estimation of the teapot depth. Moreover, [10, 11] reconstruction algorithms based on the uniform-Gaussian

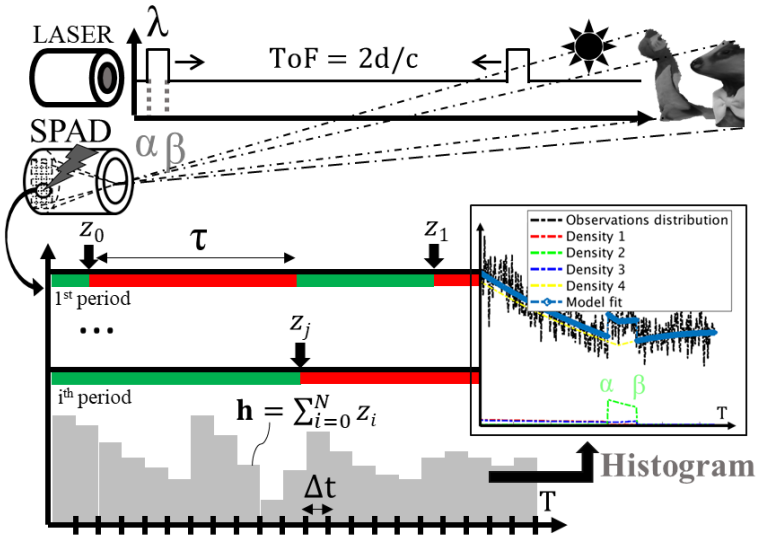


Figure 4.3: SPAD system overview.

model are not able to correctly estimate the luminance map because they do not take into account, in their Gaussian distribution model, the exponential rate parameter that encodes the luminance information with respect to our physically-plausible model.

Instead, despite of the considered complex histogram model, our EM algorithm provides high fidelity depth-luminance reconstruction under various SBRs using the appropriate distribution model and an efficient genetic-based expectation-maximization (c-GEM) method to properly manage non-convex issues. Regarding our depth reconstruction performance, the RMSE metric is not truly representative between each of our reconstructions under different noise levels because of salt-and-pepper like noise (*i.e.*, sparsely occurring white and black pixels). In addition, luminance reconstruction is more complex under high SBR (less photon noise) since it is estimated from the decaying shapes of the background illumination photon counts as well as the laser pulse photon counts. Compared to [259, 260] works, our model includes a shift estimator, $\boldsymbol{\alpha}$ that advantageously enables the proposed pixel-wise algorithm which appears to be compatible with the specific problem of depth-luminance reconstruction from TCSPC data. In addition, better reconstruction performances –for a concrete deployment of the algorithm– could be achieved taking advantage of a spatial collaboration, such as using any spatial regularization technique as Total Variation [10].

	Intensity (PSNR in dB)			Depth (RMSE in m)		
SBR	Shin [10]	Rapp [11]	Our	Shin [10]	Rapp [11]	Our
0.31	17.07	11.26	29.98	0.50	0.47	0.03
0.10	16.17	13.16	24.10	0.88	1.28	0.03
0.06	15.49	14.09	14.23	0.94	1.38	0.03

Tableau 4.2: Quantitative comparisons of average depth-luminance reconstructions of the Middlebury dataset [257] under various SBR and from the optimal photon regime of each methods (i.e., low photon counts for [10, 11] and high photon counts for our custom c-GEM algorithm.

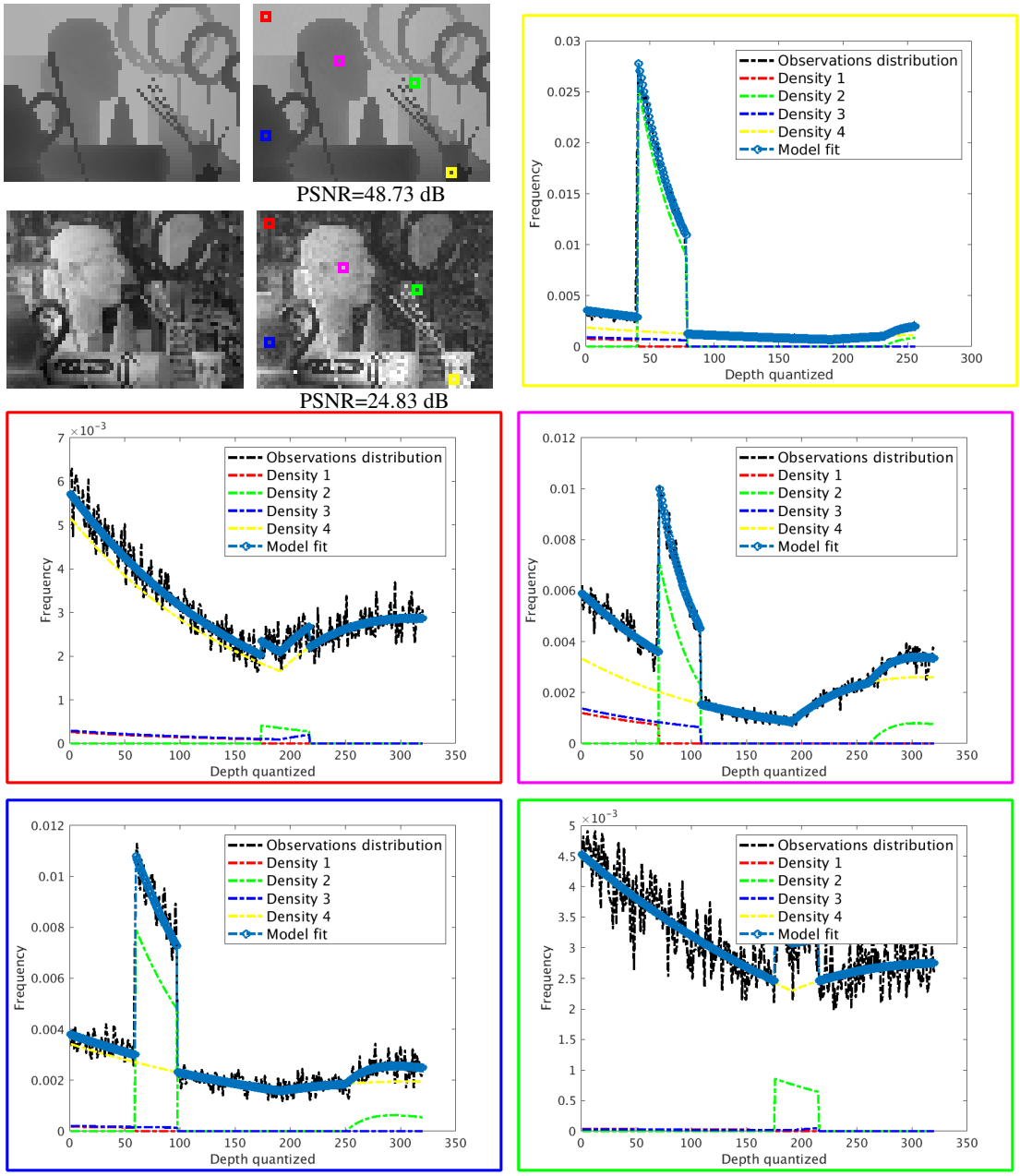


Figure 4.4: Depth-Luminance estimation from $2 \cdot 10^4$ photons ToF samples size. Note that third and forth image rows correspond to the pixels histograms surrounded in color in the upper image reconstructions.

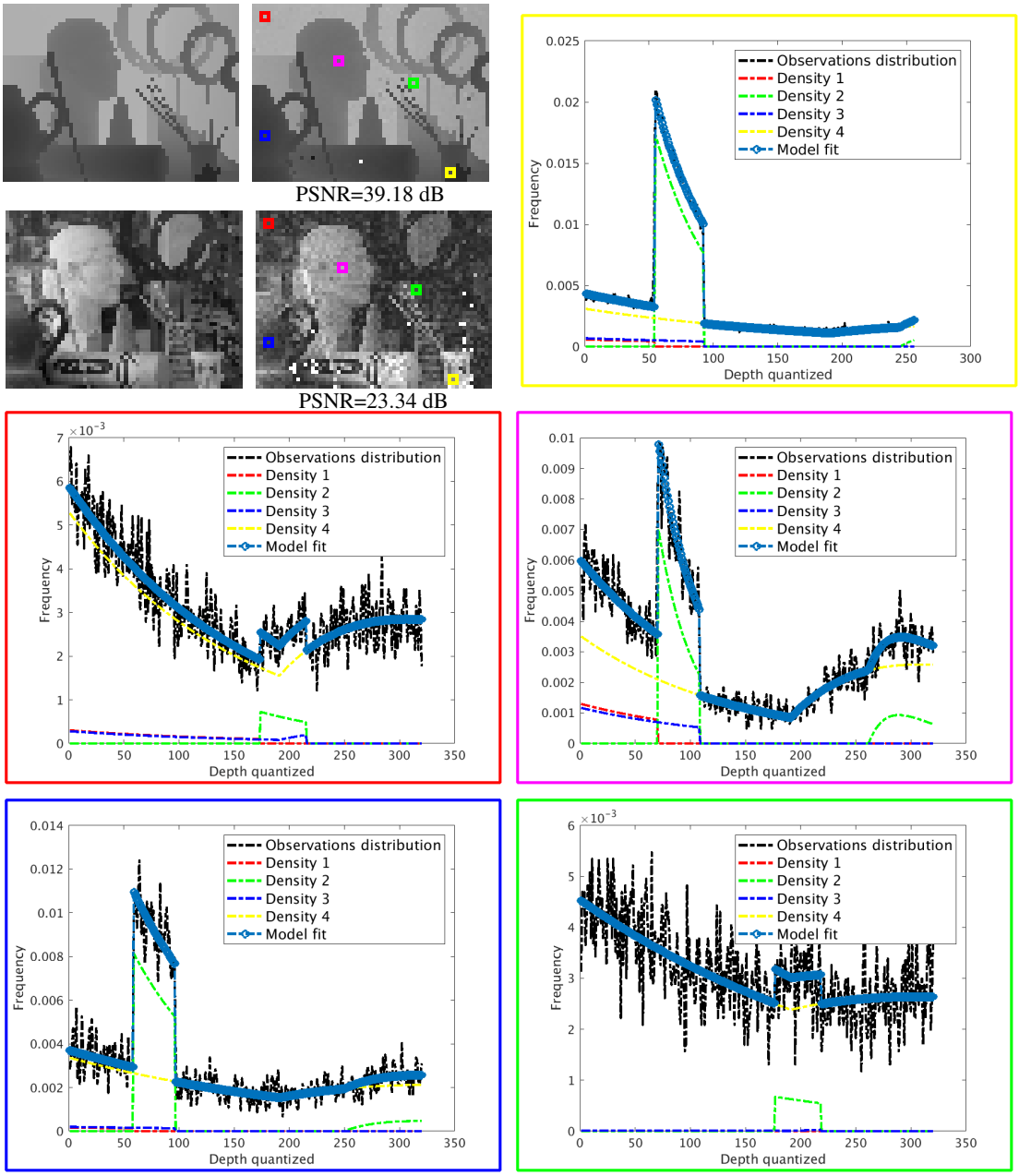


Figure 4.5: Depth-Luminance estimation from $5 \cdot 10^3$ photons ToF samples size. Note that third and forth image rows correspond to the pixels histograms surrounded in color in the upper image reconstructions.

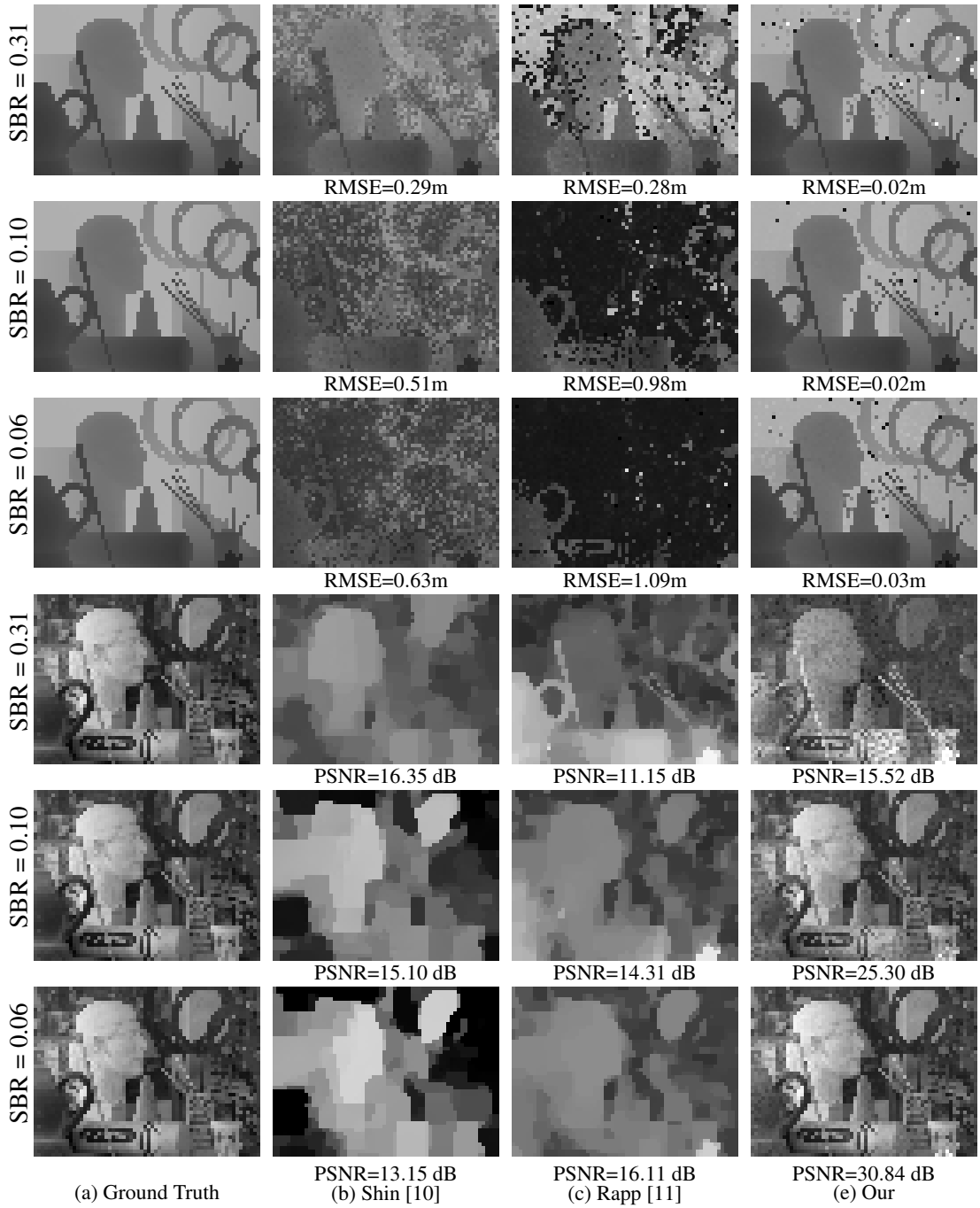


Figure 4.6: Depth-luminance reconstructions. Note that the 2^{nd} and 3^{rd} column reconstructions are under low photon counts (i.e., $100-500$ photons ToF samples size) since they were designed for these typical photon regimes and do not work in the high photon regimes. While the fourth column reconsctructions are under high photon counts (i.e., $5\cdot10^3$ photons ToF samples size).

4.5 Conclusion

This chapter focused on the statistical analysis of time distribution output by complex queueing systems, involving PCPP inputs. The extension of the EM algorithm for fitting mixtures of right truncated Shifted Erlang ($m\mathscr{E}$) probability distribution functions has been thus introduced. Although estimating such a model involves high computational complexity of the algorithms, the benefit of the high parametrization enabled by $m\mathscr{E}$ is that it provides a general model of Phase-type distribution (PH) with very flexible shapes for numerous real use cases. In order to accurately estimate the number of phases in the mixture distribution model as well as the 4 latent parameters of each component, a custom genetic based Expectation-Maximization (c-GEM) algorithm has been proposed. The core algorithmic inner loop implements a set of tools such as Maximum Likelihood Estimation [261], Moments Matching [262], Parametric Imputation [263] and Variable Neighborhood Search [264]. An outer loop additionally aims at inferring the number of model components thanks to a Genetic search approach [265] based on iterative statistical tests.

Besides providing high accuracy of depth-luminance reconstructions under various SBRs, the proposed algorithm has the advantage of being compatible with a wide range of complex models, for instance involving the detection of multiple object reflections, for super-resolution imaging or enabling multi illumination sources classifications (*e.g.*, SPAD sensors sharing memory and electronics circuitry). This parametric estimation of the PH mixture model has been shown to be relevant in the context of a specific TCSPC imaging system (long laser pulses with a large Dead Time). But the arrival process assumptions seem to span a fairly broad class, so several other settings can be addressed *e.g.*, phone calls arriving [266], patient arriving to the hospital [267], portfolio credit risk [268]. Unfortunately, because the synchronous SPAD operating mode implies complex processing method, in a view to replace TCSPC system at pixel level, the following proposed solution in chapter 5 only consider the high photon counts asynchronous SPAD operating mode resulting in Gaussian-Uniform statistical SPAD data model.

5

Expectation-Maximization Algorithm and Hardware Co-design

157

Time-Correlated Single Photon Counting (TCSPC) sensors based on the combination of Single Photon Avalanche Diodes (SPADs), Time to Digital Converters (TDCs) and histogram-builders have become a standard for depth imaging using a direct Time of Flight (D-ToF) sensing. However, one of the main bottleneck of this working mode is intrinsically due to the large amount of data required to generate pixelwise ToF histograms. Therefore, aiming at replacing TCSPC system at pixel level, the proposed solution in this chapter initially considers only the high photon counts asynchronous SPAD operating mode, resulting in Gaussian-Uniform statistical SPAD data model (cf. Fig. 5.1). Consequently, to overcome this limitation, in contrast to Chap. 4, a simpler 2-stage Expectation Maximization (EM) algorithm for online peak detection with its Register Transfer Level hardware implementation is proposed. Based on the signal processing background provided in Chap. 3.1, the online estimator consists in sequentially computing the statistical parameters that directly encode the latent spatial distance and intensity information of the illuminated scene. The evaluation on synthetic data demonstrates that with a compression ratio of 0.4%, our approach can efficiently reduce the pixel pitch by around 40% without sacrificing timing resolution compared to prior works.

Contents

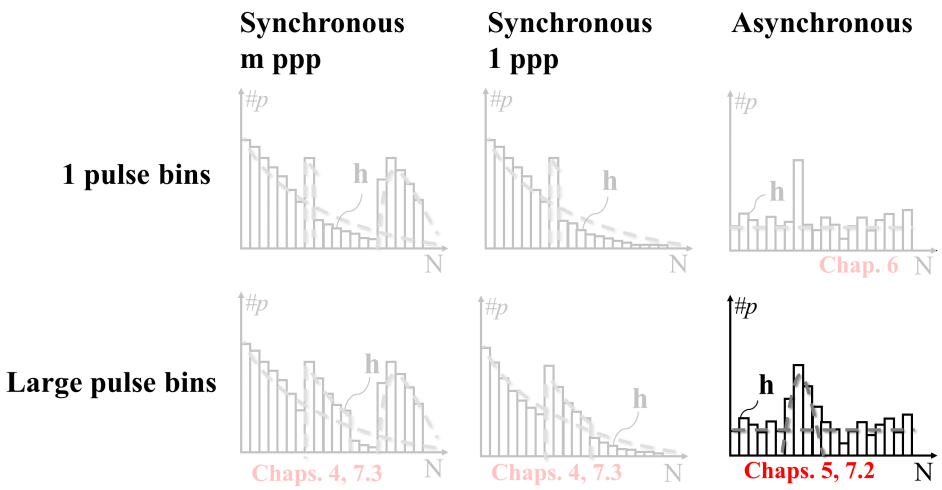


Figure 5.1: ***TCSPC distribution modeling assumptions throughout this chapter.***

5.1 Summary of existing SPAD sensing methods

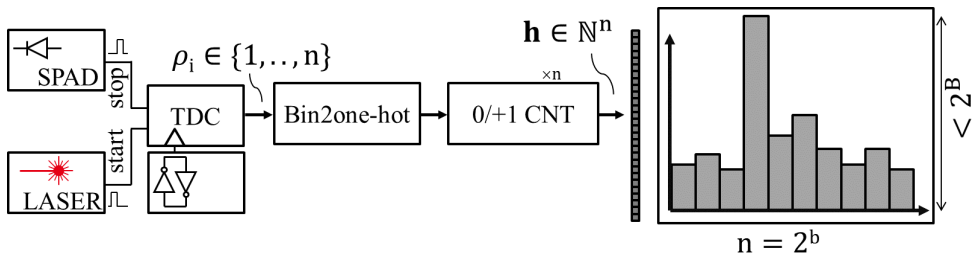
Unconventional SPAD TCSPC readout architectures emerge to cap the overall amount of data. For instance, [70] introduces "Folded inter-frame Histogram" (FiFH) which consists in building two smaller histograms, one representing the most significant bits (*i.e.*, a coarse temporal resolution) while the other is for the less significant bits (*i.e.*, a fine temporal resolution). Otherwise, [71] even proposes "Shifted inter-frame Histogram" (SiFH) to filter the second histogram using the estimated result from the first, to increase the Signal to Background Ratio (SBR) and thus the accuracy of depth estimation. Likewise, [269] implements an in-pixel zoom histogramming TDC architecture inspired by dichotomy partitioning where the bin equivalent duration is thus sequentially shrunk by half. Unfortunately, these methods imply a per-pixel complex SPAD scheduling while discarding most of the information except the peak position.

Even though peak detection is the most commonly used approach to retrieve depth information [270], advanced statistical tools have been proposed to improve system performances (*e.g.*, Gaussian fit methods [271]). Parametric estimation techniques have been developed to extract relevant features or even to reduce data throughput. Indeed, SPAD sensors intrinsically provide multimodal information while the scene characteristics shape the measured ToF histogram. Doing this, [10, 93, 272–274] present various reconstruction methods to estimate depth and reflectivity from SPAD raw data using a maximum likelihood problem definition, working under low and high flux. Recently, [8] even proposed a histogram compression framework via a sketching approach [275]. Although these algorithms provide good reconstructions, they still require either the full data histogram building or the entire photon arrival sequence storage. Therefore, when it comes to embedded systems, this approach remains highly challenging due to its computational complexity and large memory requirements.

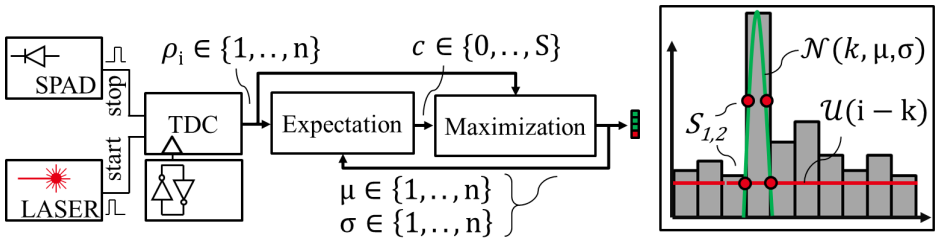
EM algorithms have already been proposed with hardware implementations for high-performance and energy-efficient designs of circuits and systems. For example in background identification, [276] proposed an implemented version of the EM algorithm that allows a fast initialization of the background model and also performs real-time background identification on high definition video sequences. The proposed approach relies first of all on the precise selection of the word length of the signals for each parameter while minimizing the error induced in the algorithm. Secondly, the standard mixture model algorithm implementation from Open source Computer Vision software library (OpenCV) involves the use of nonlinear functions. Since the nonlinear operations would be too complex, the common technique to reduce the logic for the nonlinear computations is the use of look-up tables directly stored in ROM circuits. In the same manner, [277] proposed an implemented version of the EM algorithm for pattern recognition involving a compromise between the requirements of high performance, real-time processing and hardware resources. The performance optimization in these application use cases is achieved by using a serial-parallel architecture, an efficient pipelining strategy, and a piecewise linear function to replace the most critical operations of the classifier, which is the exponential calculations. Unfortunately, these methods are not directly applicable to TCSPC sensors due to sequential data acquisition and not on a single image. In this chapter, we thus propose hardware implementation tricks for a 2-stage online EM algorithm dedicated to TCSPC SPAD data.

5.2 TCSPC imaging, system-level considerations

D-ToF sensing strategy consists in measuring the propagation time of a light pulse, emitted by a transceiver (*i.e.*, LASER), and then received by a synchronized receiver (*i.e.*, SPAD). Thanks to a high frequency Time-to-Digital Converter, every ToF $\rho_i \in [1,n]$ with $i \in [1,I]$ is stored in a ToF histogram, where the number of photon arrivals over each measured time period is accumulated in the corresponding bin of the histogram, **h**, as depicted in Fig. 5.2 (a).



(a) TCSPC baseline (straightforward approach)



(b) Classified Online-Expectation Maximization CO-EM (ours)

Figure 5.2: System-level views of baseline (a) and our variant (b).

To solve the on-chip memory issues, we propose to use a latent representation of the histogram, avoiding the need of an embedded full-scale histogram builder. The proposed method is thus based on a mixture model of distributions that provides a proper statistical representation of the collected ToF data with only a few parameters. Note that, this distribution of the observations may be corrupted by non-linear distortions (*i.e.*, the pile-up effect), but it can be efficiently bypassed thanks to an asynchronous SPAD control [72], regardless of the operating mode at low or high photon flux. In other words, temporally misaligning SPAD measurements windows smooths out pile-up distortions due to the Poisson ToF generation process [278]. Eq. 5.1 defines the considered mixture model as the weighted sum of a uniform distribution $\mathscr{U}$ due to the background illumination (π_u-weighted) and a normal distribution $\mathscr{N}$ due to the signal of interest (π_g-weighted).

$$\mathrm{P} \sim \mathscr{N}(\mu, \sigma^2, \pi_g) + \mathscr{U}(\pi_u) \tag{5.1}$$

Thanks to this model, the statistical distribution of the random ToF variable P can be described by only 4 parameters $(\mu, \sigma, \pi_g, \pi_u)$, leading to an equivalent measurement ratio of 0.4% for n=1024 (*i.e.*, Compression Ratio, CR=$\frac{4}{n}$ when considering a unique data coding). The following sections address the problem of estimating these 4 parameters during the acquisition, in order to efficiently replace the storage of histograms and its related hardware dedicated resources.

5.3 EM algorithm co-optimization

5.3.1 From real-valued responsability to bilevel classifier

The Expectation-Maximization (EM) algorithm [116] is now a canonical approach for estimating the parameters of density mixture models. It consists in maximizing a likelihood function associated to the observed data with respect to the model. The EM algorithm alternates between the E-step that computes the expectation over the conditional distribution of the latent data given the observations and the M-step that maximizes the likelihood function for the considered mixture (*e.g.*, Eq. 5.1). The main constraints of an offline EM algorithm are related to the large data observations requirement (*i.e.*, full ToF histogram in our case) and its high computational complexity. To overcome these major drawbacks, an online version of the EM algorithm has been developed, consisting in inspecting each value of the dataset only once instead of computing the entire dataset multiple times. The parameters of the mixture model are thus updated for each new sample of the measured ToF in the proposed Alg. 4. Furthermore, a "classified" version has been considered to firstly enhance the convergence rate and secondly facilitate the hardware implementation. Indeed, working with a binary classification of the ToF, $c \in \{0,1\}$, rather than floating points seems highly appropriate for a hardware implementation. Furthermore, the expectation and classification steps amount to finding intersection points of the density functions in the distribution Mixture Model (see Fig. 5.2 (b)). This problem restatement enables a significant computation complexity reduction by only solving the equality $\mathscr{N}(\mu, \sigma^2, \pi_g) = \mathscr{U}(\pi_u)$, which provides the roots, namely $s1_{lw/up}$, denoting the lower and upper intersection points.

Algorithm 4: 4 pts mémoires: μ, σ, k, i

```
   // Initialization
1  k ← I_1, μ̂_1 ← n/2, σ̂_1 ← n/16
   // Pre-warming step
2  for i = 1 : I_1 do
3      α(i) ← 1 / 2^⌈(1/2) ln2(i)⌉
4      μ̂_{i+1} ← μ̂_i + α(i)(ρ_i − μ̂_i)
5  end
6  for i = I_1 : I_2 do
       // Expectation Step
7      ζ ← Scaler(i, k, σ̂_i)
8      s1_{up/lw} ← μ̂_{i+1} ± ζ σ̂_{i+1}
9      s2_{up/lw} ← μ̂_{i+1} ± ⌈ζ/2⌉ σ̂_{i+1}
10     c ← Classifier(ρ_i | s1_{up/lw}, s2_{up/lw})
       // Maximization Step
11     k ← k + (c ≠ 0)
12     α(k) ← c / 2^⌈(1/2) ln2(k)⌉
13     σ̂_{i+1} ← σ̂_i + α(k − I_1)(|ρ_i − μ̂_i| − σ̂_i)
14     μ̂_{i+1} ← μ̂_i + α(k)(ρ_i − μ̂_i)
15 end
```

ToF ρ_i being within the interval $[s1_{lw}, s1_{up}]$ (c=1) are thus classified as signal photon ToF. If not, it is classified as noise (c=0). However, it appears that this 2-class scheme degrades the overall EM performance. We consequently proposed a bilevel classification strategy, using a second threshold pair (S=2). This second threshold pair is defined by dividing the sigma-scaling ζ by a factor of 2; allowing for a compromise between a 2-class classification and the real-valued class responsibility

(cf. Eq. 5.2).

$$\begin{aligned} s_{1_{lw/up}} &= \hat{\mu}_i \pm \zeta \hat{\sigma}_i, \\ s_{2_{lw/up}} &= \hat{\mu}_i \pm \frac{\zeta}{2} \hat{\sigma}_i, \\ \text{with } \zeta &= \sqrt{4 \ln(2) \log_2 \left(\frac{\pi_g n}{\sigma \sqrt{2\pi} \pi_u} \right)}. \end{aligned} \tag{5.2}$$

The final classification-related step thus provides the coarse discrete quantization of the density function weights c according to the $Classifier()$ function, as described in Eq. 5.3.

$$Classifier(\rho_i) = \begin{cases} 2 & \text{if } \rho_i \in [s_{2_{lw}}, s_{2_{up}}], \\ 1 & \text{if } \rho_i \in [s_{1_{lw}}, s_{1_{up}}], \\ 0 & \text{, otherwise.} \end{cases} \tag{5.3}$$

5.3.2 Parameters updates via Maximum Likelihood estimation

The proposition of a Classified Online EM algorithm (CO-EM) constitutes a recursive rewriting of the computation from which the statistical parameters can be calculated in a numerically stable fashion. Once a photon arrival is defined as coming from background illumination or laser pulse (*i.e.*, classification results c), the parameters of the related distribution (*i.e.*, uniform or normal) are directly updated. Based on the Welford's online algorithm, parameters update functions are equivalent to a first-order Infinite Impulse Response (IIR) filter with an evolving $\alpha(k)$ coefficient, as written in Eq. 5.4. Note that R_i denotes the residual updates equal to $\rho_{i+1} - \hat{\mu}_i$ and $(\rho_{i+1} - \hat{\mu}_i)^2 - \hat{\sigma}_i$ respectively when $\hat{\theta}_{i+1}$ refers to $\hat{\mu}_{i+1}$ and when $\hat{\theta}_{i+1}$ refers to $\hat{\sigma}_{i+1}$ (cf., line 13 and 14 in Alg. 4).

$$\hat{\theta}_{i+1} = \hat{\theta}_i + \alpha(k) R_i. \tag{5.4}$$

The standard deviation estimate $\hat{\sigma}$ is yet doubly biased, firstly by the misclassification of the ToFs as in the estimation of the mean $\hat{\mu}$, and secondly by the poor estimation of the mean. Consequently, the choice of the filter coefficient update function is all the more critical. Therefore, we follow the suggestion of [279] claiming that tuning the filter coefficients update function as $\alpha(k) = c \times k^{-0.5}$ enables a consistent online EM (with k the measurement index and c a scalar).

5.3.3 Pre-warming step to avoid algorithm degeneracies

The proper behavior of the EM algorithm highly depends on starting states and does not guarantee to converge to a global optimum. To overcome local convergence issues, a variety of heuristic or metaheuristic approaches exist. By concern of limited storage and computation complexity, a pre-warming initialization step, based on the Circular Mean (CM) estimation method [280], has been introduced in order to coarsely initialize $\hat{\mu}$. This CM estimator consists in centering all ToF measurements around the current estimation of $\hat{\mu}$ modulo the bounded support, $[\hat{\mu} - \frac{n}{2}, \hat{\mu} + \frac{n}{2}]$ (note that during this step, every ToF is considered as signal of interest measurements). Doing the CM instead of a standard mean estimation method thus makes the distribution of the noise measurements unbounded and allows to refine the initialization of $\hat{\mu}$ without being biased by the bounded nature of the ToF value support.

5.4 CO-EM, a hardware-friendly architecture

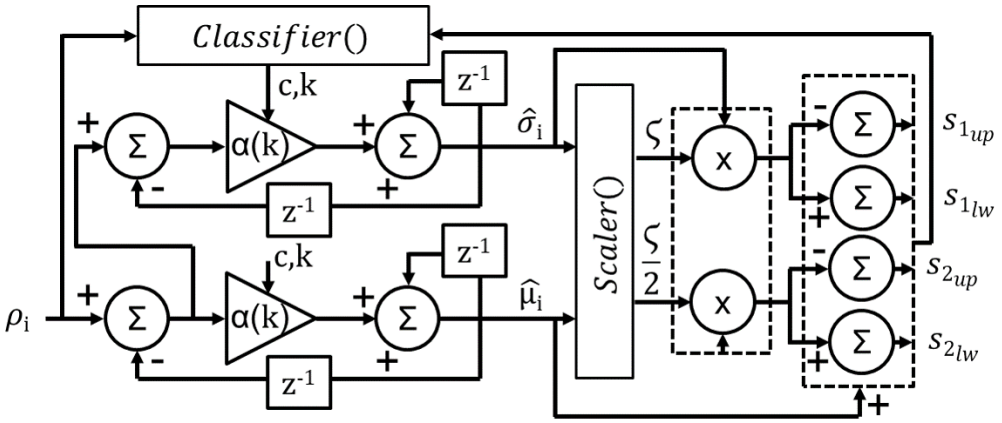


Figure 5.3: Block diagram of the proposed CO-EM circuit.

One of the major challenges when implementing the CO-EM solution is related to operator complexity. The integer approximation of the computation of floating-point operators seems to be a good candidate for a tiny digital implementation. Therefore, we proposed hardware-friendly piecewise constant functions to approximate the two complex operations that are the filter coefficient update function (Eq. 5.4) and the calculation of the threshold (Eq. 5.2).

5.4.1 Filter coefficient update function approximation

Since the filter coefficient update function is a division by a square root, only ceiling the square root function would lead to a complex implementation with an unnecessary high precision of approximation. Based on the square root notation in its logarithmic form (*i.e.*, $\sqrt{k} = 2^{\frac{log_2(k)}{2}}$), the approximation of Eq. 5.4 only consists of ceiling the logarithm plus one (operation denoted by $\underline{l_2}(k) \approx log_2(k)$) and ceiling its division by two (cf., line 3 and 12 in Alg. 4). This amounts to a simple bitshift of the residual R_i by the half position of the most significant bit of the ToF counter k plus one as illustrated in Fig. 5.4.

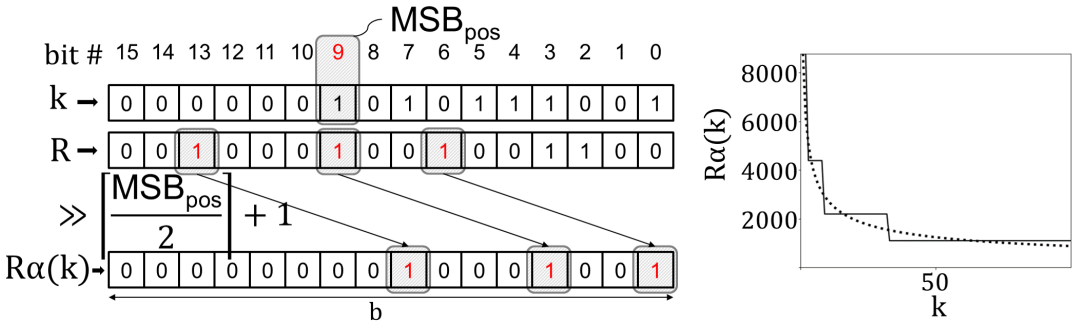


Figure 5.4: Effects of square root division approximation for the filter coefficient update function. The dotted line is for the exact inverse square root function and the solid line is for its proposed piecewise constant approximation function.

5.4.2 Threshold estimate approximation

The calculation of ζ is also based on the square root function but requires a higher precision than a single coefficient to the power of two. Therefore, based on this notation, $\sqrt{x} = x2^{-\frac{1}{2}log_2(x)-1} + 2^{\frac{1}{2}log_2(x)-1}$ and considering the limited dynamic range of x ($\zeta \in [1,5]$ in Eq. 5.2 so $x \in [1,25]$), variations of $\frac{1}{2}log_2(x)$ can be approximated by its expectancy ($\frac{1}{2}log_2(x) \approx 2$). To considerably reduce the computational complexity, ζ is computed using the linear function in Eq. 5.5 being illustrated in Fig. 5.5 for various SBR and $\hat{\sigma}$ values. Note that SBR represents the ratio of the amount of signal ToF measurements k over all ToF measurements i. To that end, a new set of constants $\frac{1}{2}$, 1, b are defined instead of $\ln(2)$, $\ln(\sqrt{2\pi})$, $log_2(n)$). Moreover, the normalized weights of the density functions $\pi_g = \frac{k_g}{i}$ and $\pi_u = \frac{i-k_g}{i}$ were simplified by removing normalization-related computations.

$$Scaler\left(\hat{\theta}\right) = \left\lceil \frac{b+7-\underline{l_2}(\hat{\sigma}_i)+\underline{l_2}(k)-\underline{l_2}(i-k)}{4} \right\rceil \tag{5.5}$$

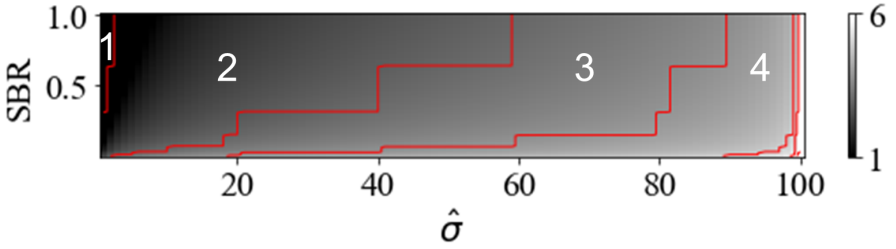


Figure 5.5: Effects of the approximated function Scaler() that corresponds to the region border red lines displayed above while the gray gradient represents the Floating Point results of the exact square root computations.

5.5 RTL synthesis and simulation results

The proposed CO-EM sensing strategy (cf. Fig. 5.3) has been evaluated on a synthetic dataset simulating a single pixel response peak per scene. We chose the parameters of the simulation model according to the ones reported in [6, 272]. Laser parameters are based on a 671-nm Picoquant pulsed laser with a pulse duration of $\approx$ 100ps and a repetition rate of 150MHz, 15MHz leading respectively to an hypothetical dynamic range of 2m and 20m. On the SPAD sensor side, the number of time bins, n, was set to 1024 leading respectively to an hypothetical depth precision of $\approx$ 2mm, 20mm and a time bin resolution of $\approx$ 6.5ps, 65.1ps. In order to only benchmark the robustness of the proposed approach, the correlation between the SBR and the measured depth was not taken into account. Doing this, the proposed sensing strategy can accommodate long depth measurements (*e.g.*, 100m) at 300 frames per second as well as shorter depth measurements at a higher frame rate (*e.g.*, 15000fps) with a SBR ranging from 0.05 to 0.5; which reflects sensing conditions for high and low-noise scenes.

We then ran a Monte-carlo simulation with 1000 trials of stochastically drawn histograms composed of I=10000 ToF measurements to evaluate and compare the performance of our 2-stage EM approach (CR=0.4%) against state-of-the-art approaches *i.e.*, histogram peak detector (Peak) (n=1024, CR=100%), coarse to fine histogram (SiFH) [71] (n=32, CR=6.25%), folded histogram (FiFH) [70] (n=32, CR=6.25%), CM estimator (CR=0.1%). The Compression Ratio (CR) refers

to the size of the measurement vector (or the number of parameters) that allows for a complete reconstruction of the histogram over the uncompressed TCSPC measurement vector size (*i.e.*, n=1024 bins). The number of iterations of the pre-warming CM step was set to $I1$=2500, while Tab. 5.1 reports results after 2500 additional CO-EM iterations.

Dyn. (m)	SBR	Previous works						CO-EM (ours)			
		Peak	CM	FiFH [70]		SiFH [71]		Float32		Int16	
				B=8	B=16	B=8	B=16	S=1	S=2	S=1	S=2
20	0.5	0,60	17.6	7.31	3.32	1.71	0.67	0.41	0.44	0.36	**0.34**
	0.1	0,83	54.8	7.78	5.07	3.39	0.91	23.3	0.50	18.4	**0.47**
	0.05	0.93	82.2	10.4	6.45	5.68	1.09	26.0	0.88	24.9	**0.51**
2	0.5	3,37	17.4	13.06	13.1	7.58	3.92	2.13	1.82	1.79	**1.51**
	0.1	5,17	53.5	13.04	12.7	5.97	5.72	32.8	2.69	23.3	**2.27**
	0.05	6.78	83.7	14.85	13.1	15.28	7.43	41.2	8.12	33.8	**8.17**
Pitch (µm)		389	-	50.3	68.8	50.3	68.8	-	-	-	**44.7**

Tableau 5.1: Comparisons of several depth reconstruction methods under various SBR. Note all results are reported as an average RMSE on binned data.

Hardware architectures presented in Fig. 5.2 were synthesized using SYNOPSYS® Design-Compiler™ using a 40nm standard cell technology to obtain the area results reported in Tab. 5.1.For high SBRs, CO-EM outperforms even a standard peak detection, benefiting from inter-bin collaborations (peak width larger than a bin) while reducing the pixel pitch by a large factor. The CO-EM architecture –with an estimated pitch of 44.7µm and a CR of 0.4%– provides better performance in terms of RMSE than all reported previous works allowing an estimated pixel pitch reduction of around 35% compared to [70, 71] (with n=32 number of bins, and B=16 counter bit width) and an estimated pixel pitch reduction of 10% even when B=8.

5.6 Conclusion

In this chapter, we proposed a novel 2-stage online EM algorithm to handle the major hardware limitations causing a data storage bottleneck. Our approach introduces an online parameter estimation instead of building the whole ToF pixel-wise histograms. To highlight the interest of the proposed solution, quantitative results demonstrate a superior depth reconstruction accuracy compared to previous works while enabling a reduction of the pixel pitch. In addition, compared to [70] and [71], CO-EM outputs parameters allow for a direct estimation of the albedo and the SBR, not only depth. The next chapter also addresses the problem related to the TCSPC data involving large pixel pitch in other SPAD configurations through a completely different approach based on compressive sensing theoretical background presented in Chap. 3.2.

6

Histogram Compressive Sensing using Shuffled Cellular Automata

The proposed approach in Chap. 5 may encountered some convergence problem that can't be fixed. Therefore, a conversion from high-dimensional TCSPC data, **h***, to low-dimensional data in the form of a vector in such a way that the two are semantically similar is presented based on compressive sensing. This chapter presents a pixel-wise histogram CS scheme from its mathematical formulation to its possible implementation using an in-pixel Cellular Automaton (CA), replacing the counter-based TDC. In addition, we propose a dedicated reconstruction algorithm to handle the specifics of TCSPC histograms. In order to support reconstruction, numerical simulations depending on SNR and number of measured ToF N have been conducted and reported under various configurations. Note that, like other unconventional TCSPC histogram acquisition strategies, our CS scheme relaxes hardware memory constraints but at the cost of a computationally intensive reconstruction. However, the point is that this approach does not limit the remote interpretation of the collected data to a simple peak detector but allows for understanding more complex underlying information, for example using deep learning tools as it is claimed in Chap. 7.*

Contents

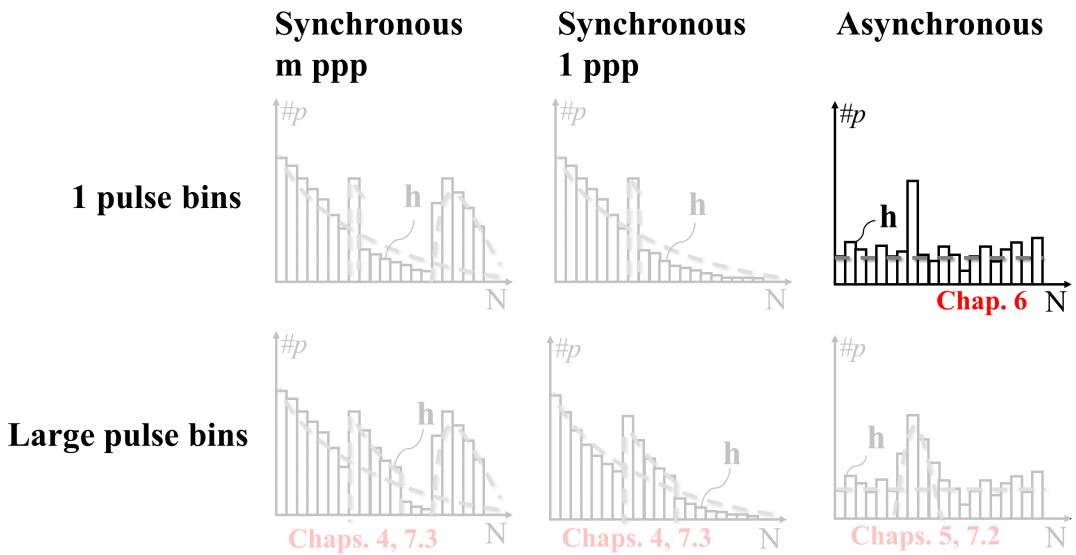


Figure 6.1: ***TCSPC distribution modeling assumptions throughout this chapter.***

6.1 ToF histogram Compressive Sensing

6.1.1 Histogram CS, underlying mathematics

Mathematically speaking, we denote the Compressive Sensing of a measured ToF histogram $\mathbf{h} \in \mathbb{N}^{\mathrm{n}}$ as a linear projection using the matrix $\mathbf{\Phi} \in \{-1,+1\}^{\mathrm{m}\times\mathrm{n}}$ (Rademacher-like distribution considered for its universality property), providing the CS measurement vector $\mathbf{y} \in \mathbb{Z}^{\mathrm{m}}$ (Eq. 6.1). Since the histogram is generated under the assumption of an asynchronous SPAD control [72], it can be written as a sum of the signal of interest $\mathbf{s}$ and an independent identically distributed (i.i.d.) Poisson noise vector $\boldsymbol{\nu}$ (without any kind of pile-up nor dead-time effects). Thanks to this model, $\mathbf{s}$ is considered k-sparse (even 1-sparse if assuming a single laser reflection per pixel, cf. Fig. 6.1).

$$\mathbf{y} = \mathbf{\Phi}\mathbf{h} = \mathbf{\Phi}(\mathbf{s}+\boldsymbol{\nu}). \tag{6.1}$$

From the sensor point-of-view, a raw D-ToF measurement $v(\boldsymbol{\rho}_i) \in \{1,...,n\}$ provided by a TDC is a time index corresponding to a one-hot vector $\boldsymbol{\rho}_i \in \{0,1\}^{\mathrm{n}}$ for the i^{th} ToF acquisition, $i \in \{1,...,N\}$. In order to actually build the pixel-wise histogram $\mathbf{h}$, this vector $\boldsymbol{\rho}_i$ only needs to be summed over N TDC successive acquisitions thanks to its intrinsic position coding (*i.e.*, $\mathbf{h} = \sum_{i=1}^{N} \boldsymbol{\rho}_i$). $\boldsymbol{\rho}_i$ has a unique non-zero coefficient equal to 1 at the position $v(\boldsymbol{\rho}_i)$ and knowing that the multiplication by $\mathbf{\Phi}$ is distributive with respect to the addition; building $\mathbf{y}$ is equivalent to calculating the sum over i of the columns $\Phi_{v(\boldsymbol{\rho}_i)}$ of $\mathbf{\Phi}$ at $v(\boldsymbol{\rho}_i)$ positions (cf. Eq. 6.2).

$$\mathbf{y} = \mathbf{\Phi}\sum_{i=1}^{N}\boldsymbol{\rho}_i = \sum_{i=1}^{N}\Phi_{v(\boldsymbol{\rho}_i)} \tag{6.2}$$

It thus enables a direct acquisition of CS measurements without the need for an explicit representation of $\mathbf{h}$ at any time, making the approach highly relevant in terms of hardware implementation. It means that the only requirement is to replace the one-hot encoding of the measured ToF by a said "chaotic" encoding (*i.e.*, the columns of $\mathbf{\Phi}$, $\Phi_{v(\boldsymbol{\rho}_i)}$).

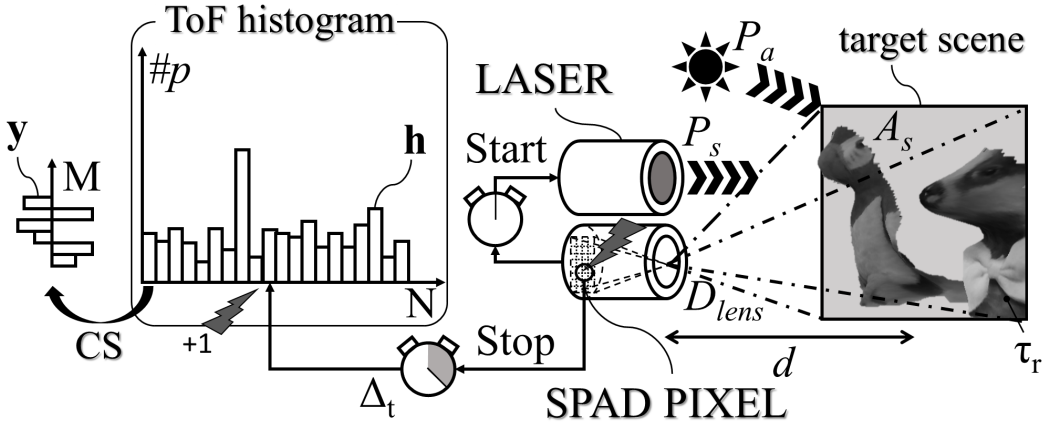


Figure 6.2: TCSPC D-ToF SPAD Operation System Overview with it CS-TCSPC counterpart illustration.

6.1.2 On the use of Shuffled Cellular Automata

The main goal of this chapter is to demonstrate that an on-the-fly compressive acquisition of SPAD time-of-flight histogram can be performed using a basic Cellular Automaton replacing the

commonly used TDC plus one-hot encoding. A binary CA is composed of a finite number of cells that have a single binary state at each cell and discrete time step. For a regular Elementary CA (ECA), each cell state only depends on a logic rule taking as inputs the previous states of the cell itself and its two neighbors. The main advantage of CAs is that a complex global behavior can be obtained using only very few digital logic gates. Since the CA is dedicated to the generation of a pseudo-random sequence, a slight modification has been made to the structure of the ECA rule 30. As depicted in Fig. 6.3, a simple routing for static shuffling is added to further increase the statistical independence of the binary states produced.

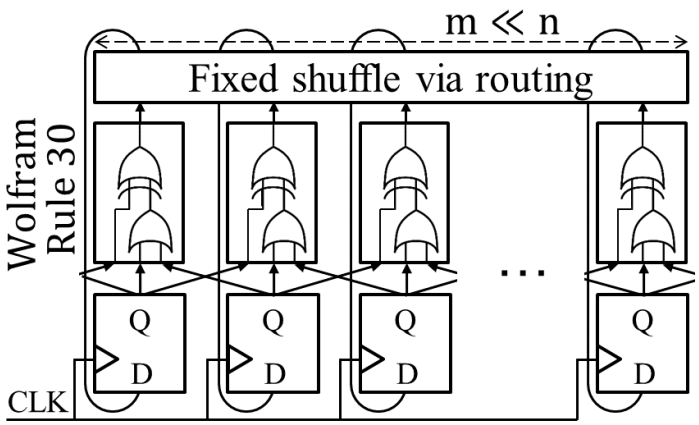
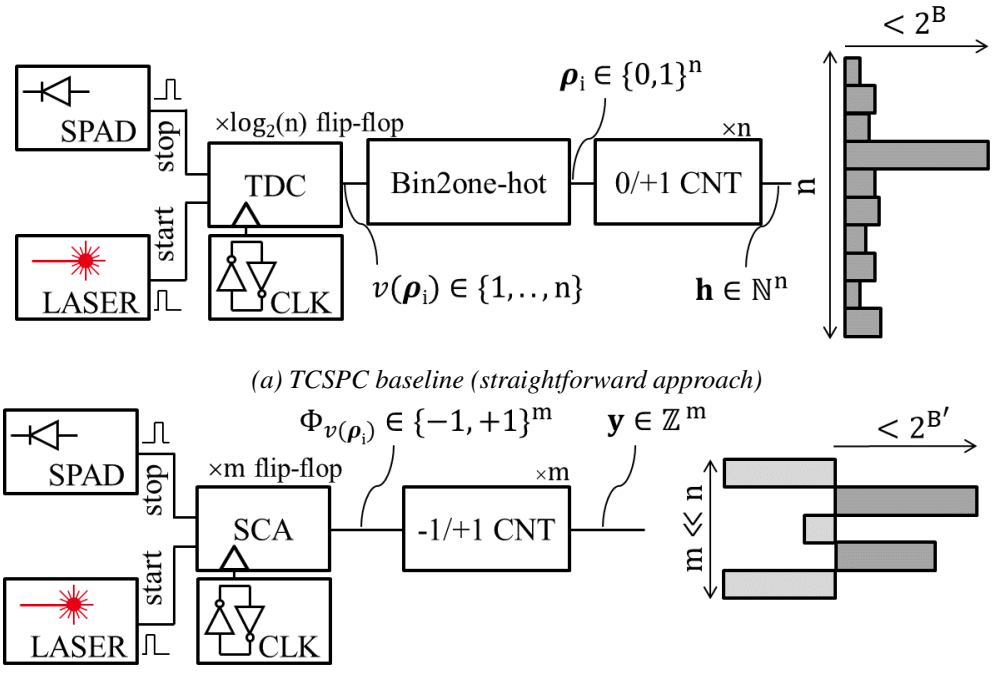


Figure 6.3: Structure of a Shuffled Cellular Automaton (SCA).

As presented in Fig. 6.4 and according to Eq. 6.2, the construction of the histogram **h** can thus be formally replaced by the direct construction of a CS measurement vector **y**. This way, the number of clock cycles between the laser shot and the SPAD trigger (*i.e.*, ToF index) is equal to the previously denoted $v(\boldsymbol{\rho}_i)$ and the state vector of the SCA is considered to be equal to $\Phi_{v(\boldsymbol{\rho}_i)}$ in its signed representation.



(a) TCSPC baseline (straightforward approach)

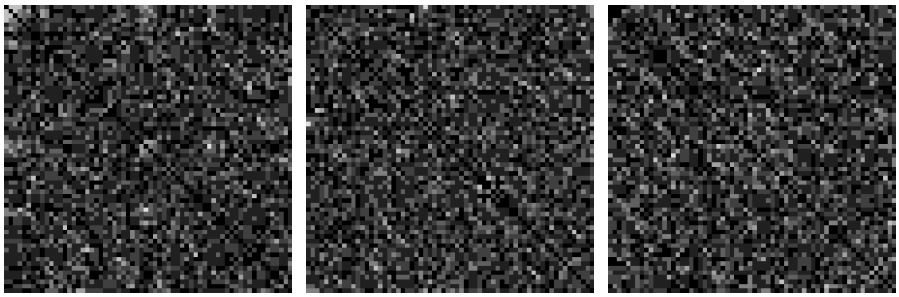
(b) TCSPC histogram Compressive Sensing (ours)

Figure 6.4: System-level views of baseline (a) and its variant (b).

The same way as a TDC starts counting from zeros, the SCA starts updating its states from the initial state triggered by the laser shot and stops on the SPAD trigger. However, this deterministic process (*i.e.*, the state sequence (columns of $\mathbf{\Phi}$) only depends on the logic rule, the shuffle map and the initial states, see Fig. 6.3) requires a proper choice of Wolfram's rule, the initialization and the shuffle pattern, to ensure having the longest possible cycle to provide a full rank $\mathbf{\Phi}$. Even if the CS community mainly focused on the Class 3 cellular automaton rule 30 which exhibits a proper chaotic behavior [281], using a shuffle stage does not imply any additional hardware component (only based on CA cells interconnections) while improving the properties of the generated pseudo-random vectors (see Fig. 6.5 for cell states ($\sim \mathbf{\Phi}$) and Fig. 6.6 for Gram matrices ($\sim \mathbf{\Phi}^T\mathbf{\Phi}$)).

(a) Rule 30 with canonical initialization. *(b) Rule 30 with random initialization.* *(c) Rule 30 SCA with random initialization.*

Figure 6.5: Cell states for various CA configurations ($\sim \mathbf{\Phi}$).



(a) $\mu_M = 0.875$, $\mu_A = 0.1959$ *(b)* $\mu_M = 0.75$, $\mu_A = 0.1956$ *(c)* $\mu_M = 0.75$, $\mu_A = 0.1938$

Figure 6.6: Normalized Gram matrix ($\sim \mathbf{\Phi}^T\mathbf{\Phi}$ with zero diagonal for proper rendering) of the sensing matrices as presented in Fig. 6.5, where μ_M corresponds to mutual coherence and μ_A is the average of the Gram matrix except the diagonal.

6.1.3 Bi-objective histogram reconstruction

On the decoder side, the signal to recover $\hat{\mathbf{h}}$ is considered to be a vector of non-negative values composed of a reconstructed signal of interest $\hat{\mathbf{s}}$ and an estimated Poisson noise $\hat{\boldsymbol{\nu}}$. In practice, additional priors are used in the Algorithm 5 to reach better reconstruction performances rather than a canonical CS approach. Firstly, $\hat{\mathbf{s}}$ is supposed to be sparse because of describing only the ToF signal from laser pulses, leading to the standard ℓ_1 regularization relaxation with parameter $\alpha_h \in \mathbb{R}^+$. Secondly, due to the intrinsic nature of a histogram, the positivity constraint has been added to $\hat{\mathbf{s}}$ and $\hat{\boldsymbol{\nu}}$. Then, since Poisson noise is consequently projected in the compressed domain in which the resulting noise distribution is almost a centered normal distribution [282], it makes the standard CS denoising technique using a ℓ_2 fidelity term applicable [283]. Finally, assuming that $\hat{\boldsymbol{\nu}}$ is a Poisson noise with sufficient amplitude (*i.e.*, close enough to a nonzero-centered Gaussian-like distribution) two additional terms are added to the $\hat{\boldsymbol{\nu}}$ estimation. The first one (parameterized by $\beta_\nu \in \mathbb{R}^+$) enforces the mean-variance equality and the other (parameterized by $\alpha_\nu \in \mathbb{R}^+$) limits the excursion of $\hat{\boldsymbol{\nu}}$ around the estimated mean, using the following matrix operators $E = [1/n;...;1/n] \in \mathbb{R}^{1\times n}$, $E_n = [E^T;...;E^T] \in \mathbb{R}^{n\times n}$, and $I_n \in \mathbb{R}^{n\times n}$ the identity.

Algorithm 5: Sparse-Poisson bi-objectives iterative reconstruction algorithm under positivity constraints.

while *stopping criterion not reached* **do**

$$\hat{\boldsymbol{\nu}} = \underset{\tilde{\boldsymbol{\nu}}}{\arg\min} \|\Phi(\hat{\mathbf{s}} + \tilde{\boldsymbol{\nu}}) - \mathbf{y}\|_2^2 + \alpha_\nu \|(I_n - E_n)\tilde{\boldsymbol{\nu}}\|_2^2 + \beta_\nu \|E(\tilde{\boldsymbol{\nu}}^{\cdot 2} - \tilde{\boldsymbol{\nu}}) - (E\tilde{\boldsymbol{\nu}})^{\cdot 2}\|_2^2, \tilde{\boldsymbol{\nu}} \succeq 0;$$

$$\hat{\mathbf{s}} = \underset{\tilde{\mathbf{s}}}{\arg\min} \|\Phi(\tilde{\mathbf{s}} + \hat{\boldsymbol{\nu}}) - \mathbf{y}\|_2^2 + \alpha_h \|\tilde{\mathbf{s}}\|_1, \tilde{\mathbf{s}} \succeq 0;$$

end

return $\hat{\mathbf{s}}, \hat{\boldsymbol{\nu}}$.

This algorithm only demonstrates the possibility of reconstruction from compressed TCSPC histograms in the context of a high level of Poisson noise. However, to further improve reconstruction performances, an alternative approach based on data-driven deep-learning could also be investigated [284].

6.2 Hardware synthesis results

n\B	4	5	6	7	8	9	10	11	12	13	14
8	19.1	20.8	22.4	23.9	25.3	26.6	27.9	29.1	30.3	31.4	32.5
16	25.3	27.9	30.3	32.5	34.5	36.5	38.3	40.1	41.8	43.4	**44.9**
32	35.3	**39.0**	42.4	45.5	48.5	51.2	**53.9**	56.4	58.8	61.1	63.3
256	97.8	108	118	127	136	143	151	158	165	171	178
1024	194	**216**	235	253	270	286	**301**	316	329	342	355

Tableau 6.1: Equivalent pixel pitch (µm) required obtained from TCSPC baseline architecture synthesis (Fig. 6.4 (a)).

m\B′	4	5	6	7	8	9	10	11	12	13	14
8	15.3	19.9	22.0	23.9	25.6	27.3	28.8	30.3	31.6	33.0	34.3
16	21.6	**28.1**	31.1	33.7	36.1	38.5	**40.7**	42.8	44.8	46.7	48.6
32	30.4	**39.7**	43.8	47.6	51.1	54.4	**57.5**	60.5	63.3	66.0	68.6

Tableau 6.2: Equivalent pixel pitch (µm) required obtained from SCA architecture synthesis (Fig. 6.4 (b)).

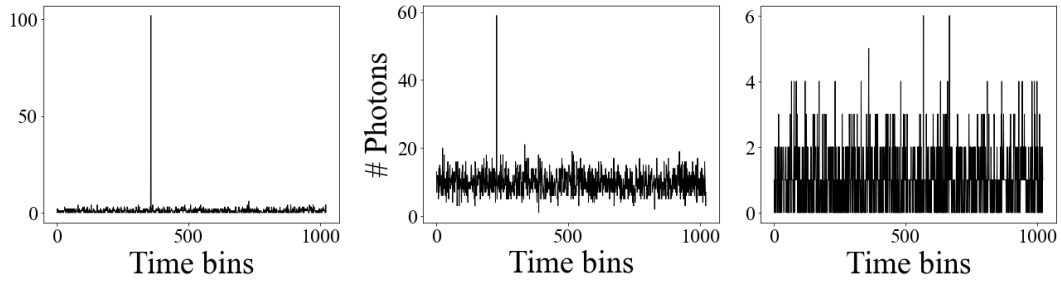
Both hardware architectures in Fig. 6.4 were synthesized using SYNOPSYS® Design-Compiler™ over a 40nm standard cell technology to obtain the area results reported in Tab. 6.1 and Tab. 6.2. For the sake of clarity, synthesis-estimated digital areas are given in equivalent pixel pitches (µm) instead of areas (µm^2). Our synthesis results are to be put in perspective against the exceedingly well optimized design of [6] that presents a tiled full custom designed histogram builder with a pitch of 36.72µm targeting 16 bins of B=14 bitwidth each. Those results which are obtained using an equivalent technology node show a size difference factor of 1.2 (*i.e.*, 44.9µm for the same configuration as in [6]). This gap would be bridged by redesigning our architecture using a full custom design methodology with specific attention to area minimization. The main takeaway is that the relative results between Tab. 6.1 and Tab. 6.2 demonstrate that SCA configurations im-

ply a marginal size increase (lower than 10%) compared to an uncompressed TCSPC baseline that explicitly builds a ToF histogram for the same number of measurements (*i.e.*, m=n) and with identical counters bitwidths (*i.e.*, B=B').

6.3 Simulated reconstruction results

6.3.1 ToF estimation under single laser reflection assumption

The proposed sensing strategy has been first evaluated for a single laser pulse per acquisition period (*i.e.*, $\hat{\mathbf{s}}$ is 1-sparse, see Fig. 6.7) in 9 configurations (250 stochastically drawn histograms for each), for three D-ToF measurements quantity N (10^3 and 10^4) and three SNR (0.1, 0.02 and 0.005) that is defined as the ratio between the number of photons related to the signal and those due to the noise.



(a) N=10^3, *SNR*=0.1 *(b)* N=10^4, *SNR*=0.005 *(c)* N=10^3, *SNR*=0.005

Figure 6.7: Illustrations of simulated histograms.

To compare our approach with a TCSPC baseline, FiFH [70] and SiFH [71], the silicon area estimation was conducted using a standard digital design flow, intrinsically limiting its ability to operate at a very high frequency (*i.e.*, $>$ 3GHz). To put in perspective our simulated D-ToF histograms shown in Fig. 6.7 with n=1024bins, it results in a time bin resolution of $\approx$ 330ps with a hypothetical depth range of 100m. Synthetic ToF measurements do not take into account the correlation between the SNR and the measured depth since the goal here is to only bench the robustness of the proposed approach.

Fig. 6.8[1] thus shows the reconstruction performance of the different methods investigated as a function of the Compression Ratio (CR) which corresponds to a certain silicon area estimation reported in Tab. 6.1 (purple for TCSPC baseline and orange for [70, 71]). The CR in Fig. 6.8[1] is related to the size of the measurement vector that allows a complete reconstruction of the histogram over the uncompressed TCSPC measurement vector size (*i.e.*, n=1024 bins). Our CS TCSPC acquisition method was evaluated for various CRs corresponding to m=16, 24, 32, 48, 64, 96. While [70, 71] approaches were evaluated for a measurement vector size of n=32. Since, they need two acquisition periods to estimate the full histogram (resulting in 64 measurements), the related CR is only 6%. Due to a noise floor effect invalidating $\hat{\mathbf{h}}$ sparsity, the standard Lasso [286] shows a low performance. Our Algorithm 5 yet succeeds to estimate the noise floor combining a Lasso with a least-squares optimizer [287]. For the reported experimental setup, we set $\alpha_h = 10^{-8}$, $\alpha_v = 10^{-4}$ and $\beta_v = 10^{-7}$ leading to less than 4 outer loop iterations with a stopping criterion set

[1] Erratum for paper [285]

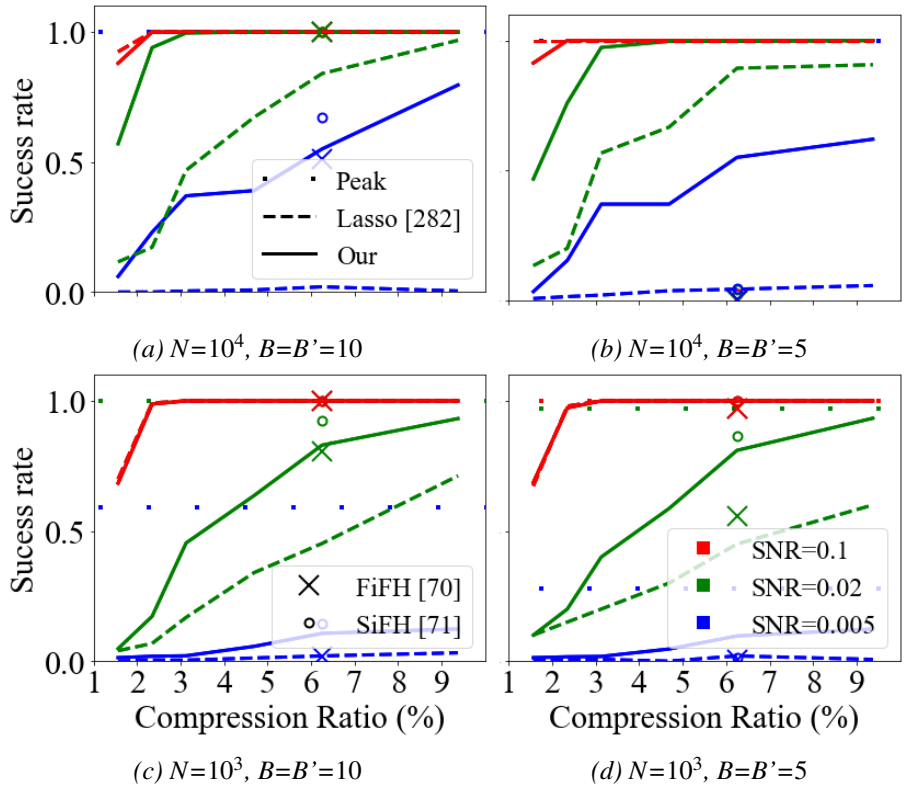


Figure 6.8: Peak detection success rate for a set of methods, SNRs, CRs and ToF measurements numbers N.

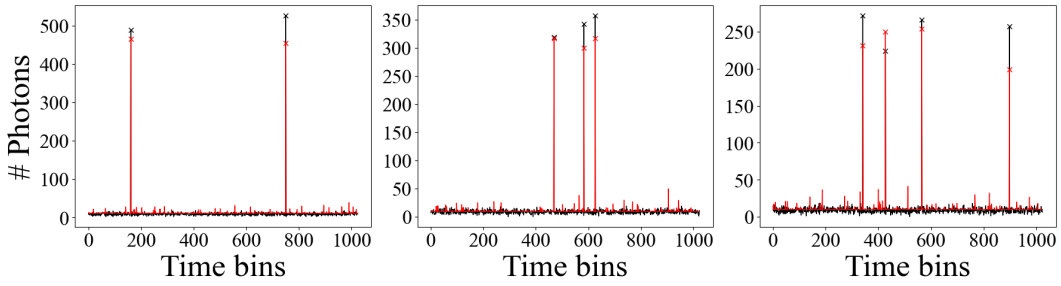
to 10^{-4} on the absolute relative change (L1-norm) of the estimated signal $\hat{\mathbf{s}}$. According to Eq. 6.2, Poisson noise is projected into the CS domain and its statistical distribution becomes Gaussian and independent of signal strength [288]. This feature advantageously enables a reduction of the number of bits B' required to capture the information in $\mathbf{y}$ (without accuracy loss) despite counter saturation issues that have been carefully taken into account in our testbench. Histogram CS allows to reach a highly accurate reconstruction, while being able to reduce both the size of measurement vectors and their bitwidths (B'=5), therefore reducing the total memory needed (acting on the pixel pitch, cf. subsection 6.2). On the other hand, when decreasing the bitwidth (*i.e.*, to B=5), we can see in Fig. 6.8 that [70,71] are significantly degraded. Tab. 6.3[1] highlights our CS approach which provides a similar peak detection success rate to [71] (B=10) and higher than [70] (B=10), with an estimated pitch reduction of 26%. It leads to a better performance at iso-surface whatever noise level while being compatible with a multiple peak detection.

6.3.2 Extension to *k*-sparse latent histograms

As reported in Tab. 6.3[1], unlike prior works [70, 71], the proposed method is compatible with *k*-sparse signals (*e.g.*, multiple laser reflections detection) without any modification of the hardware nor the reconstruction algorithm. Fig. 6.9 exhibits qualitative *k*-sparse reconstructions. Note that it could provide a better information retrieval compared to a single peak detector, for example to improve full-resolution depth reconstructions and feeding a possible super-resolution stage. Although the off-sensor reconstruction is not the main contribution of this chapter, this work paves the way to more advanced reconstruction methods [284], enabling an increase of the spatial resolution allowed by our CS TCSPC sensing scheme (*e.g.*, leading up to a 82% reduction in terms of silicon area while sharing the hardware for a subgroup of 4 pixels).

N	SNR	Previous works						CS-TCSPC (ours)			
		Peak		FiFH [70]		SiFH [71]		m=32		m=16	
		B=5	B=10	B=5	B=10	B=5	B=10	B'=5	B'=10	B'=5	B'=10
10^3	0.1	1.00	1.00	1.00	**1.00**	**1.00**	1.00	**1.00**	1.00	1.00	1.00
	0.02	0.98	1.00	0.43	**0.51**	**0.63**	0.90	**0.79**	0.81	0.40	0.45
	0.005	0.26	0.57	0.01	**0.02**	**0.04**	0.11	**0.09**	0.11	0.02	0.02
10^4	0.1	1.00	1.00	0.02	**1.00**	**0.04**	1.00	**1.00**	1.00	1.00	1.00
	0.02	1.00	1.00	0.01	**0.95**	**0.01**	0.99	**1.00**	1.00	0.97	0.99
	0.005	1.00	1.00	0.01	**0.30**	**0.01**	0.67	**0.55**	0.55	0.36	0.37
Pitch (µm)		216	301	39.0	**53.9**	**39.0**	53.9	**39.7**	57.5	28.1	40.7
k-sparse		Yes		**No**		**No**		**Yes**			

Tableau 6.3: Comparison of depth reconstruction methods in terms of reconstruction success rate, pixel pitch estimation and k-sparse compatibility for multiple peak detection.



(a) k=2 *(b) k*=3 *(c) k*=4

Figure 6.9: k-sparse signals with SNR=0.1*, N*=10^4*, m*=64*. Black is for original, red for reconstructions.*

6.4 Conclusion

CS TCSPC reduces the number of stored values and the number of bits per values, thereby reducing the pixel pitch required for an in-pixel implementation. The proposed acquisition method thus allows to reach a reconstruction performance similar to existing methods with an estimated pitch reduction of 26%. In addition, by further taking advantage of signals' structure assumptions, unlike previous works, our approach is compatible with more complex signals (*e.g.*, *k*-sparse), opening the way to more advanced reconstruction/inference strategies (using deep learning techniques [256, 289]). The next Chap. 7 will therefore introduce the use of CS in combination with deep learning techniques allowing multimodal, super resolution reconstructions as well as to extend the use of CS to all SPAD configurations.

7

Extended CS D-ToF imaging using deep learning

The last chapter of this thesis presents two deep learning models to fully exploit the potential of CS by reconstructing both depth and intensity maps in all SPAD operating modes (cf. Fig. 7.1). A deep learning model based on low photon counts SPAD data has been firstly developed in order to fairly compared to SoA DL-based algorithms. Our topology introduces a compression pattern layer which operates a learned projection into a smaller latent representation space. Subsequently, the hardware implementation of the DNN's compression pattern layer for pixel-wise compressive sensing (cf., Chap. 6) is introduced, extending the use of SPAD sensors to high photon counts, as SPAD no longer suffers from hardware constraints related to the amount of data. SoA processing methods [12, 13] and SoA sensing methods [70, 71] are completely outperformed in high photon counts at equivalent signal-to-background-ratio (SBR).
On the other hand, the second model developed aims to get rid of the sparsity constraint and to provide efficient multimodal and super-resolution reconstructions, thus extending the use of the CS to all the SPAD operating modes (e.g.., synchronous) but without counters overshoot consideration. We propose a two-part deep generative model (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 the introduction of proper normalization layers with a learned pile-up effect compensation, multidimensional-multiscale filtering and the concatenation 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%.

Contents

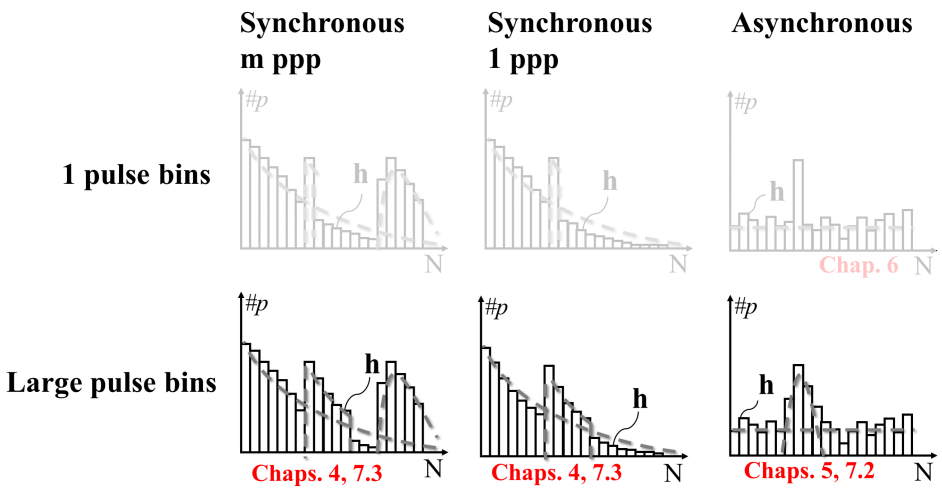


Figure 7.1: ***TCSPC distribution modeling assumptions throughout this chapter.***

7.1 Summary of existing SPAD processing methods

Direct time-of-flight (D-ToF) sensors are now key devices for a wide range of imaging applications [31, 80, 82]. However, D-ToF measurements are subject to high level background noise illumination which poses a great challenge for state-of-the-art computational reconstruction algorithms for 3D imaging. To avoid noisy image reconstructions, traditional imaging algorithmic methods based on pixel-wise maximum likelihood (ML) estimation are proposed [272]. In addition, several works have also added spatial constraints, such as introducing a total variation (TV) regularization [10] or block-matching and 3-D filtering [290]. [11] even proposed an additional pixel-wise adaptive gating ML estimation method to discard photon detection times that fall outside the gating interval. Furthermore, deep learning approaches that outperform by far prior works [12, 13] recently succeed to achieve high reconstruction fidelity under extremely low photon counts and a very low Signal-to-Background noise ratio (SBR).

On the other hand, in addition to being highly relevant in terms of depth reconstructions accuracy, deep learning data processing methods will most likely play an important role in overcoming SPAD hardware limitations (*i.e.*, data throughput, data storage, photon detection efficiency). For example, [218, 291–293] developed Deep Learning algorithms –namely image-guided depth upsampling– which consists in reconstructing a high resolution depth map from a full-scale RGB frame combined with its associated low resolution depth map, thereby indirectly relaxing hardware constraints. Besides, the Compressive Sensing (CS) strategy [294] has been investigated in the context of SPAD imaging in combination with DL data processing methods, by the use of a spatial light modulator (*e.g.*, Digital micro-mirror device, DMD) in front of the sensor [14, 295–297]. Unfortunately, optical-CS seems to be impractical for consumer electronics as long as they rely on the use of bulky optical systems which makes the system sensitive to process and temperature variations, while requiring complex calibrations.

7.2 Asynchronous CS-SPAD data as a latent space

7.2.1 Deep Neural Network topology

To improve the performance of TCSPC LiDAR systems, es- pecially under noisy measurements, this paper proposes a Deep Neural Network model that aims at jointly inferring depth, D, and intensity maps, T, from raw TCSPC SPAD data $\mathbf{h}$. This topology is divided into modules; the compression stage, the depth and intensity decoders, both providing multiple depth and intensity initial reconstructions, and the multiscale filtering (MF) that adaptively tune the weights of each reconstruction sub-channel. Last layers of this topology perform pixel- wise reconstructions selections combined with residual skip connections among the depth and intensity decoders outputs filtered by the MF module (*i.e.*, respectively lower part and upper part of Fig. 7.9) which stack several reconstructions from various filters, gathering both local, region-based information and pixel-wise temporal information. Luminance/Intensity T reconstructions additionally rely on a physical model with assumptions on frame-based normalization. Regarding the linear compression module (red block in Fig. 7.9), this layer converts high-dimensional TCSPC data to low-dimensional data in the shape of a pixel-wise vector aiming at preserving semantic information. In other words, this layer corresponds to a Conv2D1x1 in the sense of CNN, *i.e.*, a pixel- wise Fully Connected (FC) layer. Note that, this will be the layer considered as the CS layer in the next sections.

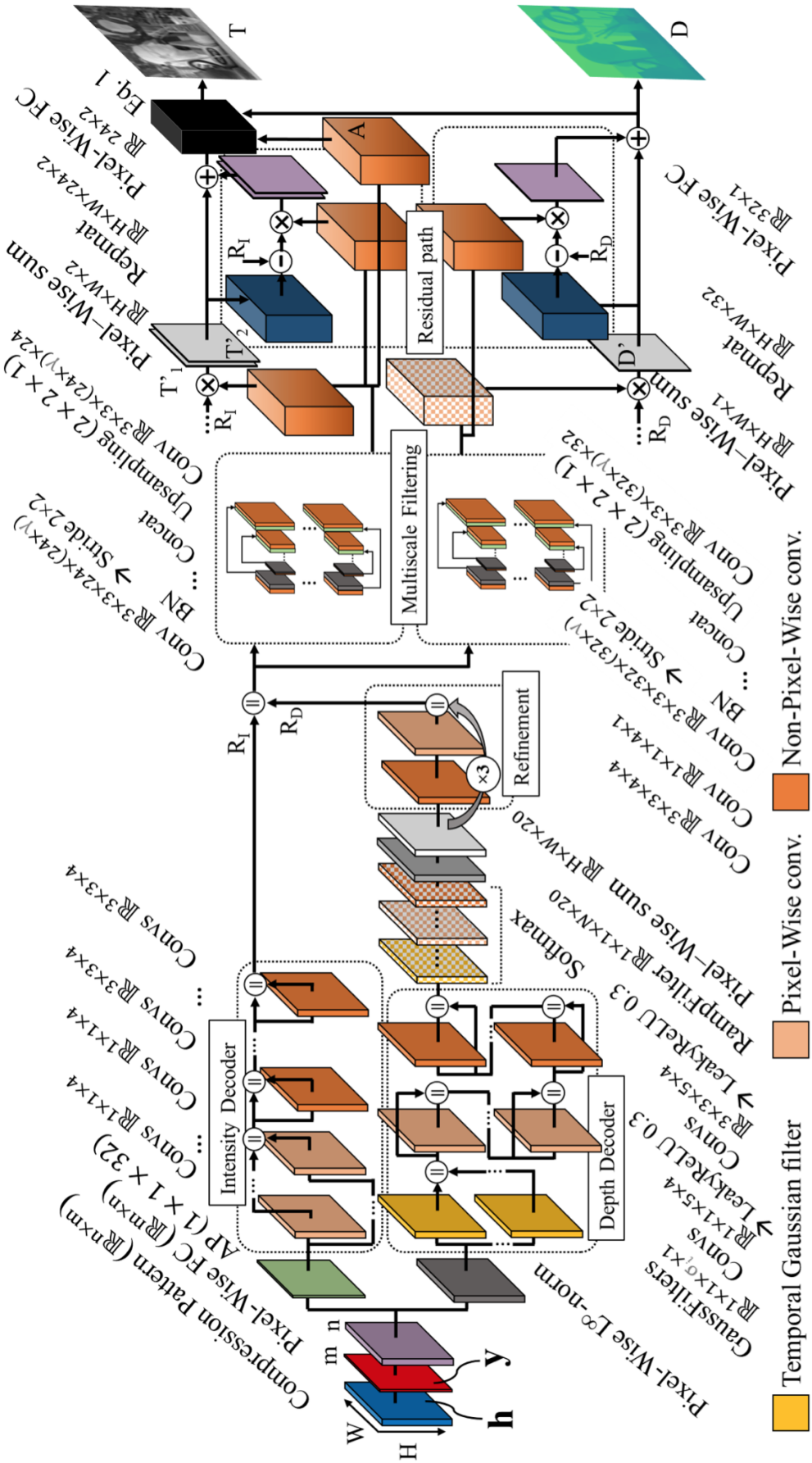


Figure 7.2: Deep Neural Network topology for depth-intensity reconstruction from compressed histograms.

On the depth decoder side, to remove histogram scaling variability from one pixel to another, this module first embeds a pixel-wise L normalization (dark gray block in Fig. 7.9). Then, "Gaussian filters" in the temporal domain with various i radius, ranging from 0 to 14 (N.B. pulse width is about 16 bins) is performed (yellow blocks in Fig. 7.9). A second filter, which is a per-pixel cascaded temporal filter, is then applied (orange blocks in Fig. 1). In order to take advantage of local temporal and spatial information simultaneously, 3D filters (Conv3D) are also cascaded afterwards (deep orange blocks in Fig. 7.9). The LeakyReLU activation is used rather than the canonical ReLU activation function in order to alleviate the "dying ReLU" problem. Every filter then provides a depth estimation output thanks to the combination of a Softmax (gridded block in Fig. 7.9) and a ramp layer (light gray block in Fig. 7.9) that act as a trainable peak detector (as a softargmax [298] would do). Finally, three layers of Conv2D 3×3 are used for an additional refinement stage.

For the intensity decoder module, since the original temporal dimension is considered irrelevant for intensity reconstructions with respect to the induced complexity on the topology, an average pooling (third-axis AP) layer (green block in Fig. 7.9) first reduces the scale of the input tensor in the temporal axis. Then in the same way as the depth decoder module, several convolutional layers are cascaded, each one providing intensity reconstructions that are concatenated in order to provide a set of intensity reconstructions. Note that for a proper behaviour of the convolutions, a custom 3D/2D padding has been implemented with a reflect padding instead of a zero padding for all modules of our model.

MF modules both take as input the intensity and depth maps provided by the decoders. It leverages local collaborations between pixels reconstructions through downscaling and then upscaling as a U-net structure [299] with an expanding coefficient value that relates to the number of channels increase at each scale. These modules for depth and intensity reconstructions selection incorporate, Conv2D 3×3 with a stride of 2×2 for downscaling steps (orange blocks in Fig. 7.9), and Conv2D 3×3 with an $2 \times 2 \times 1$ upsampling (US) (green blocks in Fig. 7.9). The MF attention module for depth selection ends with two separate Conv2D1x1 layers, one with a Softmax activation and the other with a Tanh activation respectively represented by a grid pattern and a orange-black gradient (see Fig. 7.9). The output of the softmax enables to further control the pixel-wise channel selection among the variety of depth reconstructions R_D, through a pixel-wise multiplication layer followed by a summation performed by a pixel-wise FC, providing the selected reconstruction D. Besides, the output of the tanh activation function adds the depth reconstruction residual error to the selected reconstruction D. Instead, the MF attention module for intensity estimation is terminated by three 1×1 convolutional layers with tanh activation functions, in order to estimate the number of photons per pixel from the laser source T_1 and from the background illumination source T_2. In the same fashion as for the depth module, the second tanh activation function performs the signal and noise photon number estimation residual error to be added to the previous estimation T_1 and T_2 . An additional physically-driven layer is used for the final intensity estimation (illustrated by the black block in Fig. 7.9) taking as input the depth estimation D, the photons quantity estimated from laser pulse T_1 and from background illumination T_2 and the third convolutionnal layer outputs of the MF attention module. This black block embeds the knowledge of physical laws (cf. Eq. 7.1 where J is a all-ones matrix) that govern the given imaging measurement system dataset in the learning process (*e.g.*, distance inverse-square law and frame-based intensity normalization). Finally, the last convolutional layer with tanh activation functions (denoted A) performs a pixel-wise selection of the intensity estimator from the weighted sum of the background photon counts estimator or the laser pulse photon counts estimator. This pixel-wise selection enables an accurate intensity reconstruction even when there is few photon coming from the laser pulse due to the distance measurement inverse square law and conversely when there is few photon coming from background illumination.

$$T = A \odot \frac{T'_2}{\overline{T'_2}} + (J - A) \odot \frac{T'_1}{\overline{T'_1}} \odot \frac{D^2}{\overline{D^2}}. \tag{7.1}$$

7.2.2 Experimental results using learned-compression pattern

To highlight the interest of the proposed solution, a quantitative benchmark has been conducted on "pseudo- realistic" SPAD raw data generated from the simulation model presented in [12] with SPAD control asynchronicity. This simulation ToFs model encapsulates several physical parameters such as the laser source power, distance inverse-square law, scene point brightness, SPAD control asynchronicity, dark count rate [47] and background noise measurements [300]. The simulation model generates the SPAD TCSPC measurements train dataset from the NYU V2 dataset [301], and the test dataset from the Middlebury dataset [257]. To ensure a proper matching between the train dataset and the test dataset, and to avoid unnecessary training for long range depths, all images whose dynamic range is greater than the one of the test dataset have been removed. In addtion, a data augmentation is performed, consisting in generating 12 samples of each NYU image with a dynamic range below 3m under 12 SBRs (*i.e.*, an average of 1, 2, 3, 5, 10 signal photon counts and 2, 10, 50, 100 noise photon counts). The simulation model finally provides a 3D output data volume with the 2D pixel array resolution and a third axis of 256 values (*i.e.*, a time bin resolution of 80ps and a dynamic range of 20ns (6m)).

DNN using learned-compression pattern

This subsection therefore reports our proposed DNN performances in comparison to prior works [10–13] without any further hardware considerations (*i.e.*, considering an optimized learned-compression pattern) and under low photon counts. For a fair comparison, our proposed DNN and that of [12, 13] were trained with the same training data *i.e.*, $64 \times 64 \times 256$ tensors and similarly to the original work, *i.e.*, using PyTorch, with a batch size of 4, a random initialization, an Adam optimizer [302], and a learning rate of 10^{-4} with a learning rate decay of 0.9 after each epoch. Instead, our DNN was trained using TensorFlow2, with a batch size of 8 and a number of epochs of 40 with a learning rate of 10^{-3} and a decay of 0.95 after each epoch, starting from the 10^{th} epoch. Note that for a proper convergence of our model, the training process of our neural network is performed in a 2-stage fashion so that the "Gaussian filters" and the ramps are made trainable only during the second stage (*i.e.*, 80 epochs in total). For the sake of simplicity, a Mean Squared Error (MSE) loss has been used with the Adam optimizer for all experiments reported in this paper. In order to limit the proposed DNN model size, γ is fixed to 1.5.

Even though the proposed DNN only achieves an average depth reconstructions root mean square error (RMSE) of 0.0382 m, *i.e.*, a similar accuracy to [12, 13] cf., Tab. 7.4. It provides decent depth reconstructions (cf. Fig. 7.3) in compar- ison to [10] and [11] with respectively a 80% and 46% lower average depth RMSE for m = 256 (uncompressed case). With a compression corresponding to m = 32, our method still improves the depth reconstruction performance by respectively 69% and 18%, compared to [10, 11]. Note that, our method clearly outperforms [10, 11] for any SBR, except in the 2:10 SBR case. In addition, an acceptable intensity reconstruction is obtained for $m \geq 16$, especially when compared to prior works. These results pave the way to possible data dimensionnality reduction through the use of a learned compression pattern (*i.e.*, a linear projection represented by the the red block in Fig. 7.9). Consequently, to reduce the pixel pitch in practice, while enabling the high photon flux operating mode, we propose to implement a data-agnostic linear projection in the name of a CS scheme, instead of a learned compression pattern (cf. the following section).

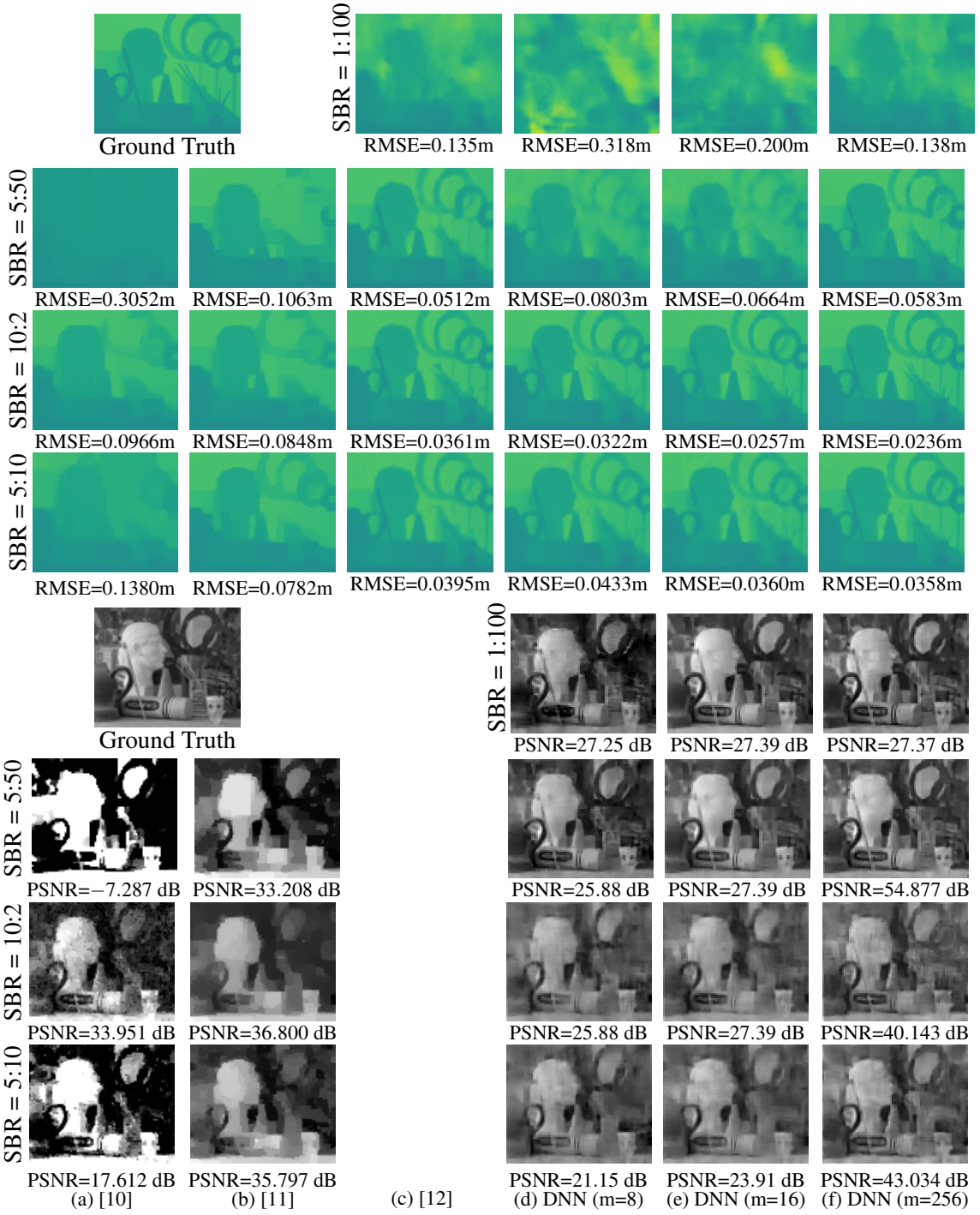


Figure 7.3: Intensity-Depth reconstructions for various SBR, reported with RMSE (in meters) and PSNR (in dB) metrics. Note that, [12] does not provide any estimation of the intensity and under the 1:100 SBR [10] and [11] completely fail for both tasks.

Modality	SBR	Shin [10]	Rapp [11]	Peng [12]	Yao [13]	ours					
						m=256	m=64	m=32	m=16	m=12	m=8
Depth	10 : 2	0.0570	0.0479	0.0198	0.0182	0.0159	0.0207	0.0204	0.0176	0.0186	0.0202
	2 : 10	0.1906	0.0612	0.0494	0.0528	0.0596	0.0836	0.1016	0.0770	0.0893	0.1381
	5 : 50	0.2502	0.0592	0.0264	0.0241	0.0311	0.0349	0.0454	0.0347	0.0371	0.0473
	2 : 100	0.3188	0.0939	0.0308	0.0293	0.0357	0.0373	0.0482	0.0386	0.0394	0.0480
	Avg	0.1849	0.0703	0.0335	0.0327	0.0382	0.0483	0.0586	0.0467	0.0509	0.0686
Intensity	10 : 2	12.68	16.91	N/A	N/A	19.96	21.54	20.52	20.28	20.38	20.70
	2 : 10	8.136	15.10	N/A	N/A	24.63	27.31	28.52	10.59	14.10	17.87
	5 : 50	8.708	15.95	N/A	N/A	25.41	28.00	27.76	13.45	18.98	21.52
	2 : 100	8.301	11.61	N/A	N/A	21.27	22.64	20.82	20.74	20.08	21.05
	Avg	9.322	14.90	N/A	N/A	22.47	24.61	24.29	16.24	18.20	20.05

*Tableau 7.1: Quantitative comparisons of several Intensity-Depth reconstruction methods under various SBR. Note that depth results are reported as an average RMSE (*m*) and intensity results are reported as an average PSNR (in dB).*

DNN combined with data agnostic compressive sensing

This last section presents the CS design performances in comparison to existing, SPAD-optimized sensing designs, *i.e.*, the SiFH [71] and FiFH [70] methods. For the sake of fair comparisons, we propose to replace the canonical argmax peak detection usually used for FiFH and SiFH by our own DNN reconstruction algorithm in order to properly characterize the CS scheme itself, decorrelated from the effect of the reconstruction strategy. Note that one CS input channel is replaced by a dummy photon counter. The proposed DNN model was therefore trained with FiFH [70], SiFH [71] as well as with CS data inputs, in the same way as described in subsec. 7.2.2 , except for the considered SBRs. The data augmentation here consists in the generation of 9 samples for each image of the NYU dataset having a dynamic range below 3m (under 9 SBRs, *i.e.*, an average of 40, 80, 120, 200, 400 signal photon counts and 2000, 4000 noise photon counts).

SBR	Fifh [70] (m=16)		Sifh [71] (m=16)		CS (m=8)		CS (m=16)	
	B=5	B=7	B=5	B=7	B=5	B=7	B=5	B=7
40 : 2000	0.40	0.04	0.33	0.03	0.08	0.09	0.05	0.05
40 : 4000	0.64	0.15	0.38	0.14	0.11	0.11	0.06	0.06
120 : 4000	0.62	0.14	0.38	0.12	0.07	0.05	0.03	0.04
Avg	0.51	0.09	0.34	0.07	0.07	0.06	0.03	0.04
Pitch (μm)	27.9	32.5	27.9	32.5	19.9	23.9	28.1	33.7

Tableau 7.2: Depth comparisons of several acquisition methods under various SBR, reported as an average RMSE in m.

SBR	Fifh [70] (m=16)		Sifh [71] (m=16)		CS (m=8)		CS (m=16)	
	B=5	B=7	B=5	B=7	B=5	B=7	B=5	B=7
40 : 2000	12.2	33.1	28.2	10.7	28.5	25.5	35.1	29.5
40 : 4000	10.7	20.6	27.5	12.6	27.1	26.5	37.4	30.6
120 : 4000	10.7	20.8	27.6	12.9	29.9	26.5	38.5	31.7
Avg	11.4	26.2	28.0	13.4	29.2	26.9	36.9	31.1
Pitch (μm)	27.9	32.5	27.9	32.5	19.9	23.9	28.1	33.7

Tableau 7.3: Intensity comparisons of several acquisition methods under various SBR, reported as an average PSNR in dB.

In case of the absence of counter overflows, Tabs. 7.2 and 7.3 reports that FiFH and SiFH [70, 71] can outperform our solution in some conditions. However, considering a practical bit depth, our proposed design shows better performances than [70, 71]. Even with same reconstruction algorithm, histogram CS allows to reach a highly accurate reconstruction, while being able to reduce both the size of measurement vectors and their bitwidths ($B = 5$), therefore reducing the total memory needs (acting on the pixel pitch). On the other hand, when decreasing the bitwidth (*i.e.*, to $B = 5$), the performances of [70, 71] are significantly downgraded cf., Fig. 7.4 and Fig. 7.5 . These Tabs. also highlight that our CS approach provides a better depth estimation RMSE compared to [70, 71] ($B = 7$), with an estimated pitch reduction of 39%. It leads to the conclusion that –in terms of depth estimation accuracy and for the considered configurations– our proposed system improves by a $11\times$ factor the depth RMSE at iso-surface (a pitch of 28μm) or a pixel surface reduction by a $2\times$ factor at iso-performance (a RMSE of 0.07m).

Similar conclusions can be drawn when considering intensity reconstruction results.

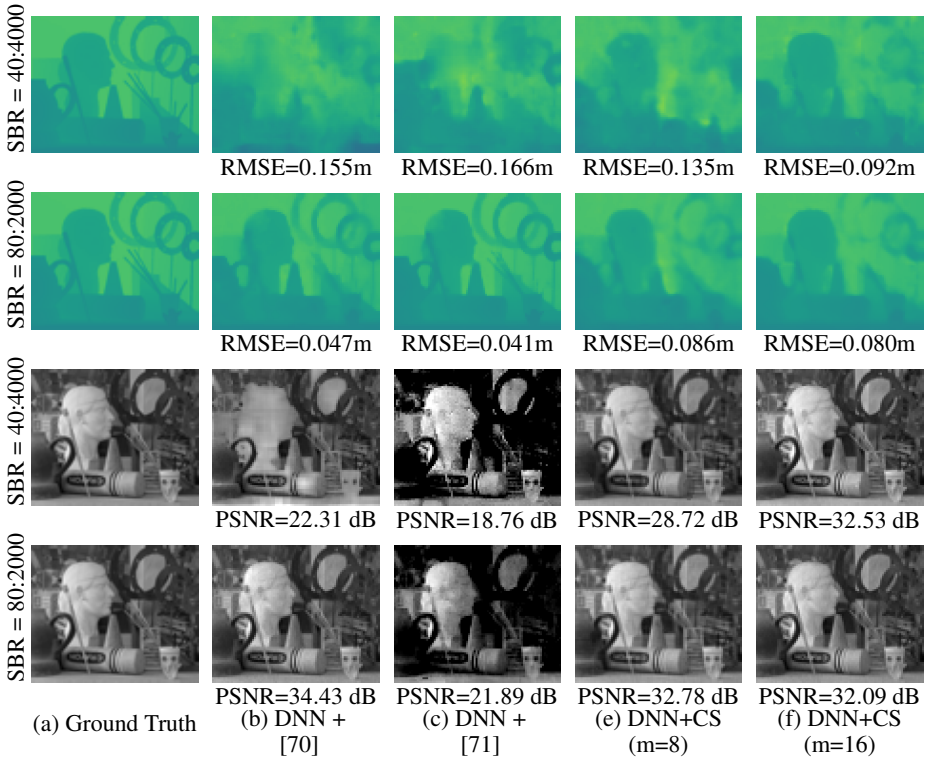


Figure 7.4: Intensity-Depth reconstructions under hardware constraints and high photon counts, with counters bitwidth of B=7.

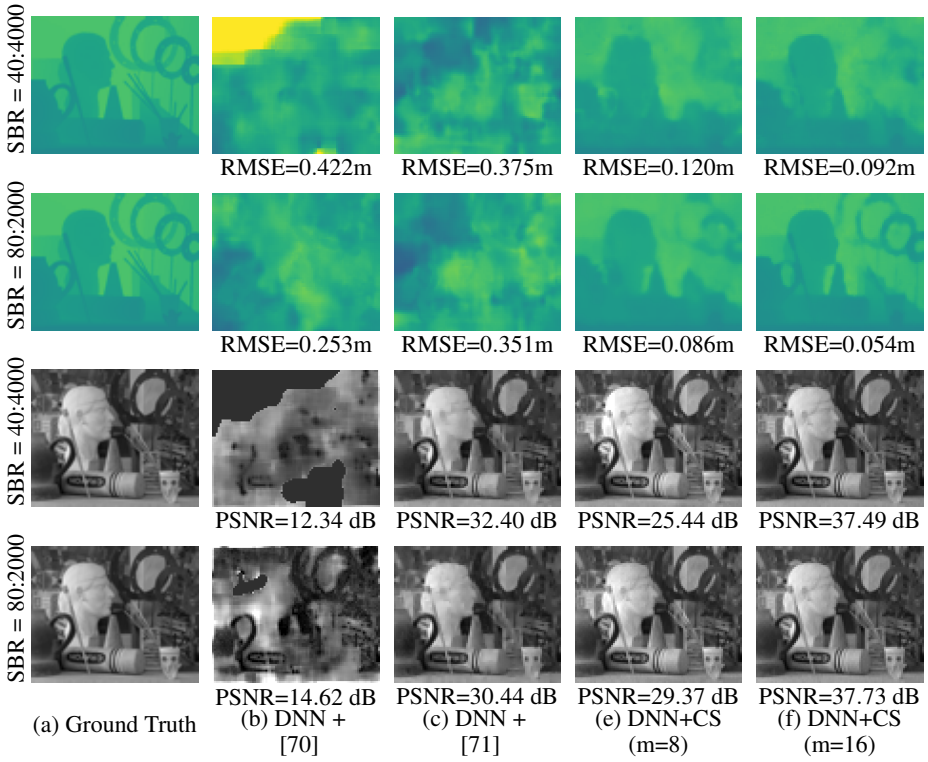


Figure 7.5: Intensity-Depth reconstructions under hardware constraints with counters bitwidth of B=5 and high photon counts.

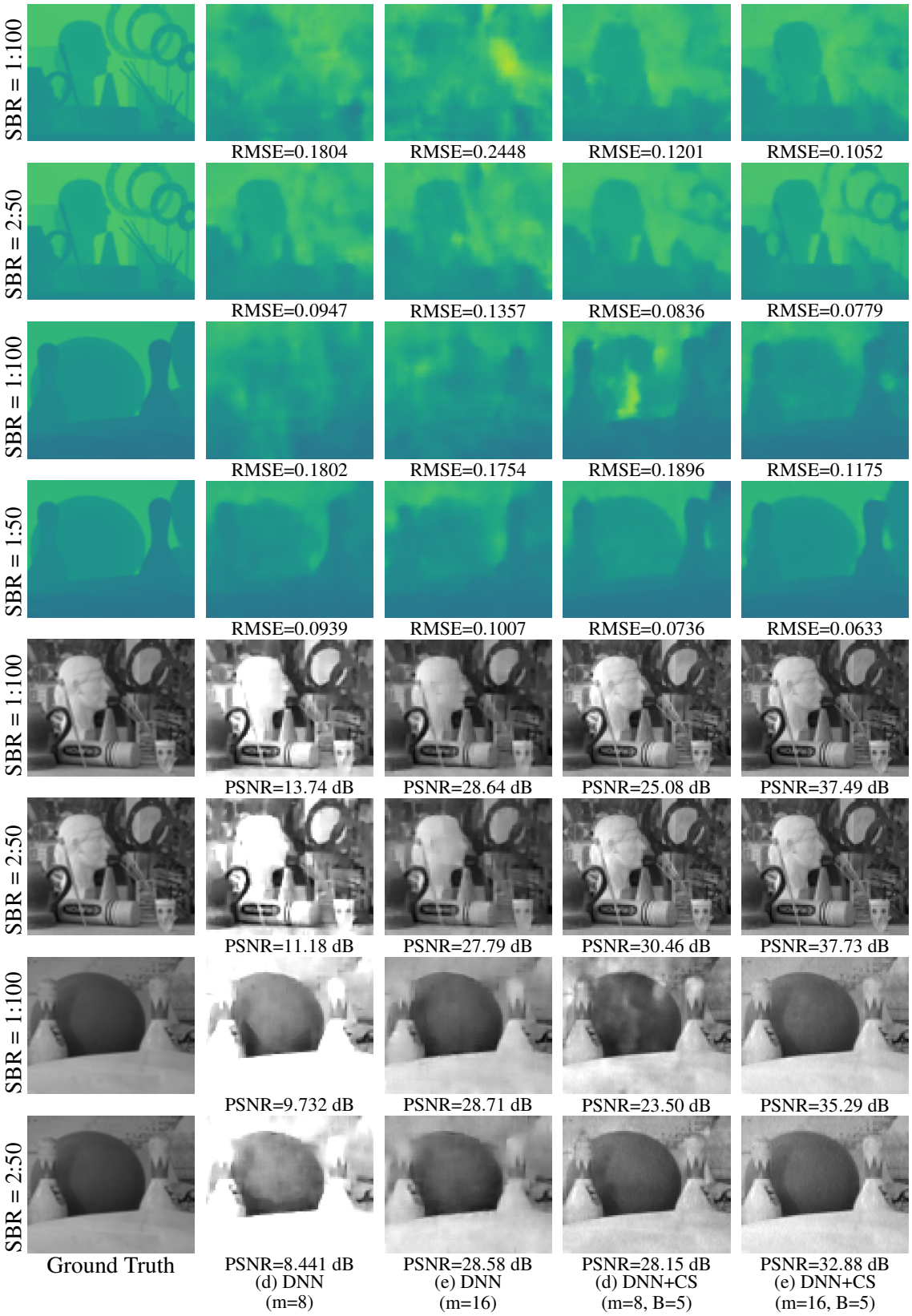


Figure 7.6: Intensity-Depth reconstructions. The 1^{st}, 2^{nd}, 3^{rd} and 4^{th} rows of depth and intensity reconstructions are respectively under 1:100, 2:50, 1:100 and 1:50 SBR while the 2^{nd} and 3^{rd} columns reconstructions are under low photon counts and the 4^{th} and 5^{th} columns reconstructions are under high photon counts.

Finally, Fig. 7.6 puts into perspectives the results with a learned-compression pattern under low photon counts with the ones obtained using the proposed CS scheme, under high photon counts. Although the SCA implementation highly constrains the measurement vector with a small bitwidth of $B = 5$, counters do not much overflow even under high photon counts. Fig. 7.6 demonstrates that the proposed DNN combined with the CS provides more accurate depth/intensity reconstructions than the DNN topology integrating learned- compression pattern (if saturation issues are not taken into consideration directly during the training stage).

7.3 Synchronous CS-SPAD data as a latent space

On the other hand, a second model aims to get rid of the sparsity constraint thus extending the use of CS to synchronous SPAD operating modes. Furthermore, this second 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 latent pixels.

7.3.1 Pixel wise compressive sensing deep generative model

Data-driven topology design

Consistently with recent works introducing deep learning approaches [12], our depth reconstruction from SPAD-CS data **y** instead of the histogram **h** is also based on a specific deep learning topology as illustrated in Fig. 7.7.

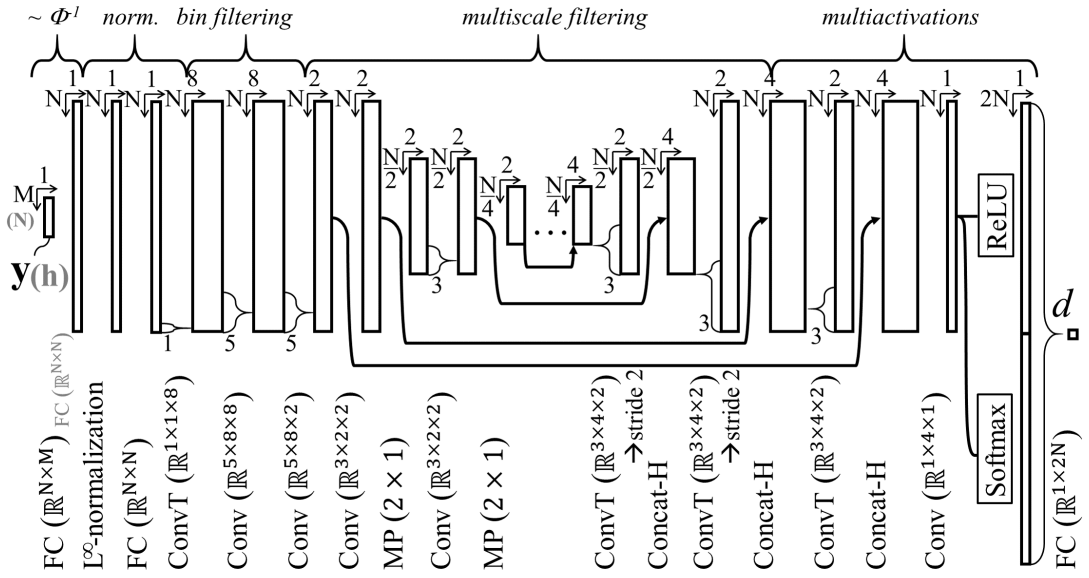
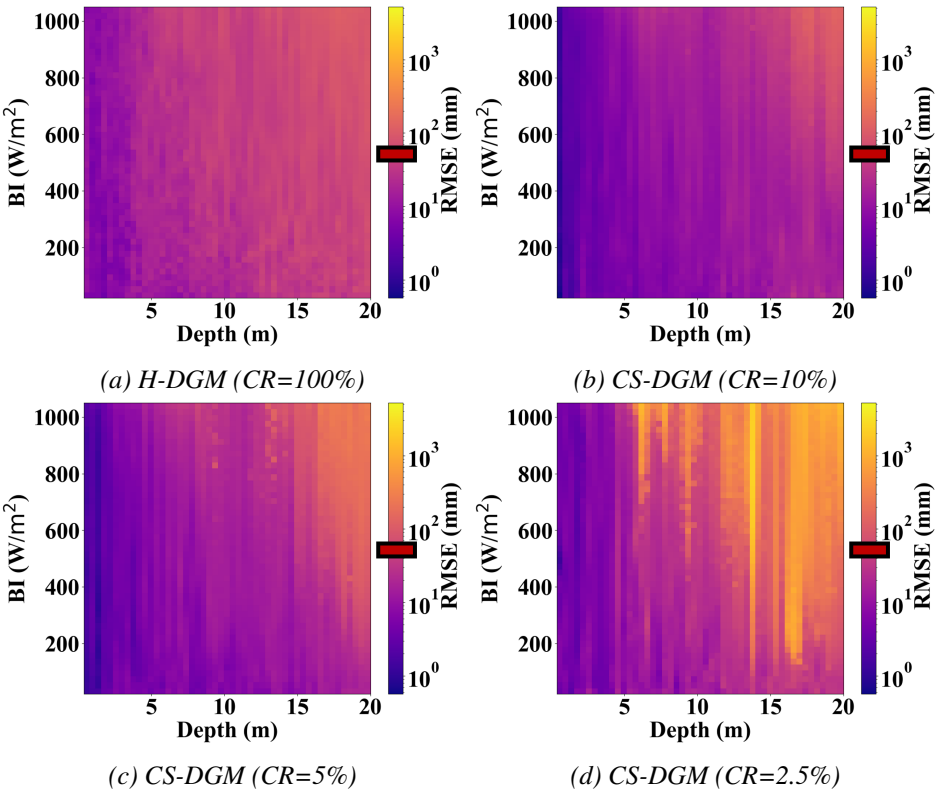


Figure 7.7: Deep Generative Model topology for pixel-wise depth reconstruction. Note that the only difference between reconstruction from histogram inputs and from CS is the first Fully Connected (FC) layer. Concat-H corresponds to horizontal concatenation and MP corresponds to Max Pooling.

This topology is designed to reconstruct a single depth measurement per pixel, corresponding to its related CS measurement vector **y**. Apart from being relevant to decompress CS data (*i.e.*,

somehow inverting Φ), this 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) first 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 the compensation of the residual pile-up effect and facilitate the detection of specific waveforms (*e.g.*, shaped laser pulses) whereas multiscale filtering allows to enforce local bin collaborations to efficiently perform denoising operations. Finally, the goal of the last layer is equivalent to a peak position detector as a Softargmax would be [298]. 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 training stage without implying unwanted overfitting. In addition, concatenating the ReLU output with Softmax further improves the reconstruction allowing the final projection to 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 function.

Pixel wise depth estimation results



(a) H-DGM (CR=100%) *(b) CS-DGM (CR=10%)*

(c) CS-DGM (CR=5%) *(d) CS-DGM (CR=2.5%)*

*Figure 7.8: 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 (78*mm*).*

Since the above reconstruction model –illustrated in Fig. 7.7– is a pixel-wise method, the use of a realistic dataset with spatial correlation between pixels to validate our approach is not necessary. 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 range. Knowing that the reconstruction 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. 2.7 and in Eq. 2.6. We thus estimate the depth root mean square error (RMSE) according to BI for various CR acquisition methods. Results reported in Fig. 7.8 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 each pixel coordinate of the 50×50 RMSE map corresponds to the ground-truth depth at 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 a low SBR (cf. Fig. 7.8, the RMSE increases with both depth and BI). Fig. 7.8 shows how the performance of compression-reconstruction methods are affected by changing the depth and the BI compared to the uncompressed counterpart. Furthermore, since the RMSE for some BI-Distance configurations is far below the simulated TDC resolution (78mm), it means that the proposed CS acquisition system would allow an even greater temporal resolution, without the expense of additional memory. Indeed, the CS measurement vector $\mathbf{y}$ size is advantageously not correlated to the temporal resolution. In other words, our CS acquisition system could benefit from higher TDC clock frequency without any hardware extra constraints.

SBR	High photon flux		Low photon flux		
	CS-DGM (CR=10%)	H-DGM	Rapp [11]	Lindell [218]	Peng [12]
0.2	**0.0035**	**0.0091**	0.0267	0.0295	0.0153
0.1	**0.0056**	**0.0153**	0.0359	0.0380	0.0185
0.04	**0.0137**	0.0377	0.0890	0.0748	0.0266
0.02	0.0402	0.0667	0.1163	0.2435	0.0328

Tableau 7.4: Quantitative comparisons of several depth reconstruction methods under various SBR. Note all results are reported as an average RMSE (in meters). [218] uses RGB sensor fusion, [11] refers to an unmixing algorithm from signal and noise sources while [12] uses a non-pixel-wise neural network for depth reconstruction.

Despite the lack of state-of-the-art works using a similar SPAD sensor operating mode due to its complex data post processing, we can still coarsely evaluate our proposed method. Indeed, by combining works detailed in [72] (*i.e.*, that mitigates pile-up effect distortions by temporally misaligning SPAD measurement windows) with [9] (*i.e.*, that takes advantage of photon arrivals coincidence), we can derive an extension of our considered operating mode to be equivalent to the one considered in [10]. Please note that these low photon flux imaging working modes still involve a non-negligible footprint in the hardware implementation of the pixel design in comparison to our CS acquisition method. Considering the restrictive case of a pixel-wise approach, Table 7.4 reports the reconstruction RMSEs with respect to the pixel-related SBRs.

7.3.2 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 insensitive as possible to BI. Our multimodal reconstruction algorithm directly takes advantage of the deep learning framework presented in the previous section. Our SPAD data model mainly takes into account the most relevant physical characteristics denoted d and τ_r. The DGM topology presented in Fig. 7.9 therefore aims at jointly inferring d

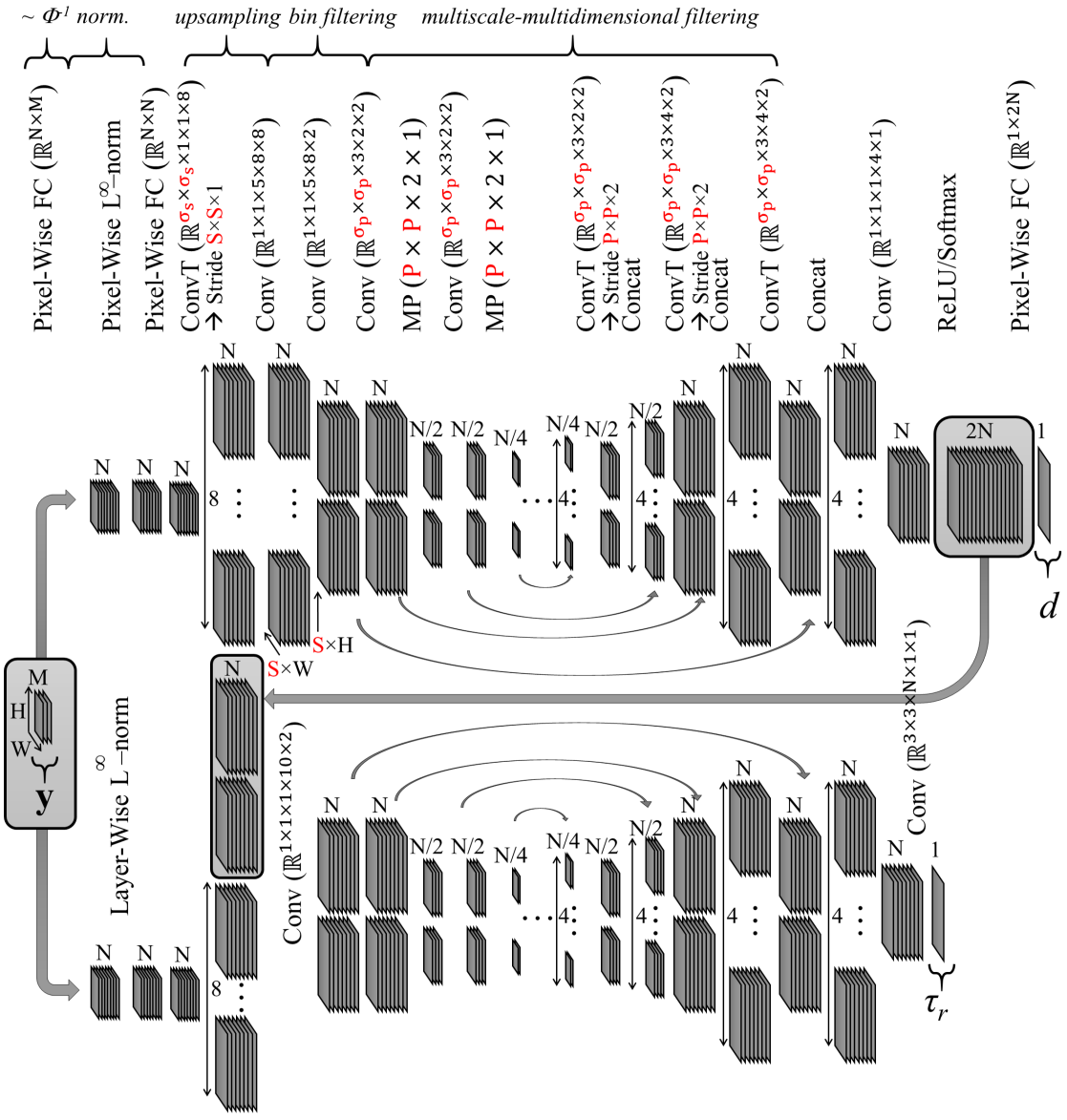


Figure 7.9: Deep Generative Model topology for depth-luminance reconstruction from compressed histograms.

and τ_r from $\mathbf{y}$, for all pixels simultaneously. The Depth part of this topology (*i.e.*, upper parts of Fig. 7.9) slightly differs from the one reported in Fig. 7.7 by including a local spatial collaboration during the multiscale filtering stage while using $\sigma_P = 3$ and $P = 2$ ($\sigma_S = 1$, $S = 1$) removing the strict pixel-wise characteristic of the system. For the luminance part (*i.e.*, lower parts of Fig. 7.9), 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 only be performed using $\mathbf{y}$ as input (which is highly efficient in case of a high BI).

7.3.3 Extension to Super-Resolution reconstructions

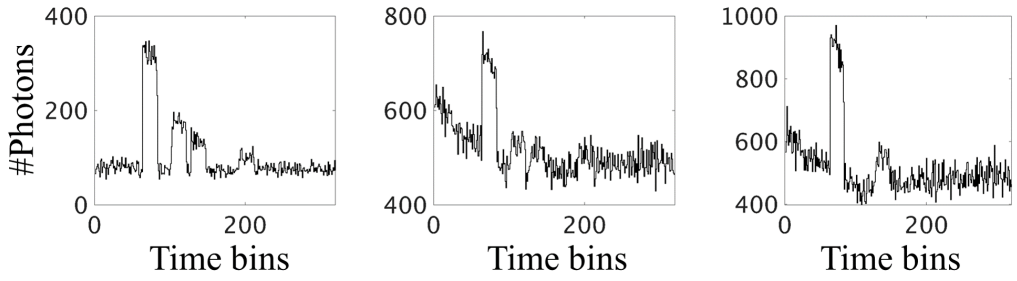


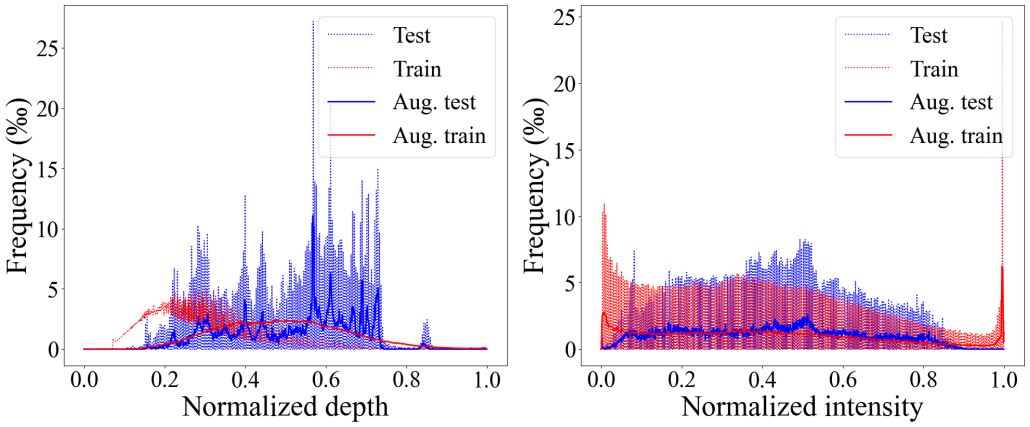
Figure 7.10: Examples of ToF histogram in super resolution. The first and second histograms refer to the same object measurements (i.e., 5, 8, 10, 15 m*) but with more BI for the second image. The third plot is the ToF histogram of 5, 5, 10, 10*m *depth object measurements.*

Several works have been conducted on the use of SR approaches in the SPAD context. For example, [303] combines data acquisitions from the SPAD sensor with data from a standard CMOS camera to improve the spatial resolution of the SPAD sensor. As we proposed for the multimodal reconstruction from SPAD data (*i.e.*, without any additional sensor), it is relevant to take advantage of the intrinsic nature of the data to also perform SR, namely because SPAD imaging systems suffer from a limited spatial resolution. Therefore, we decided to emulate each pixel as a group of latent pixels, thus leading to the acquisition of shared histograms. It would theoretically enable it to perform a finer mutual information retrieval compared to what would happen in case of a simple averaging, this with more realistic synthetic data. However, it also implies histogram peaks masking (*i.e.*, short range pixels are "shadowing" long range ones at the same albedo) which makes the reconstruction more complex in practice. Fig. 7.10 illustrates simulated histograms where 2×2 neighbor pixels do not share the 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. 7.9. Note that even if this SR reconstruction is not the main contribution of this work, it still demonstrates that our framework is compatible with any deep learning SR techniques, with possible further extensions to image-guided upsampling or pixel-shuffle [304].

7.3.4 Experimental setup

Multimodal and SR reconstructions with the NYU V2 dataset [301] for training, and Middlebury [257] for testing are presented in Fig. 7.12, Fig.7.13 and Fig. 7.14. Since Deep Learning performances rely on the training dataset, we applied data augmentation on both datasets, namely

to limit overfitting issues. It aids in training a model that better generalizes to the test dataset. Indeed, a relevant data augmentation has been considered in order to provide a proper evaluation of the 12 test scenes of Middlebury reconstructions under 7 different BIs. To that end, 3 samples of each NYU image are generated (*i.e.*, 1449 $\times$ 3 RGB-D images) with various BIs, randomly drawn from 50 to 1050 $\mathrm{W/m^2}$.



(a) Depth distributions *(b) Intensity distributions*

Figure 7.11: Depth-Intensity distribution of values from images of train and test datasets (NYU V2 and Middlebury), before and after data augmentation (i.e., depth holes filling, scale normalization, noise adding and re-quantization).

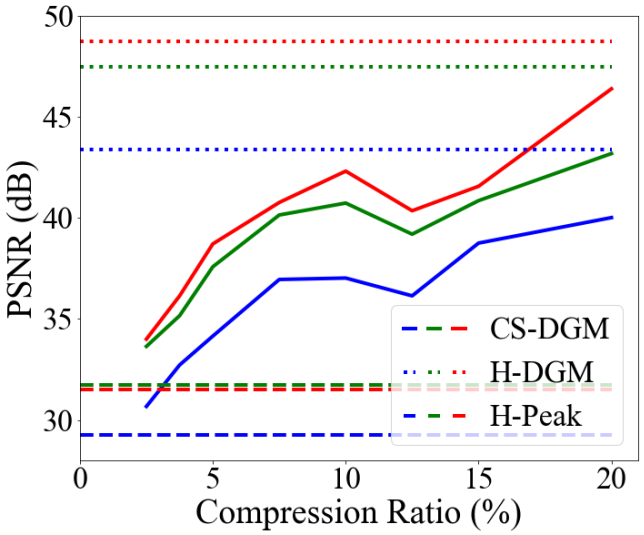
As depicted in Fig. 7.11, data preprocessing is performed on the original datasets to adapt the depth range distribution and to properly handle the data quantization effects. To fix the depth range distribution misalignment, we apply a data augmentation which consists in applying depth values transformation to improve the generalization between training and test datasets over the entire dynamic range. Since the NYU V2 depth information is quantized in a log-based fashion with a limited number of bits, a log-scaled uniform quantization noise has been added to avoid any unrealistic physical behavior that could limit the DGM performances by introducing reconstruction artifacts. The same correction has been applied to the Middlebury dataset but with uniform noise ranging below the initial dataset quantization step. Note that Middlebury images also contain depth holes that have been filled using a median-diffusion algorithm [305] [306]. Since all datasets must have the same scene scale and in order to be consistent with the simulated TDC resolution, the training and test datasets are similarly normalized (cf., Table 7.7) leading to an equivalent depth quantization step of 78mm.

Finally, all images are subsampled (with border removal) to 48x64, or to 96x128 for low resolution (LR) and SR reconstruction test benches. Neural network models reported in this section are all trained using TensorFlow2, with a batch size of 5 frames (3 in subsection 7.3.3) 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 the Adam optimizer [302] using an output frame-based Mean Squared Error (MSE) loss function.

7.3.5 Experimental results

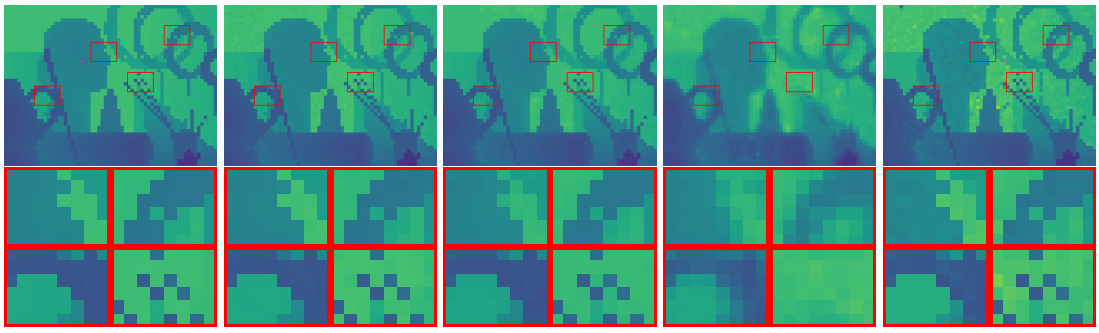
The operating mode extension from [9, 72] previously performed in section 7.3.1 for pixel-wise depth reconstruction benchmark (cf. Tab. 7.4) does not yet allow for a fair and quantitative comparison in the context of non pixel-wise reconstruction with previously mentioned works in Tab. 7.4 (*i.e.*, [11, 12, 218]). Indeed, these works do not consider the effects of the inverse-square law (*i.e.*, photons quantity inversely proportional to the square of the distance). Consequently, the

SBR variation between pixels depending on depth measurements is not taken into account. For instance, under bright outdoor lighting conditions ($P_a = 1050\text{W}/\text{m}^2$), the average SBR level of all the Middlebury images is of 0.1, but the per-pixel SBR can vary from 0.0005 to 0.2. To roughly put results in perspective, we must then consider that despite a modest average RMSE of 0.0380m under 0.1 SBR, [218] starts from an average RMSE of 0.0295m under 0.2 SBR with a dramatic decrease to an average RMSE of 1.3650m under 0.001 SBR. Therefore, a fair comparison with our proposed multimodal reconstruction method (CS-DGM with CR=20%) under a 0.1 average SBR that achieves an average RMSE of 0.24m, would be to consider a weighted ratio of average RMSE results of [218] related to the pixel quantity ratio under a SBR level between 0.0005 and 0.2.



	100	350	1050
CR=2.50%	34.00	33.64	30.67
CR=3.75%	36.14	35.16	32.72
CR=5.00%	38.71	37.58	34.15
CR=7.50%	40.76	40.14	36.95
CR=10.0%	42.31	40.73	37.02
CR=12.5%	40.35	39.19	36.14
CR=15.0%	41.56	40.85	38.75
CR=20.0%	46.38	43.18	40.01
H-DGM	48.72	47.47	43.40
H-Peak	31.51	31.75	29.28

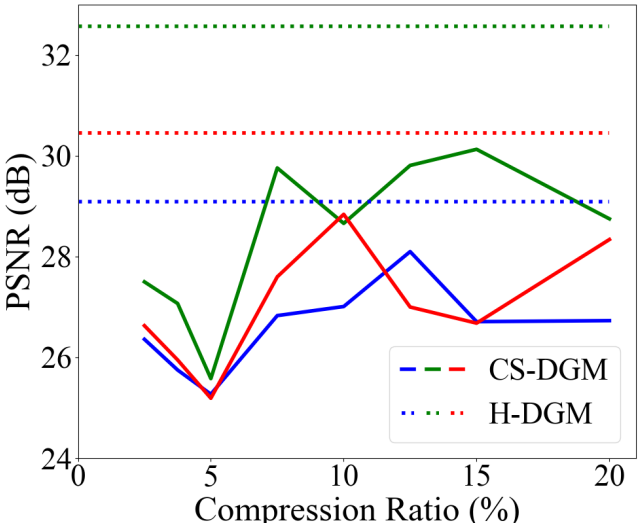
Tableau 7.5: Average depth reconstructions PSNR (over the 12 test images of Middlebury), for three P_a) (i.e., $100\text{W}/\text{m}^2$ in red, $350\text{W}/\text{m}^2$ in green and $1050\text{W}/\text{m}^2$ in blue). Horizontal dotted lines are for Histogram based reconstruction (H-DGM). The horizontal dashed line (H-Peak) is for a learned histogram filter combined with a peak detector after bin-wise re-weighting (≈ adapted log-matched filter).



(a) GT 48×64 *(b) H-DGM (CR=100%)* *(c) CS-DGM (CR=20%)* *(d) CS-DGM (CR=2.5%)* *(e) H-Peak (CR=100%)*

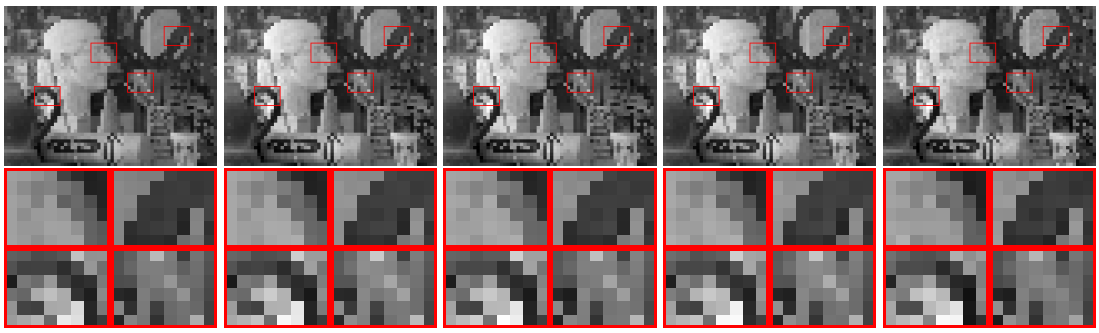
Figure 7.12: Depth reconstructions under a BI of P_a=$1050\text{W}/\text{m}^2$. H-DGM refers to our histogram based reconstruction and H-Peak is based on a learned histogram filter combined with a peak detector after bin-wise re-weighting (≈ adapted log-matched filter).

Therefore, we quantitatively evaluate the histogram CS with uncompressed reconstruction



	100	350	1050
CR=2.50%	26.63	27.50	26.36
CR=3.75%	25.95	27.07	25.75
CR=5.00%	25.19	25.58	25.27
CR=7.50%	27.60	29.76	26.83
CR=10.0%	28.84	28.66	27.01
CR=12.5%	27.00	29.81	28.10
CR=15.0%	26.68	30.13	26.71
CR=20.0%	28.34	28.75	26.73
H-DGM	30.46	32.57	29.09

Tableau 7.6: Average intensity reconstructions PSNR (over the 12 test images of Middlebury), for three P_a) (i.e., $100\mathrm{W/m^2}$ in red, $350\mathrm{W/m^2}$ in green and $1050\mathrm{W/m^2}$ in blue). Horizontal dotted lines are for Histogram based reconstruction (H-DGM).



(a) GT 48×64 *(b) H-DGM (CR=100%)* *(c) CS-DGM (CR=20%)* *(d) CS-DGM (CR=10%)* *(e) CS-DGM (CR=2.5%)*

Figure 7.13: Intensity reconstructions under a BI of $P_a = 1050\mathrm{W/m^2}$. H-DGM refers to our histogram based reconstruction.

baselines (see Tab. 7.5 and 7.6), which are firstly a simple peak detector performed after a learned filter (*i.e.*, H-Peak) performed on our "pseudo-realistic" simulated raw histograms and secondly our proposed DGM reconstruction performed on the same raw data, before performing CS (H-DGM). The average reconstruction PSNR curves (cf. Tab. 7.5 and 7.6) are obtained using an additional MSE loss function applied on the luminance part (5x more weighted). Compared to the results of a pixel-wise depth-only reconstruction (cf. Fig. 7.8), this multimodal collaborative reconstruction enables to further improve BI robustness, increasing the accuracy for depth estimation (cf. Tab. 7.5), 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 lights on, because both light sources (laser and BI) then participate in estimating τ_r.

In spite of a small degradation in comparison to the uncompressed counterpart (*i.e.*, H-DGM), the proposed collaborative reconstruction substantially outperforms H-Peak which is equivalent to an adapted log-matched filter. Indeed, Fig. 7.12 shows noisy depth reconstructions in case of a low reflection τ_r and long distance measurements d when using H-Peak whereas CS-DGM provides robust reconstruction of low intensity signals and with a low CR of 2.5%. Nonetheless, note that

as reported in Tab. 7.6 and 7.5, the intensity-depth map reconstructions are significantly degraded for a CR below 7.5%. Fig. 7.12 and Fig. 7.13 typically illustrates this, where most of the fine details are lost in case of a CR of 2.5%.

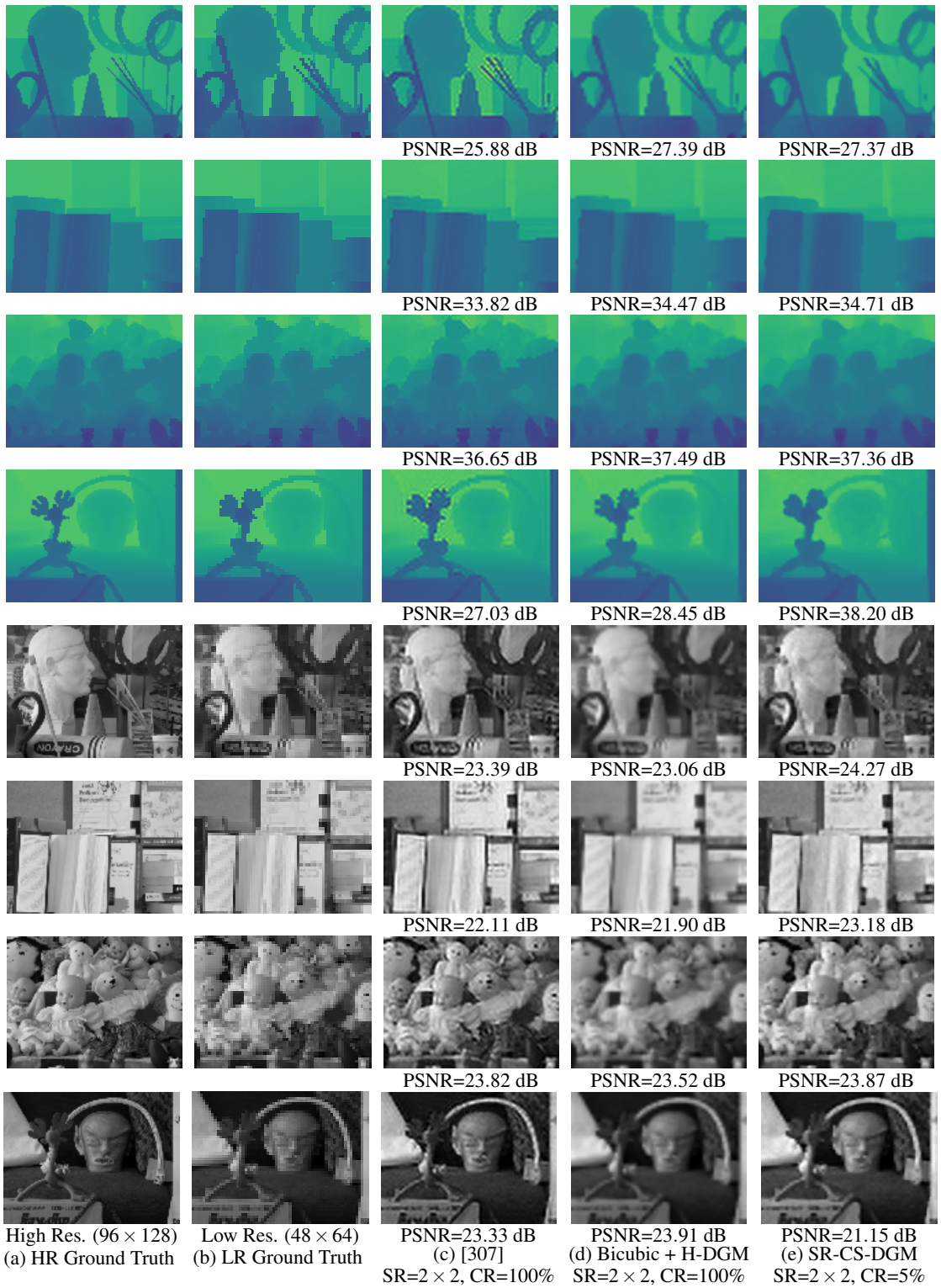


Figure 7.14: Intensity-Depth SR reconstructions under P_a=350W/m^2.

Fig. 7.14 highlights the performance of the upsampling stage extension described in section 7.3.3, comparing our SR-CS-DGM reconstructions from 48×64 CS ToF histograms (*i.e.*, 5% ($M = 16$)) with SR reconstructions using a residual dense network [307] and with SR reconstruc-

tions using a bicubic interpolation (both performed on 48×64 intensity-depth map reconstructions using H-DGM described in section 7.3.2). Reconstruction artifacts from both compressed and uncompressed histograms (*i.e.*, SR-CS-DGM, H-DGM with [307] and H-DGM with bicubic) are yet visible (see Fig. 7.14, holes in circles row 1 and ears details row 4). These are likely due to the combination of compression-related loss and background noise effects while long distance pixel-level details at low-resolution scales are masked by foreground objects (*i.e.*, with a higher contribution to low-resolved pixel-shared histograms). We can also observe ringing artifacts on reconstructions using [307] (see Fig. 7.14 col. 3) so as to blur artifacts when using a bicubic interpolation (see Fig. 7.14 col. 4), while both artifacts are mitigated using our SR extension achieving a better accuracy in terms of PSNR even at an extreme CR of 5%.

7.4 Conclusion

The first part of this chapter introduces a Deep Neural Network (DNN) integrating a learned-compression pattern layer in order to demonstrate a possible input data reduction without too much depth/intensity information loss for future data agnostic compressive sensing replacing hardware TCSPC sensing design. In the typical state-of-the-art (SoA) low photon counts SPAD operating mode, this study provides promising results with an average RMSE depth reconstruction loss of only 0.016 m and an average PSNR intensity reconstruction increase of 10 dB in comparison to the best performing SoA.

Subsequently, a Compressive Sensing (CS) hardware implementation scheme replacing the DNN learned-compression pattern layer is proposed which allows to relax hardware constraints on the SPAD sensor. The CS design consequently reduces the number of memory words and the number of bits per words, thereby reducing the pixel pitch required for an in-pixel implementation with the use of a pixel-wise shuffled Cellular Automaton (CA). Based on the same reconstruction algorithm, the proposed acquisition method thus allows higher reconstruction performance to existing sensing methods [70, 71] with an estimated pitch reduction of approximately 40%. Finally, the CS scheme avoids counters overshoot even at high photon counts, which, in combination with the proposed DNN, provides a higher reconstruction accuracy than the best-in-class remote processing work that limits its mode of operation to low photon counts.

The second neural network topology aims to get rid of the sparsity constraint thus extending the use of CS to all SPAD operating modes. From the topology view point, a pixel-wise/layer-wise L^{∞} normalization layer is applied in order to make the reconstruction as robust as possible against background illumination variations. Implementing depth-luminance DGM interconnections with two types of 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 only 32 CS measurements, the proposed topology enables an average PSNR depth reconstruction loss of only 7dB, while its luminance counterpart is less than 3dB for background illumination levels up to 1050 $\mathrm{W/m^2}$.

In addition, the second part of this chapter demonstrates that this in-pixel histogram CS can be combined with a super-resolved multimodal reconstruction thanks to a dedicated DGM variant, even under the most challenging signal processing conditions, such as in the case of high photon flux that leads to distortion of the incident temporal waveform due to non-linearities in the image formation model. Even though these results are preliminary, disambiguating local ToF informa-

tion collected by SPAD pixels paves the way for even higher resolved images from D-ToF type of sensors. Future work to improve SR reconstructions could include pixel-shuffle [304] with deactivatable skip connections [308].

Finally, since this chapter is centered on specific deep generative models, the use case treated is limited to an intensity/depth map reconstruction. However the proposed histogram CS scheme is based on a universal sensing matrix, so it should in principle be compatible with a wide range of applications (*e.g.*, gesture recognition, object detection, NLOS, FLIM). Note that reconstruction models are assumed to be performed remotely and can therefore be replaced by other specific reconstruction algorithms. Yet, it can advantageously be implemented in-sensor depending on the application use case and available computing resources. However, it appears that deep learning is efficient and generic enough for latent parameter estimation based on compressed data. As a result, assuming that SPAD histogram data is structurally shaped and admits a small set of latent variables, our proposed CS scheme –in combination with a properly designed deep learning model– should be taken into consideration to relax sensor hardware constraints.

Conclusion

This thesis explored new acquisition and processing methods based on Expectation-Maximization (EM) algorithm, Compressive Sensing (CS) and Deep Learning (DL) algorithm for SPAD-based TCSPC imaging systems. All these algorithmic co-design solutions address two of the main challenges in the development of SPAD D-ToF sensors: spatial resolution limited by the amount of data to store and transfer, robustness to noise related to the new application which is 3D imaging in outdoor conditions. As written in Tab. 7.7, several modes of SPAD operation, resulting in various histograms distribution model, are discussed in this manuscript. due to the wide variability of the SPAD sensor application framework, *e.g.*. biological, semiconductor, spatial, automotive etc.

	Chap. 4	Chap. 5	Chap. 6	Chap. 7.2	Chap. 7.3
Related TCSPC distribution modeling	Mixture of Truncated - Shifted Erlang distributions	Mixture of Gaussian - Uniform distributions	Sparse	Sparse/ Mixture of Gaussian - Uniform distributions	Mixture of Truncated - Shifted Erlang distributions
Photon counts	High	High	High	Low and High	High
Compression ratio	100%	0.4%	1% - 10%	2.5% - 20%	6% - 20%
Pixel pitch (µm)	∅ ($\sim$ 355)	44.7	39 - 53.9	20 - 28	∅ ($\sim$ 28)
Overshoot consideration	No	Yes	Yes	Yes	No
Multimodal reconstruction	Yes	No	No	Yes	Yes
Super-resolution reconstruction	No	No	No	No	Yes
Depth accuracy	0.03 m (RMSE)	0.41 - 3 m (RMSE)	0.5 - 0.9 % (Success rate)	0.038 - 0.068 m (RMSE)	30 - 40 dB (PSNR)
Albedo accuracy	>15 dB (PSNR)	∅	∅	24.61 - 16.24 dB (PSNR)	>25dB (PSNR)

Tableau 7.7: This table reports all the reconstruction performances and contributions of this manuscript.

- **Time distribution modeling for piecewise-constant Poisson process rate (Chapter 4):** This chapter focused on the reliability of the 3D reconstruction taking into account the drawbacks of SPAD sensors (in worst case) such as the dead time involving the pile-up effect in a

synchronous working mode and considering the high photon flux mode of operation. These considerations are generally neglected in the literature due to the complexity of the statistical analysis of the time distribution of the outputs of such a queueing system with PCPP inputs. Although estimating parameters of such a model involves high computational complexity of the algorithms. The benefit of the high parametrization of $m\mathscr{E}$ is that it provides a general model of Phase-type distribution (PH) for numerous and various real use cases. A custom genetic based Expectation-Maximization (c-GEM) algorithm has been proposed to accurately estimate the number of phases in the mixture distribution model as well as the 4 latent parameters of each component. Even though, this problem statement has been shown to be relevant in the context of a specific TCSPC imaging system, the proposed algorithm has the advantage of being compatible with a broad class of complex model, for instance with other SPAD configurations, involving the detection of multiple object reflections or enabling multi-illumination sources rejections. Although the proposed approach takes into account a wide range of application parameters and a robust noise filtering method, all these considerations do not allow a pixel-level hardware implementation of the proposed solution.

- **Expectation-Maximization Algorithm and Hardware Co-design (Chapter 5):** Instead, this chapter addressed the major hardware limitations caused by a "data deluge" and the inherent data throughput bottleneck. This is done in focusing on the high photon counts asynchronous SPAD operating mode, resulting in a simpler regression model, in the form of a Gaussian-Uniform statistical SPAD data model, than that considered in Chap. 4. Then, we proposed an online parameter estimation instead of building the whole ToF pixel-wise histograms. To highlight the proposed solution, quantitative results demonstrate a superior depth reconstruction accuracy compared to previous works while enabling a reduction of the pixel pitch. In addition, compared to existing sensing approaches [70] and [71], CO-EM outputs parameters allow for a direct estimation of the albedo and the SBR, not only depth. However, when the algorithm may encounter convergence problems, in such case the SPAD raw data is definitely lost since parameters do not fit to the latent SPAD measurement histograms. For a concrete deployment of the algorithm, remote filtering could be achieved, since the CO-EM is pixelwise, taking advantage of a spatial collaboration, such as using any spatial regularization technique as Total Variation.

- **Histogram Compressive Sensing using Shuffled Cellular Automata (Chapter 6):** Therefore, this chapter also proposes an alternative to the conventional TCSPC data but based on compressive sensing theory, limiting information loss compared to solution presented in Chap. 5 in case of convergence problem. The proposed solution CS TCSPC reduces the number of stored values and the number of bits per values, thereby reducing the pixel pitch required for an in-pixel implementation by using an in-pixel Cellular Automaton. We noticed that similar reconstruction performance to existing methods with an estimated pitch reduction of 26%, compared to alternative most advanced SoA for TCSPC compressing, are reached with this acquisition method. This approach is also able to take advantage of signals' structure assumptions with more complex signals through the use of more advanced reconstruction strategies.

- **Extended CS D-ToF imaging using deep learning for super resolution and multimodal reconstructions (Chapter 7):** This last chapter aims at extending the use of the compressive sensing (CS) approach to various SPAD operating modes. Because, low photon regimes have been considered as a general assumption in the previous literature, suitable for long range imaging systems (over a range of several kilometers), thus neglecting noise photon

counts pileup effect and the hardware limitations of the SPAD sensor imaging system. We thus consider this low photon regimes in order to be able to address both long and short range measurement in combination with our proposed CS scheme. To do so, a Deep Neural Network (DNN) integrating a learned-compression pattern layer is introduced in the first part of this chapter. In this typical state-of-the-art (SoA) low photon counts SPAD operating mode, the proposed DNN provides promising results with an average RMSE depth reconstruction loss of only 0.016 m and an average PSNR intensity reconstruction increase of 10 dB in comparison to the best performing SoA DL-based works. Because, the low photon operating mode is chosen by default in the literature due to the hardware constraints imposed by the SPAD sensor. High photon regimes are rather considered when the CS scheme based on the pixel-wise shuffled cellular automaton replaces the learning embedding. The evaluation of the proposed approach was done in two parts. A first evaluation validated the CS acquisition scheme in comparison to existing sensing methods [70, 71], resulting in an estimated pitch reduction of 39%. The second evaluation states that the CS scheme avoids counters overshoot even at high photon counts. Which what in combination with the proposed DNN, provides a higher reconstruction accuracy than the SoA works that currently limit its mode of operation to low photon counts.

The second constraint on the SPAD sensor operating mode is its asynchronization for sparse regression models of the data histograms. This may imply a high power consumption, a complex electronic circuitry and one photon-per-pulse detection functioning limitation. Therefore, the proposed second neural network topology aims to get rid of the sparsity constraint thus extending the use of CS to all SPAD operating modes. This second topology design provides promising numerical results for multimodal sensing which show that even for a compression using only 32 CS measurements, the proposed topology enables an average PSNR depth reconstruction loss of only 7dB, while its luminance counterpart is less than 3dB for background illumination levels up to 1050 $\mathrm{W/m^2}$. Moreover, it is compatible with super-resolved multimodal reconstruction thanks to a dedicated DGM variant, even under the most challenging signal processing conditions.

Despite good progress for the hardware limitations of SPAD sensors with the proposed variational Bayesian algorithm, compressive sensing, and deep learning algorithm, there is still much to explore to improve these solutions. An immediate continuity of the study of these algorithm-hardware enablers would be to jointly take advantage of the benefits of each of the CO-EM and CS solutions. This can be achieved by arranging the smart pixels based on CO-EM and CS processing methods alternately on the rows and columns of the sensor array. The used remote processing algorithm therefore benefits locally from two different forms of data acquisition, enhancing sensor performance in any operating mode. Due to the fact that the CO-EM method is more efficient when the laser pulse is spread over more than one bins, and conversely for the CS method.

Finally, to overcome one of the major limitations of high-quality RGB-D imaging systems, *i.e.*, online calibration and alignment, the canonical Bayer BGGR color filter pattern could be replaced in the future by an RGB-IR pattern with a CO-EM or CS-based smart pixel behind the IR filter.

Publications & Patents

Journal

- V. Poisson, V. T. Nguyen, W. Guicquero and G. Sicard, "Luminance-Depth Reconstruction From Compressed Time-of-Flight Histograms," in *IEEE Transactions on Computational Imaging*, vol. 8, pp. 148-161, 2022, doi: 10.1109/TCI.2022.3149088.

- V. Poisson, W. Guicquero, D. Coriat and G. Sicard, "A 2-Stage EM Algorithm for Online Peak Detection, an Application to TCSPC Data," in *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 69, no. 9, pp. 3625-3629, Sept. 2022, doi: 10.1109/TCSII.2022.3181687.

- V. Poisson, W. Guicquero, and G. Sicard, "TCSPC histogram data modeling: a custom EM algorithm dedicated to a mixture of truncated-shifted Erlangs" in *Springer Journal of Mathematical Imaging and Vision*. [*Submitted*]

International Conferences

- V. Poisson, W. Guicquero, D. Coriat and G. Sicard, "Histogram Compressive Sensing using Shuffled Cellular Automata: the TCSPC sensor use case," 2022 20th *IEEE Interregional NEWCAS Conference (NEWCAS)*, 2022, pp. 124-128, doi: 10.1109/NEWCAS52662.2022.9842077. **Among the 10 best papers of the student contest**

- V. Poisson, W. Guicquero, D. Coriat and G. Sicard, "A 2-Stage EM Algorithm for Online Peak Detection, an Application to TCSPC Data," in *IEEE International Midwest Symposium on Circuits and Systems (MWSCAS)*, vol. 69, no. 9, pp. 3625-3629, Sept. 2022, doi: 10.1109/TCSII.2022.3181687. **Presentation of TCAS-II,** 3^{rd} **place of the student paper contest**

Patent

- V. Poisson and W. Guicquero, ”Méthode de mesure compressive de la distribution statistique d’une grandeur physique”, WO2022/123189.

- V. Poisson and W. Guicquero, ”Méthode de mesure paramétrique en ligne de la distribution statistique d’un signal physique”, WO2022/153016.

Bibliography

[1] Juha Kostamovaara, Sahba S Jahromi, and Pekka Keränen. Temporal and spatial focusing in spad-based solid-state pulsed time-of-flight laser range imaging. *Sensors*, 20(21):5973, 2020.

[2] Maik Beer, Jan F Haase, Jennifer Ruskowski, and Rainer Kokozinski. Background light rejection in spad-based lidar sensors by adaptive photon coincidence detection. *Sensors*, 18(12):4338, 2018.

[3] 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, October 2020.

[4] M Caccia, L Nardo, R Santoro, and D Schaffhauser. Silicon photomultipliers and spad imagers in biophotonics: Advances and perspectives. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 926:101–117, 2019.

[5] Arin Can Ulku, Claudio Bruschini, Ivan Michel Antolović, Yung Kuo, Rinat Ankri, Shimon Weiss, Xavier Michalet, and Edoardo Charbon. A 512 × 512 spad image sensor with integrated gating for widefield flim. *IEEE Journal of Selected Topics in Quantum Electronics*, 25(1):1–12, 2019.

[6] Sam W. Hutchings, Nick Johnston, Istvan Gyongy, Tarek Al Abbas, Neale A. W. Dutton, Max Tyler, Susan Chan, Jonathan Leach, and Robert K. Henderson. A Reconfigurable 3-D-Stacked SPAD Imager With In-Pixel Histogramming for Flash LIDAR or High-Speed Time-of-Flight Imaging. *IEEE Journal of Solid-State Circuits*, 54(11):2947–2956, November 2019.

[7] Ion Vornicu, Ricardo Carmona-Galán, and Ángel Rodríguez-Vázquez. Real-Time Inter-Frame Histogram Builder for SPAD Image Sensors. *IEEE Sensors Journal*, 18(4):1576–1584, February 2018.

[8] Michael P. Sheehan, Julián Tachella, and Mike E. Davies. A Sketching Framework for Reduced Data Transfer in Photon Counting Lidar. *arXiv.org*, February 2021.

[9] Preethi Padmanabhan, Chao Zhang, and Edoardo Charbon. Modeling and Analysis of a Direct Time-of-Flight Sensor Architecture for LiDAR Applications. *Sensors*, 19(24):5464, January 2019.

[10] 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*, 1(2):112–125, June 2015.

[11] Joshua Rapp and Vivek K. Goyal. A Few Photons Among Many: Unmixing Signal and Noise for Photon-Efficient Active Imaging. *IEEE Transactions on Computational Imaging*, 3(3):445–459, September 2017. arXiv: 1609.07407.

[12] Jiayong Peng, Zhiwei Xiong, Xin Huang, Zheng-Ping Li, Dong Liu, and Feihu Xu. Photon-Efficient 3d Imaging with A Non-Local Neural Network. August 2020.

[13] Gongxin Yao, Yiwei Chen, Yong Liu, Xiaomin Hu, and Yu Pan. Robust photon-efficient imaging using a pixel-wise residual shrinkage network. *arXiv preprint arXiv:2201.01453*, 2022.

[14] 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*, pages 273–282, June 2018.

[15] Heon-Cheol Lee, Seung-Hwan Lee, Tae-Seok Lee, Doo-Jin Kim, and Beom-Hee Lee. A survey of map merging techniques for cooperative-slam. In *2012 9th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*, pages 285–287, 2012.

[16] Radu Horaud, Miles Hansard, Georgios Evangelidis, and Clement Menier. An Overview of Depth Cameras and Range Scanners Based on Time-of-Flight Technologies. *arXiv.org*, December 2020.

[17] Yoav Schechner and Nahum Kiryati. Depth from defocus vs. stereo: how different really are they? volume 39, pages 1784 – 1786 vol.2, 09 1998.

[18] Jinbeum Jang, Sangwoo Park, Jieun Jo, and Joonki Paik. Depth map generation using a single image sensor with phase masks. *Optics Express*, 24(12):12868–12878, June 2016.

[19] Edmund Y Lam. Computational photography with plenoptic camera and light field capture: tutorial. *JOSA A*, 32(11):2021–2032, 2015.

[20] (PDF) Depth from defocus vs. stereo: how different really are they? In *ResearchGate*.

[21] Tyler Bell, Beiwen Li, and Song Zhang. Structured Light Techniques and Applications. *Wiley Encyclopedia of Electrical and Electronics Engineering*, pages 1–24, February 2016.

[22] Andreas Quirrenbach. Optical interferometry. *Annual Review of Astronomy and Astrophysics*, 39:353, 2001.

[23] François Piron, Daniel Morrison, Mehmet Rasit Yuce, and Jean-Michel Redouté. A review of single-photon avalanche diode time-of-flight imaging sensor arrays. *IEEE Sensors Journal*, 21(11):12654–12666, 2020.

[24] François Blais. Review of 20 years of range sensor development. *Journal of electronic imaging*, 13(1):231–243, 2004.

[25] 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)*, pages 1–4, December 2018.

[26] Thinal Raj, Fazida Hanim Hashim, Aqilah Baseri Huddin, Mohd Faisal Ibrahim, and Aini Hussain. A survey on lidar scanning mechanisms. *Electronics*, 9(5):741, 2020.

[27] You Li and Javier Ibanez-Guzman. Lidar for autonomous driving: The principles, challenges, and trends for automotive lidar and perception systems. *IEEE Signal Processing Magazine*, 37(4):50–61, 2020.

[28] Olaf T Von Ramm and Stephen W Smith. Beam steering with linear arrays. *IEEE transactions on biomedical engineering*, (8):438–452, 1983.

[29] Craig Benko, Emil Kadlec, Zeb Barber, and Randy Reibel. FMCW lidar for autonomous vehicles. In *High-Power Diode Laser Technology XIX*, volume 11668, page 116680N. International Society for Optics and Photonics, March 2021.

[30] Min-Sun Keel, Daeyun Kim, Yeomyung Kim, Myunghan Bae, Myoungoh Ki, Bumsik Chung, Sooho Son, Hoyong Lee, Heeyoung Jo, Seung-Chul Shin, Sunjoo Hong, Jaeil An, Yonghun Kwon, Sungyoung Seo, Sunghyuck Cho, Youngchan Kim, Young-Gu Jin, Youngsun Oh, Yitae Kim, JungChak Ahn, Kyoungmin Koh, and Yongin Park. 7.1 a 4-tap 3.5 m 1.2 mpixel indirect time-of-flight cmos image sensor with peak current mitigation and multi-user interference cancellation. In *2021 IEEE International Solid- State Circuits Conference (ISSCC)*, volume 64, pages 106–108, 2021.

[31] 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*, 36(12):2435–2442, June 2018.

[32] 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*, 74(7):3263–3270, June 2003.

[33] Bayarto K Lubsandorzhiev. On the history of photomultiplier tube invention. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 567(1):236–238, 2006.

[34] AG Wright. *The photomultiplier handbook*. Oxford University Press, 2017.

[35] David R Smith, David M Walton, Richard Ingley, Andrew D Holland, Mark Cropper, and Peter Pool. Emccds for space applications. In *High Energy, Optical, and Infrared Detectors for Astronomy II*, volume 6276, pages 185–196. SPIE, 2006.

[36] Lijian Zhang, Leonardo Neves, Jeff S Lundeen, and Ian A Walmsley. A characterization of the single-photon sensitivity of an electron multiplying charge-coupled device. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 42(11):114011, 2009.

[37] Afrin Sultana, Ehsan Kamrani, and Mohamad Sawan. Cmos silicon avalanche photodiodes for nir light detection: a survey. *Analog Integrated Circuits and Signal Processing*, 70(1):1–13, 2012.

[38] F Lin, M Mac Sweeney, MM Sheehan, and A Mathewson. A protein biosensor using geiger mode avalanche photodiodes. In *Journal of Physics: Conference Series*, volume 10, page 082. IOP Publishing, 2005.

[39] Yasser El-Batawy, Farseem M Mohammedy, and M Jamal Deen. Resonant cavity enhanced photodetectors: theory, design and modeling. In *Photodetectors*, pages 415–470. Elsevier, 2016.

[40] Feiyang Sun, Yue Xu, Zhong Wu, and Jun Zhang. A simple analytic modeling method for spad timing jitter prediction. *IEEE Journal of the Electron Devices Society*, 7:261–267, 2019.

[41] 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*, 34(15):3610–3615, August 2016.

[42] Matteo Perenzoni, Nicola Massari, Leonardo Gasparini, Manuel Moreno Garcia, Daniele Perenzoni, and David Stoppa. A Fast 50 × 40-Pixels Single-Point DTOF SPAD Sensor With Photon Counting and Programmable ROI TDCs, With <4 mm at 3 m up to 18 klux of Background Light. *IEEE Solid-State Circuits Letters*, 3:86–89, 2020.

[43] J. Kosman, O. Almer, T. A. Abbas, N. Dutton, R. Walker, S. Videv, K. Moore, H. Haas, and R. Henderson. 29.7 A 500mb/s -46.1dbm CMOS SPAD Receiver for Laser Diode Visible-Light Communications. In *2019 IEEE International Solid- State Circuits Conference - (ISSCC)*, pages 468–470, February 2019.

[44] Chockalingam Veerappan and Edoardo Charbon. A substrate isolated cmos spad enabling wide spectral response and low electrical crosstalk. *IEEE Journal of Selected Topics in Quantum Electronics*, 20(6):299–305, 2014.

[45] David Hall, Yu-Hsin Liu, and Yu-Hwa Lo. Single photon avalanche detectors: prospects of new quenching and gain mechanisms. *Nanophotonics*, 4(4):397–412, 2015.

[46] Sergio Cova, Massimo Ghioni, ARTURO Lotito, Ivan Rech, and Franco Zappa. Evolution and prospects for single-photon avalanche diodes and quenching circuits. *Journal of modern optics*, 51(9-10):1267–1288, 2004.

[47] Suhaila Isaak, M.C. Pitter, S. Bull, and I. Harrison. Design and characterisation of 16×1 parallel outputs SPAD array in 0.18 um CMOS technology. In *2010 IEEE Asia Pacific Conference on Circuits and Systems*, pages 979–982, December 2010.

[48] Claudio Bruschini, Harald Homulle, Ivan Michel Antolovic, Samuel Burri, and Edoardo Charbon. Single-photon avalanche diode imagers in biophotonics: review and outlook. *Light: Science & Applications*, 8(1):1–28, 2019.

[49] RK Henderson, BR Rae, and D-U Li. Complementary metal-oxide-semiconductor (cmos) sensors for fluorescence lifetime imaging (flim). In *High Performance Silicon Imaging*, pages 312–347. Elsevier, 2014.

[50] Dennis R Schaart, Edoardo Charbon, Thomas Frach, and Volkmar Schulz. Advances in digital sipms and their application in biomedical imaging. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 809:31–52, 2016.

[51] Christoph Krafft, Benjamin Dietzek, Jürgen Popp, and Michael Schmitt. Raman and coherent anti-stokes raman scattering microspectroscopy for biomedical applications. *Journal of biomedical optics*, 17(4):040801, 2012.

[52] Lidar for automotive and industrial applications 2019. *Yole Développement*, 2019.

[53] 3d imaging and sensing - technology and market trends 2021. *Market Technology Report*, Apr 2021.

[54] Abdullah Abuolaim, Abhijith Punnappurath, and Michael S Brown. Revisiting autofocus for smartphone cameras. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 523–537, 2018.

[55] J. H. Park, J. H. Lee, and I. H. Suh. Issues on simultaneous use of multiple structured light-based RGBD sensors in 3d environmental reconstruction. In *The 18th IEEE International Symposium on Consumer Electronics (ISCE 2014)*, pages 1–2, June 2014.

[56] Gabriella Musarra, Ashley Lyons, Enrico Conca, Yoann Altmann, Federica Villa, F Zappa, Miles J Padgett, and Daniele Faccio. Non-line-of-sight three-dimensional imaging with a single-pixel camera. *Physical Review Applied*, 12(1):011002, 2019.

[57] Ryan E Warburton, Susan Chan, Genevieve Gariepy, Yoann Altmann, Steve McLaughlin, Jonathan Leach, and Daniele Faccio. Real-time tracking of hidden objects with single-pixel detectors. In *Imaging Systems and Applications*, pages IT4E–2. Optica Publishing Group, 2016.

[58] Taerim Yoon, Chang-Seok Kim, Kyujung Kim, and Jong-ryul Choi. Emerging applications of digital micromirror devices in biophotonic fields. *Optics & Laser Technology*, 104:17–25, 2018.

[59] G. Acconcia, M. Crotti, S. Antonioli, I. Rech, and M. Ghioni. High performance time-to-amplitude converter array. In *2013 IEEE Nordic-Mediterranean Workshop on Time-to-Digital Converters (NoMe TDC)*, pages 1–5, 2013.

[60] C Priyanka and P Latha. Design and implementation of time to digital converters. In *2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, pages 1–4. IEEE, 2015.

[61] Zouyi Jiang, Lei Zhao, Xingshun Gao, Ruoshi Dong, Jinxin Liu, and Qi An. Mismatch error correction for time interleaved analog-to-digital converter over a wide frequency range. *Review of Scientific Instruments*, 89(8):084709, 2018.

[62] Luca Vercesi, Antonio Liscidini, and Rinaldo Castello. Two-dimensions vernier time-to-digital converter. *IEEE Journal of Solid-State Circuits*, 45(8):1504–1512, 2010.

[63] Poki Chen, Shen-Luan Liu, and Jingshown Wu. A cmos pulse-shrinking delay element for time interval measurement. *IEEE Transactions on Circuits and Systems II: Analog and digital signal processing*, 47(9):954–958, 2000.

[64] Matthew Z Straayer and Michael H Perrott. A multi-path gated ring oscillator tdc with first-order noise shaping. *IEEE Journal of Solid-State Circuits*, 44(4):1089–1098, 2009.

[65] Mahantesh P Mattada and Hansraj Guhilot. Time-to-digital converters—a comprehensive review. *International Journal of Circuit Theory and Applications*, 49(3):778–800, 2021.

[66] 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)*, pages 1–5, May 2019.

[67] NA Dutton12, Luca Parmesan12, Salvatore Gnecchi12, Istvan Gyongy, Neil Calder, Bruce R Rae, Lindsay A Grant, and Robert K Henderson. Oversampled itof imaging techniques using spad-based quanta image sensors. In *Proc. Int. Image Sensor Workshop*, pages 170–173, 2015.

[68] Istvan Gyongy, Neil Calder, Amy Davies, Neale A. W. Dutton, Rory R. Duncan, Colin Rickman, Paul Dalgarno, and Robert K. Henderson. A 256×256 , 100-kfps, 61time-resolved microscopy applications. *IEEE Transactions on Electron Devices*, 65(2):547–554, 2018.

[69] Cristiano Niclass, Mineki Soga, Hiroyuki Matsubara, Masaru Ogawa, and Manabu Kagami. A 0.18-μ m cmos soc for a 100-m-range 10-frame/s 200 $\times$96-pixel time-of-flight depth sensor. *IEEE Journal of Solid-State Circuits*, 49(1):315–330, 2014.

[70] 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*, 19(6):2181–2190, March 2019.

[71] 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)*, pages 1–4, May 2019.

[72] A. Gupta, A. Ingle, and M. Gupta. Asynchronous Single-Photon 3d Imaging. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7908–7917, October 2019.

[73] Jochen Arlt, David Tyndall, Bruce Rae, David Day-Uei Li, Justin A. Richardson, and Robert K. Henderson. A study of pile-up in integrated time-correlated single photon counting systems. *The Review of scientific instruments*, 84 10:103105, 2013.

[74] Cristiano Niclass, Alexis Rochas, P-A Besse, and Edoardo Charbon. Toward a 3-d camera based on single photon avalanche diodes. *IEEE Journal of selected topics in Quantum Electronics*, 10(4):796–802, 2004.

[75] Julius Schöning and Gunther Heidemann. Taxonomy of 3d sensors. *Argos*, 3(P100):9–10, 2016.

[76] 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, April 2020.

[77] 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*, 17(3):782–795, March 2016.

[78] Hakan Urey, Kishore V Chellappan, Erdem Erden, and Phil Surman. State of the art in stereoscopic and autostereoscopic displays. *Proceedings of the IEEE*, 99(4):540–555, 2011.

[79] Cristiano Niclass, Claudio Favi, Theo Kluter, Marek Gersbach, and Edoardo Charbon. A 128 $\times$ 128 single-photon image sensor with column-level 10-bit time-to-digital converter array. *IEEE Journal of Solid-State Circuits*, 43(12):2977–2989, 2008.

[80] Augusto Carimatto, Shingo Mandai, Esteban Venialgo, Ting Gong, Giacomo Borghi, Dennis R. Schaart, and Edoardo Charbon. 11.4 A 67,392-SPAD PVTB-compensated multichannel digital SiPM with 432 column-parallel 48ps 17b TDCs for endoscopic time-of-flight PET. pages 1–3. IEEE, February 2015.

[81] 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*, pages 40–41,41a, February 2009.

[82] 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)*, pages 96–98, February 2018.

[83] Robert K. Henderson, Nick Johnston, Francescopaolo Mattioli Della Rocca, Haochang Chen, David Day-Uei Li, Graham Hungerford, Richard Hirsch, David Mcloskey, Philip Yip, and David J. S. Birch. A 192 $\times$ 128 time correlated spad image sensor in 40-nm cmos technology. *IEEE Journal of Solid-State Circuits*, 54(7):1907–1916, 2019.

[84] Oichi Kumagai, Junichi Ohmachi, Masao Matsumura, Shinichiro Yagi, Kenichi Tayu, Keitaro Amagawa, Tomohiro Matsukawa, Osamu Ozawa, Daisuke Hirono, Yasuhiro Shinozuka, Ryutaro Homma, Kumiko Mahara, Toshio Ohyama, Yousuke Morita, Shohei Shimada, Takahisa Ueno, Akira Matsumoto, Yusuke Otake, Toshifumi Wakano, and Takashi Izawa. 7.3 a 189×600 back-illuminated stacked spad direct time-of-flight depth sensor for automotive lidar systems. In *2021 IEEE International Solid- State Circuits Conference (ISSCC)*, volume 64, pages 110–112, 2021.

[85] Chao Zhang, Ning Zhang, Zhijie Ma, Letian Wang, Yu Qin, Jieyang Jia, and Kai Zang. A 240 x00d7; 160 3d-stacked spad dtof image sensor with rolling shutter and in-pixel histogram for mobile devices. *IEEE Open Journal of the Solid-State Circuits Society*, 2:3–11, 2022.

[86] Ahmet T. Erdogan, Tarek Al Abbas, Neil Finlayson, Charlotte Hopkinson, Istvan Gyongy, Oscar Almer, Neale A. W. Dutton, and Robert K. Henderson. A high dynamic range 128 xd7; 120 3-d stacked cmos spad image sensor soc for fluorescence microendoscopy. *IEEE Journal of Solid-State Circuits*, 57(6):1649–1660, 2022.

[87] David Stoppa. A reconfigurable qvga/q3vga direct time-of-flight 3d imaging system with on-chip depth-map computation in 45/40 nm 3d-stacked bsi spad cmos. 2021.

[88] Seonghyeok Park, Bumjun Kim, Junhee Cho, Jung-Hoon Chun, Jaehyuk Choi, and Seong-Jin Kim. An 80x00d7;60 flash lidar sensor with in-pixel histogramming tdc based on quaternary search and time-gated x0394;-intensity phase detection for 45m detectable range and background light cancellation. In *2022 IEEE International Solid- State Circuits Conference (ISSCC)*, volume 65, pages 98–100, 2022.

[89] Enrico Manuzzato, Alessandro Tontini, Andrej Seljak, and Matteo Perenzoni. A <tex>64 × 64</tex>-pixel flash lidar spad imager with distributed pixel-to-pixel correlation for background rejection, tunable automatic pixel sensitivity and first-last event detection strategies for space applications. In *2022 IEEE International Solid- State Circuits Conference (ISSCC)*, volume 65, pages 96–98, 2022.

[90] Danilo Bronzi, Federica Villa, Simone Tisa, Alberto Tosi, Franco Zappa, Daniel Durini, Sascha Weyers, and Werner Brockherde. 100 000 frames/s 64 × 32 single-photon detector array for 2-d imaging and 3-d ranging. *IEEE Journal of Selected Topics in Quantum Electronics*, 20(6):355–364, 2014.

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

[92] Byungchoul Park, Injun Park, Woojun Choi, Yoondeok Na, and Youngcheol Chae. A 40-m range 90-frames/s cmos time-of-flight sensor using spad and in-pixel time-gated pulse counter. *IEEE Solid-State Circuits Letters*, 3:422–425, 2020.

[93] Joshua Rapp, Yanting Ma, Robin M. A. Dawson, and Vivek K Goyal. Dead Time Compensation for High-Flux Ranging. *IEEE Transactions on Signal Processing*, 67(13):3471–3486, July 2019.

[94] 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)*, pages 1–4, May 2018.

[95] A Schenk. A model for the field and temperature dependence of shockley-read-hall lifetimes in silicon. *Solid-State Electronics*, 35(11):1585–1596, 1992.

[96] Rüdiger Paschotta. Chromatic dispersion. *Encyclopedia of Laser Physics and Technology*, 2010.

[97] Anthony Mark Fox, Mark Fox, et al. *Quantum optics: an introduction*, volume 15. Oxford university press, 2006.

[98] Abraham Pais. Einstein and the quantum theory. *Reviews of modern physics*, 51(4):863, 1979.

[99] A Vidiella-Barranco and JA Roversi. Quantum superpositions of binomial states of light. *Journal of Modern Optics*, 42(12):2475–2493, 1995.

[100] Alessandro Tontini, Leonardo Gasparini, and Matteo Perenzoni. Numerical Model of SPAD-Based Direct Time-of-Flight Flash LIDAR CMOS Image Sensors. *Sensors*, 20(18):5203, January 2020.

[101] 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)*, pages 1558–1562, October 2018.

[102] 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*, 66(1):497–504, January 2019.

[103] 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*, 21(3):5–16, May 2015.

[104] 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)*, pages 1–4, May 2018.

[105] Antonino Ingargiola, Maya Segal, Angelo Gulinatti, Ivan Rech, Ivan Labanca, Piera Maccagnani, Massimo Ghioni, Shimon Weiss, and Xavier 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*, 9(12):255–258, December 2018.

[106] ON semiconductor. Sipms in direct tof ranging applications. *White paper*, (TND6254/D), June 2020.

[107] Marcel F. Neuts. Renewal processes of phase type. *Naval Research Logistics Quarterly*, 25(3):445–454, September 1978.

[108] Amelia Carolina Sparavigna. Poissonian distributions in physics: Counting electrons and photons. 2021.

[109] Marcel F. Neuts. Renewal processes of phase type. *Naval Research Logistics Quarterly*, 25(3):445–454, 1978.

[110] Oliver C. Ibe. 1 - Basic Concepts in Probability. In Oliver C. Ibe, editor, *Markov Processes for Stochastic Modeling (Second Edition)*, pages 1–27. Elsevier, Oxford, second edition edition, 2013.

[111] Erhan Cinlar and RA Agnew. On the superposition of point processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30(3):576–581, 1968.

[112] Harry Crane and Peter Mccullagh. Poisson superposition processes. *Journal of Applied Probability*, 52(4):1013–1027, 2015.

[113] M. F. NEUTS. Probability distributions of phase type. *Liber Amicorum Prof. Emeritus H. Florin*, 1975.

[114] Oliver C. Fundamentals of applied probability and random processes. *Academic Press*, 2014.

[115] LM McGrath, B Mustanski, A Metzger, DS Pine, E Kistner-Griffin, E Cook, and LS Wakschlag. A latent modeling approach to genotype–phenotype relationships: Maternal problem behavior clusters, prenatal smoking, and maoa genotype. *Archives of women's mental health*, 15(4):269–282, 2012.

[116] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, September 1977.

[117] James V Stone. Bayes' rule: a tutorial introduction to bayesian analysis. 2013.

[118] Kenneth Lange. A gradient algorithm locally equivalent to the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(2):425–437, 1995.

[119] D Michael Titterington. Recursive parameter estimation using incomplete data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 46(2):257–267, 1984.

[120] Christopher M Bishop. Mixture models and the em algorithm. *Microsoft Research, Cambridge*, 2006.

[121] J MacQueen. Classification and analysis of multivariate observations. In *5th Berkeley Symp. Math. Statist. Probability*, pages 281–297, 1967.

[122] Gilles Celeux and Jean Diebolt. L'algorithme sem: un algorithme d'apprentissage probabiliste pour la reconnaissance de mélange de densités. *Revue de statistique appliquée*, 34(2):35–52, 1986.

[123] Sparsh Mittal. A survey of techniques for approximate computing. *ACM Computing Surveys (CSUR)*, 48(4):1–33, 2016.

[124] Jie Han and Michael Orshansky. Approximate computing: An emerging paradigm for energy-efficient design. In *2013 18th IEEE European Test Symposium (ETS)*, pages 1–6. IEEE, 2013.

[125] Haoliang Li, Jianhao Hu, and Jienan Chen. A novel low-power filter design via reduced-precision redundancy for voltage overscaling applications. In *2013 IEEE Global Communications Conference (GLOBECOM)*, pages 3282–3287. IEEE, 2013.

[126] Luigi Dadda. Some schemes for fast serial input multipliers. In *1983 IEEE 6th Symposium on Computer Arithmetic (ARITH)*, pages 52–59, 1983.

[127] Vaibhav Gupta, Debabrata Mohapatra, Anand Raghunathan, and Kaushik Roy. Low-power digital signal processing using approximate adders. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 32:124–137, 2013.

[128] Joshua Yung Lih Low and Ching Chuen Jong. Non-iterative high speed division computation based on mitchell logarithmic method. In *2013 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 2219–2222. IEEE, 2013.

[129] Stelios Sidiroglou-Douskos, Sasa Misailovic, Henry Hoffmann, and Martin Rinard. Managing performance vs. accuracy trade-offs with loop perforation. In *Proceedings of the 19th ACM SIGSOFT symposium and the 13th European conference on Foundations of software engineering*, pages 124–134, 2011.

[130] Duy Thanh Nguyen, Hyun Kim, Hyuk-Jae Lee, and Ik-Joon Chang. An approximate memory architecture for a reduction of refresh power consumption in deep learning applications. In *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5. IEEE, 2018.

[131] Weiqiang Liu, Jing Li, Tao Xu, Chenghua Wang, Paolo Montuschi, and Fabrizio Lombardi. Combining restoring array and logarithmic dividers into an approximate hybrid design. In *2018 IEEE 25th Symposium on Computer Arithmetic (ARITH)*, pages 92–98, 2018.

[132] Emiel Por, M v Kooten, and Vanja Sarkovic. Nyquist-shannon sampling theorem. *Leiden University*, 1:1, 2019.

[133] Robert H Walden. Analog-to-digital converter survey and analysis. *IEEE Journal on selected areas in communications*, 17(4):539–550, 1999.

[134] Emmanuel J Candès et al. Compressive sampling. In *Proceedings of the international congress of mathematicians*, volume 3, pages 1433–1452. Citeseer, 2006.

[135] Pierre Duhamel and Martin Vetterli. Fast fourier transforms: a tutorial review and a state of the art. *Signal processing*, 19(4):259–299, 1990.

[136] AM Raid, WM Khedr, Mohamed A El-Dosuky, and Wesam Ahmed. Jpeg image compression using discrete cosine transform-a survey. *arXiv preprint arXiv:1405.6147*, 2014.

[137] Christopher E Heil and David F Walnut. Continuous and discrete wavelet transforms. *SIAM review*, 31(4):628–666, 1989.

[138] Emmanuel J Candès and Michael B Wakin. An introduction to compressive sampling. *IEEE signal processing magazine*, 25(2):21–30, 2008.

[139] Ljubisa Stankovic, Danilo P Mandic, Milos Dakovic, and Ilya Kisil. Demystifying the coherence index in compressive sensing [lecture notes]. *IEEE Signal Processing Magazine*, 37(1):152–162, 2020.

[140] Emmanuel J Candes and Terence Tao. Decoding by linear programming. *IEEE transactions on information theory*, 51(12):4203–4215, 2005.

[141] Joel Tropp, Anna C Gilbert, et al. Signal recovery from partial information via orthogonal matching pursuit. *IEEE Trans. Inform. Theory*, 53(12):4655–4666, 2007.

[142] Hadi Zayyani, Massoud Babaie-Zadeh, and Christian Jutten. Bayesian pursuit algorithm for sparse representation. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1549–1552. IEEE, 2009.

[143] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

[144] Manya V Afonso, José M Bioucas-Dias, and Mário AT Figueiredo. An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE transactions on image processing*, 20(3):681–695, 2010.

[145] Stephen J Wright. Coordinate descent algorithms. *Mathematical Programming*, 151(1):3–34, 2015.

[146] Bjorn Johansson, Tamás Keviczky, Mikael Johansson, and Karl Henrik Johansson. Subgradient methods and consensus algorithms for solving convex optimization problems. In *2008 47th IEEE Conference on Decision and Control*, pages 4185–4190. IEEE, 2008.

[147] Bradley Efron, Trevor Hastie, Iain Johnstone, and Robert Tibshirani. Least angle regression. *The Annals of statistics*, 32(2):407–499, 2004.

[148] Neal Parikh, Stephen Boyd, et al. Proximal algorithms. *Foundations and trends® in Optimization*, 1(3):127–239, 2014.

[149] Diego Vidaurre, Concha Bielza, and Pedro Larranaga. A survey of l1 regression. *International Statistical Review*, 81(3):361–387, 2013.

[150] Emmanuel Candes and Terence Tao. The dantzig selector: Statistical estimation when p is much larger than n. *The annals of Statistics*, 35(6):2313–2351, 2007.

[151] Noah Simon, Jerome Friedman, Trevor Hastie, and Robert Tibshirani. A sparse-group lasso. *Journal of computational and graphical statistics*, 22(2):231–245, 2013.

[152] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, 67(2):301–320, 2005.

[153] Ming Yuan and Yi Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67, 2006.

[154] Peng Zhao, Guilherme Rocha, and Bin Yu. The composite absolute penalties family for grouped and hierarchical variable selection. *The Annals of Statistics*, 37(6A):3468–3497, 2009.

[155] Xin Yuan, Tsung-Han Tsai, Ruoyu Zhu, Patrick Llull, David Brady, and Lawrence Carin. Compressive hyperspectral imaging with side information. *IEEE Journal of selected topics in Signal Processing*, 9(6):964–976, 2015.

[156] Fengqiang Li, Huaijin Chen, Chia-Kai Yeh, Adithya Pediredla, Kuan He, Ashok Veeraghvan, and Oliver Cossairt. Compressive time-of-flight imaging. In *Applied Industrial Optics: Spectroscopy, Imaging and Metrology*, pages AM2A–5. Optica Publishing Group, 2018.

[157] CJ Miosso, R Von Borries, and JH Pierluissi. Compressive sensing method for improved reconstruction of gradient-sparse magnetic resonance images. In *2009 conference record of the forty-third Asilomar conference on signals, systems and computers*, pages 799–806. IEEE, 2009.

[158] Kévin Degraux, Valerio Cambareri, Bert Geelen, Laurent Jacques, and Gauthier Lafruit. Multispectral compressive imaging strategies using fabry–pérot filtered sensors. *IEEE Transactions on Computational Imaging*, 4(4):661–673, 2018.

[159] Marco F Duarte, Mark A Davenport, Dharmpal Takhar, Jason N Laska, Ting Sun, Kevin F Kelly, and Richard G Baraniuk. Single-pixel imaging via compressive sampling. *IEEE signal processing magazine*, 25(2):83–91, 2008.

[160] Wai Lam Chan, Kriti Charan, Dharmpal Takhar, Kevin F Kelly, Richard G Baraniuk, and Daniel M Mittleman. A single-pixel terahertz imaging system based on compressed sensing. *Applied Physics Letters*, 93(12):121105, 2008.

[161] Prasan A Shedligeri, Sreyas Mohan, and Kaushik Mitra. Data driven coded aperture design for depth recovery. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 56–60. IEEE, 2017.

[162] Hoover Rueda, Henry Arguello, and Gonzalo R Arce. Colored coded aperture compressive spectral imaging: Design and experimentation. In *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 601–604. IEEE, 2015.

[163] Dikpal Reddy, Ashok Veeraraghavan, and Rama Chellappa. P2c2: Programmable pixel compressive camera for high speed imaging. In *CVPR 2011*, pages 329–336. IEEE, 2011.

[164] Munir M El-Desouki, Ognian Marinov, M Jamal Deen, and Qiyin Fang. Cmos active-pixel sensor with in-situ memory for ultrahigh-speed imaging. *IEEE Sensors Journal*, 11(6):1375–1379, 2010.

[165] Munir El-Desouki, M Jamal Deen, Qiyin Fang, Louis Liu, Frances Tse, and David Armstrong. Cmos image sensors for high speed applications. *Sensors*, 9(1):430–444, 2009.

[166] Vahid Majidzadeh, Laurent Jacques, Alexandre Schmid, Pierre Vandergheynst, and Yusuf Leblebici. A (256× 256) pixel 76.7 mw cmos imager/compressor based on real-time in-pixel compressive sensing. In *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*, pages 2956–2959. IEEE, 2010.

[167] Ryan Robucci, Jordan D Gray, Leung Kin Chiu, Justin Romberg, and Paul Hasler. Compressive sensing on a cmos separable-transform image sensor. *Proceedings of the IEEE*, 98(6):1089–1101, 2010.

[168] Yusuke Oike and Abbas El Gamal. Cmos image sensor with per-column σ δ adc and programmable compressed sensing. *IEEE Journal of Solid-State Circuits*, 48(1):318–328, 2012.

[169] Hsuan-Tsung Wang and Walter D. Leon-Salas. An incremental sigma delta converter for compressive sensing applications. In *2011 IEEE International Symposium of Circuits and Systems (ISCAS)*, pages 522–525, 2011.

[170] William Guicquero, Arnaud Verdant, and Antoine Dupret. High-order incremental sigma–delta for compressive sensing and its application to image sensors. *Electronics Letters*, 51(19):1492–1494, 2015.

[171] William Guicquero, Antoine Dupret, and Pierre Vandergheynst. An algorithm architecture co-design for cmos compressive high dynamic range imaging. *IEEE Transactions on Computational Imaging*, 2(3):190–203, 2016.

[172] Jean-Jacques Greffet, Patrick Bouchon, Giovanni Brucoli, Emilie Sakat, and François Marquier. Generalized kirchhoff law. *arXiv preprint arXiv:1601.00312*, 2016.

[173] Wissam Benjilali, William Guicquero, Laurent Jacques, and Gilles 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*, 67(4):1218–1231, 2020.

[174] John McCarthy. Artificial intelligence, logic and formalizing common sense. In *Philosophical logic and artificial intelligence*, pages 161–190. Springer, 1989.

[175] Timothy O'Shea and Jakob Hoydis. An introduction to deep learning for the physical layer. *IEEE Transactions on Cognitive Communications and Networking*, 3(4):563–575, 2017.

[176] Subhankar Roy, Willi Menapace, Sebastiaan Oei, Ben Luijten, Enrico Fini, Cristiano Saltori, Iris Huijben, Nishith Chennakeshava, Federico Mento, Alessandro Sentelli, Emanuele Peschiera, Riccardo Trevisan, Giovanni Maschietto, Elena Torri, Riccardo Inchingolo, Andrea Smargiassi, Gino Soldati, Paolo Rota, Andrea Passerini, Ruud J. G. van Sloun, Elisa Ricci, and Libertario Demi. Deep learning for classification and localization of covid-19 markers in point-of-care lung ultrasound. *IEEE Transactions on Medical Imaging*, 39(8):2676–2687, 2020.

[177] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3431–3440, 2015.

[178] Tannu Chauhan, Surbhi Rawat, Samrath Malik, and Pushpa Singh. Supervised and unsupervised machine learning based review on diabetes care. In *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, volume 1, pages 581–585, 2021.

[179] Muhammad Usama, Junaid Qadir, Aunn Raza, Hunain Arif, Kok-lim Alvin Yau, Yehia Elkhatib, Amir Hussain, and Ala Al-Fuqaha. Unsupervised machine learning for networking: Techniques, applications and research challenges. *IEEE Access*, 7:65579–65615, 2019.

[180] M. Belkin, I. Matveeva, and P. Niyogi. Tikhonov regularization and semi-supervised learning on large graphs. In *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 3, pages iii–1000, 2004.

[181] Raniah Zaheer and Humera Shaziya. A study of the optimization algorithms in deep learning. In *2019 third international conference on inventive systems and control (ICISC)*, pages 536–539. IEEE, 2019.

[182] Derya Soydaner. A comparison of optimization algorithms for deep learning. *International Journal of Pattern Recognition and Artificial Intelligence*, 34(13):2052013, 2020.

[183] Sebastian Ruder. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*, 2016.

[184] Robert Hecht-Nielsen. Theory of the backpropagation neural network. In *Neural networks for perception*, pages 65–93. Elsevier, 1992.

[185] Christian Darken, Joseph Chang, John Moody, et al. Learning rate schedules for faster stochastic gradient search. In *Neural networks for signal processing*, volume 2. Citeseer, 1992.

[186] Mahesh Chandra Mukkamala and Matthias Hein. Variants of rmsprop and adagrad with logarithmic regret bounds. In *International conference on machine learning*, pages 2545–2553. PMLR, 2017.

[187] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[188] Michael Mathieu, Camille Couprie, and Yann LeCun. Deep multi-scale video prediction beyond mean square error. *arXiv preprint arXiv:1511.05440*, 2015.

[189] Harold C Burger, Christian J Schuler, and Stefan Harmeling. Image denoising: Can plain neural networks compete with bm3d? In *2012 IEEE conference on computer vision and pattern recognition*, pages 2392–2399. IEEE, 2012.

[190] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz. Loss functions for image restoration with neural networks. *IEEE Transactions on Computational Imaging*, 3(1):47–57, 2017.

[191] Andrea Asperti and Matteo Trentin. Balancing reconstruction error and kullback-leibler divergence in variational autoencoders. *IEEE Access*, 8:199440–199448, 2020.

[192] Peter L Bartlett and Marten H Wegkamp. Classification with a reject option using a hinge loss. *Journal of Machine Learning Research*, 9(8), 2008.

[193] William S Noble. What is a support vector machine? *Nature biotechnology*, 24(12):1565–1567, 2006.

[194] Kunihiko Fukushima. Neocognitron: A hierarchical neural network capable of visual pattern recognition. *Neural networks*, 1(2):119–130, 1988.

[195] David H Hubel and Torsten N Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160(1):106, 1962.

[196] Alexander Waibel, Toshiyuki Hanazawa, Geoffrey Hinton, Kiyohiro Shikano, and Kevin J Lang. Phoneme recognition using time-delay neural networks. *IEEE transactions on acoustics, speech, and signal processing*, 37(3):328–339, 1989.

[197] Yann LeCun, Bernhard Boser, John Denker, Donnie Henderson, Richard Howard, Wayne Hubbard, and Lawrence Jackel. Handwritten digit recognition with a back-propagation network. *Advances in neural information processing systems*, 2, 1989.

[198] Vivienne Sze, Yu-Hsin Chen, Tien-Ju Yang, and Joel S Emer. Efficient processing of deep neural networks: A tutorial and survey. *Proceedings of the IEEE*, 105(12):2295–2329, 2017.

[199] P Sibi, S Allwyn Jones, and P Siddarth. Analysis of different activation functions using back propagation neural networks. *Journal of theoretical and applied information technology*, 47(3):1264–1268, 2013.

[200] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

[201] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[202] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[203] Shuiwang Ji, Wei Xu, Ming Yang, and Kai Yu. 3d convolutional neural networks for human action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):221–231, 2013.

[204] Di Wu, Lionel Pigou, Pieter-Jan Kindermans, Nam Do-Hoang Le, Ling Shao, Joni Dambre, and Jean-Marc Odobez. Deep dynamic neural networks for multimodal gesture segmentation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(8):1583–1597, 2016.

[205] Pavlo Molchanov, Xiaodong Yang, Shalini Gupta, Kihwan Kim, Stephen Tyree, and Jan Kautz. Online detection and classification of dynamic hand gestures with recurrent 3d convolutional neural networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4207–4215, 2016.

[206] Qi Dou, Hao Chen, Lequan Yu, Lei Zhao, Jing Qin, Defeng Wang, Vincent CT Mok, Lin Shi, and Pheng-Ann Heng. Automatic detection of cerebral microbleeds from mr images via 3d convolutional neural networks. *IEEE Transactions on Medical Imaging*, 35(5):1182–1195, 2016.

[207] Clara Callenberg, Ashley Lyons, D den Brok, Areeba Fatima, Alex Turpin, Vytautas Zickus, Laura Machesky, Jamie Whitelaw, Daniele Faccio, and MB Hullin. Super-resolution time-resolved imaging using computational sensor fusion. *Scientific reports*, 11(1):1–8, 2021.

[208] Qilin Sun, Jian Zhang, Xiong Dun, Bernard Ghanem, Yifan Peng, and Wolfgang Heidrich. End-to-end learned, optically coded super-resolution spad camera. *ACM Transactions on Graphics (TOG)*, 39(2):1–14, 2020.

[209] Olivier Rukundo and Hanqiang Cao. Nearest neighbor value interpolation. *arXiv preprint arXiv:1211.1768*, 2012.

[210] Kim T Gribbon and Donald G Bailey. A novel approach to real-time bilinear interpolation. In *Proceedings. DELTA 2004. Second IEEE international workshop on electronic design, test and applications*, pages 126–131. IEEE, 2004.

[211] Shengkui Gao and Viktor Gruev. Bilinear and bicubic interpolation methods for division of focal plane polarimeters. *Optics express*, 19(27):26161–26173, 2011.

[212] Robert Keys. Cubic convolution interpolation for digital image processing. *IEEE transactions on acoustics, speech, and signal processing*, 29(6):1153–1160, 1981.

[213] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer, 2014.

[214] Matthew D Zeiler, Dilip Krishnan, Graham W Taylor, and Rob Fergus. Deconvolutional networks. In *2010 IEEE Computer Society Conference on computer vision and pattern recognition*, pages 2528–2535. IEEE, 2010.

[215] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883, 2016.

[216] Sebastian Schuon, Christian Theobalt, James Davis, and Sebastian Thrun. Lidarboost: Depth superresolution for tof 3d shape scanning. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 343–350. IEEE, 2009.

[217] Qingxiong Yang, Ruigang Yang, James Davis, and David Nistér. Spatial-depth super resolution for range images. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007.

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

[219] Time series analysis and forecasting techniques for municipal solid waste management. *Resources, Conservation and Recycling*, 35(3):201–214, May 2002.

[220] Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50:159–175, January 2003.

[221] Gen Sakoda, Hideki Takayasu, and Misako Takayasu. Tracking Poisson Parameter for Non-Stationary Discontinuous Time Series with Taylor's Abnormal Fluctuation Scaling. *Stats*, 2(1):55–69, March 2019.

[222] Efficient simulation of non-Poisson non-stationary point processes to study queueing approximations. *Statistics & Probability Letters*, 109:202–207, February 2016.

[223] A. Cominelli, G. Acconcia, P. Peronio, I. Rech, and M. Ghioni. Readout Architectures for High Efficiency in Time-Correlated Single Photon Counting Experiments—Analysis and Review. *IEEE Photonics Journal*, 9(3):1–15, June 2017.

[224] Jie Chen and Arjun K. Gupta. *Parametric Statistical Change Point Analysis*. Birkhäuser Boston, Boston, 2012. DOI: 10.1007/978-0-8176-4801-5.

[225] Søren Asmussen, Olle Nerman, and Marita Olsson. Fitting Phase-Type Distributions via the EM Algorithm. *Scandinavian Journal of Statistics*, 23(4):419–441, 1996.

[226] A. Thummler, P. Buchholz, and M. Telek. A Novel Approach for Phase-Type Fitting with the EM Algorithm. *IEEE Transactions on Dependable and Secure Computing*, 3(3):245–258, July 2006.

[227] Hiroyuki Okamura, Tadashi Dohi, and Kishor S. Trivedi. Improvement of expectation–maximization algorithm for phase-type distributions with grouped and truncated data. *Applied Stochastic Models in Business and Industry*, 29(2):141–156, March 2013.

[228] Wei Xie, Barry L. Nelson, and Russell R. Barton. A bayesian framework for quantifying uncertainty in stochastic simulation. *Operations Research*, 62(6):1439–1452, November 2014.

[229] Sheldon H. Jacobson and Enver Yücesan. Analyzing the Performance of Generalized Hill Climbing Algorithms. *Journal of Heuristics*, 10(4):387–405, July 2004.

[230] Gilles Celeux, Didier Chauveau, and Jean Diebolt. On Stochastic Versions of the EM Algorithm. report, INRIA, 1995.

[231] F. Pernkopf and D. Bouchaffra. Genetic-based EM algorithm for learning Gaussian mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1344–1348, August 2005.

[232] Genetic-based EM algorithm for learning Gaussian mixture models.

[233] R. Eberhart and J. Kennedy. A new optimizer using particle swarm theory. In *MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, pages 39–43, October 1995.

[234] M. Srinivas and L. M. Patnaik. Genetic algorithms: a survey. *Computer*, 27(6):17–26, June 1994.

[235] Geoffrey J. McLachlan and Suren Rathnayake. On the number of components in a Gaussian mixture model. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 4(5):341–355, September 2014.

[236] C. Biernacki, G. Celeux, and G. Govaert. Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(7):719–725, July 2000.

[237] Donald B. Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, December 1976.

[238] R. Zhou, J. Liu, S. Kumar, and D. P. Palomar. Student's t VAR Modeling With Missing Data Via Stochastic EM and Gibbs Sampling. *IEEE Transactions on Signal Processing*, 68:6198–6211, 2020.

[239] J. Zhang and P. Yin. Multivariate Time Series Missing Data Imputation Using Recurrent Denoising Autoencoder. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 760–764, November 2019.

[240] Daniel J. Stekhoven and Peter Bühlmann. MissForest - nonparametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1):112–118, January 2012. arXiv: 1105.0828.

[241] Muhammad S. Osman, Adnan M. Abu-Mahfouz, and Philip R. Page. A Survey on Data Imputation Techniques: Water Distribution System as a Use Case. *IEEE Access*, 6:63279–63291, 2018.

[242] Cold Deck Imputation for Survey Non-response Through Record Linkage | Request PDF. In *ResearchGate*.

[243] Rebecca R. Andridge and Roderick J. A. Little. A Review of Hot Deck Imputation for Survey Non-response. *International statistical review = Revue internationale de statistique*, 78(1):40–64, April 2010.

[244] Christophe Crambes and Yousri Henchiri. Regression imputation in the functional linear model with missing values in the response. *Journal of Statistical Planning and Inference*, 201:103–109, 2019.

[245] Hyun Kang. The prevention and handling of the missing data. *Korean Journal of Anesthesiology*, 64(5):402–406, May 2013.

[246] Søren Asmussen, Olle Nerman, and Marita Olsson. Fitting phase-type distributions via the em algorithm. *Scandinavian Journal of Statistics*, pages 419–441, 1996.

[247] Axel Thummler, Peter Buchholz, and Miklós Telek. A novel approach for phase-type fitting with the em algorithm. *IEEE Transactions on dependable and secure computing*, 3(3):245–258, 2006.

[248] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment Matching for Multi-Source Domain Adaptation. pages 1406–1415, 2019.

[249] Pierre Hansen and Nenad Mladenović. An Introduction to Variable Neighborhood Search. In *Meta-Heuristics*, pages 433–458. Springer, Boston, MA, 1999. DOI: 10.1007/978-1-4615-5775-3_30.

[250] François Bavaud. Relative Entropy and Statistics. *arXiv:0808.4111 [cs, math, stat]*, August 2008. arXiv: 0808.4111.

[251] Rebecca R Andridge and Roderick JA Little. A review of hot deck imputation for survey non-response. *International statistical review*, 78(1):40–64, 2010.

[252] T. J. Ayoub, K. Basu, and D. J. Jeffrey. Recent results on the lambert w function. In *2020 22nd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, pages 36–39, 2020.

[253] Yingjie Tian and Yuqi Zhang. A comprehensive survey on regularization strategies in machine learning. *Inf. Fusion*, 80(C):146–166, apr 2022.

[254] Mary L. McHugh. The Chi-square test of independence. *Biochemia Medica*, 23(2):143–149, June 2013.

[255] Shamsul Huda, John Yearwood, and Roberto Togneri. Hybrid Metaheuristic Approaches to the Expectation Maximization for Estimation of the Hidden Markov Model for Signal Modeling. *IEEE Transactions on Cybernetics*, 44(10):1962–1977, October 2014.

[256] Valentin Poisson, William Guicquero, Gilles Sicard, et al. Luminance-depth reconstruction from compressed time-of-flight histograms. *IEEE Transactions on Computational Imaging*, 8:148–161, 2022.

[257] H. Hirschmuller and D. Scharstein. Evaluation of Cost Functions for Stereo Matching. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2007.

[258] Philip Gatt, Steven Johnson, and Terry Nichols. Geiger-mode avalanche photodiode ladar receiver performance characteristics and detection statistics. *Applied optics*, 48(17):3261–3276, 2009.

[259] Roel Verbelen, Lan Gong, Katrien Antonio, Andrei Badescu, and Sheldon Lin. Fitting mixtures of erlangs to censored and truncated data using the em algorithm. *ASTIN Bulletin: The Journal of the IAA*, 45(3):729–758, 2015.

[260] Tsz Chai Fung, Andrei L Badescu, and X Sheldon Lin. Fitting censored and truncated regression data using the mixture of experts models. *North American Actuarial Journal*, pages 1–25, 2021.

[261] In Jae Myung. Tutorial on maximum likelihood estimation. *Journal of mathematical Psychology*, 47(1):90–100, 2003.

[262] Mary A Johnson and Michael R Taaffe. Matching moments to phase distributions: Mixtures of erlang distributions of common order. *Stochastic Models*, 5(4):711–743, 1989.

[263] Hiroyuki Okamura, Tadashi Dohi, and Kishor S Trivedi. Improvement of expectation–maximization algorithm for phase-type distributions with grouped and truncated data. *Applied Stochastic Models in Business and Industry*, 29(2):141–156, 2013.

[264] Pierre Hansen and Nenad Mladenović. An introduction to variable neighborhood search. In *Meta-heuristics*, pages 433–458. Springer, 1999.

[265] Mandavilli Srinivas and Lalit M Patnaik. Genetic algorithms: A survey. *computer*, 27(6):17–26, 1994.

[266] LAWRENCE M. LEEMIS. TECHNICAL NOTE: Nonparametric estimation and variate generation for a nonhomogeneous Poisson process from event count data. *IIE Transactions*, 36(12):1155–1160, December 2004.

[267] M.G. Thomason and J.A. Whittaker. Rare failure-state in a Markov chain model for software reliability. In *Proceedings 10th International Symposium on Software Reliability Engineering (Cat. No.PR00443)*, pages 12–19, November 1999. ISSN: 1071-9458.

[268] Alexander Herbertsson. Modelling default contagion using multivariate phase-type distributions. *Review of Derivatives Research*, 14(1):1–36, April 2011.

[269] Bumjun Kim, Seonghyeok Park, Jung-Hoon Chun, Jaehyuk Choi, and Seong-Jin Kim. 7.2 A 48×40 13.5mm Depth Resolution Flash LiDAR Sensor with In-Pixel Zoom Histogramming Time-to-Digital Converter. In *2021 IEEE International Solid- State Circuits Conference (ISSCC)*, volume 64, pages 108–110, February 2021.

[270] H. Seo and J. Choi. Histogram-based mixed-signal time-to-digital-converter array for direct time-of-flight depth sensors. *Electronics Letters*, 55(6):310–312, March 2019.

[271] Mihail Georgiev, Robert Bregović, and Atanas Gotchev. Fixed-Pattern Noise Modeling and Removal in Time-of-Flight Sensing. *IEEE Transactions on Instrumentation and Measurement*, 65(4):808–820, April 2016.

[272] Julián Tachella, Yoann Altmann, Ximing Ren, Aongus McCarthy, Gerald S. Buller, Jean-Yves Tourneret, and Steve McLaughlin. Bayesian 3d Reconstruction of Complex Scenes from Single-Photon Lidar Data. *SIAM Journal on Imaging Sciences*, 12(1):521–550, January 2019. arXiv: 1810.11633.

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

[274] Ahmed Kirmani, Dheera Venkatraman, Dongeek Shin, Andrea Colaço, Franco N. C. Wong, Jeffrey H. Shapiro, and Vivek K. Goyal. First-Photon Imaging. *Science*, 343(6166):58–61, January 2014.

[275] Nicolas Keriven, Anthony Bourrier, Rémi Gribonval, and Patrick Pérez. Sketching for large-scale learning of mixture models. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6190–6194, 2016.

[276] Mariangela Genovese and Ettore Napoli. ASIC and FPGA Implementation of the Gaussian Mixture Model Algorithm for Real-Time Segmentation of High Definition Video. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 22(3):537–547, March 2014.

[277] Minghua Shi and Amine Bermak. An Efficient Digital VLSI Implementation of Gaussian Mixture Models-Based Classifier. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 14(9):962–974, September 2006.

[278] Elham Sarbazi, Majid Safari, and Harald Haas. Statistical Modeling of Single-Photon Avalanche Diode Receivers for Optical Wireless Communications. *IEEE Transactions on Communications*, 66(9):4043–4058, September 2018.

[279] Olivier Cappé. Online EM Algorithm for Hidden Markov Models. *arXiv:0908.2359 [stat]*, August 2009. arXiv: 0908.2359.

[280] Nicolas Keriven, Anthony Bourrier, Rémi Gribonval, and Patrick Pérez. Sketching for Large-Scale Learning of Mixture Models. *Information and Inference*, 7(3):447–508, September 2018.

[281] Theory of cellular automata: A survey. *Theoretical Computer Science*, 334(1-3):3–33, April 2005.

[282] Ivan Nourdin, Giovanni Peccati, and Gesine Reinert. Stein's method and stochastic analysis of Rademacher functionals. *arXiv.org*, October 2008.

[283] Maxim Raginsky, Rebecca M. Willett, Zachary T. Harmany, and Roummel F. Marcia. Compressed sensing performance bounds under Poisson noise. *IEEE Transactions on Signal Processing*, 58(8):3990–4002, August 2010. arXiv: 0910.5146.

[284] Morteza Mardani, Enhao Gong, Joseph Y Cheng, Shreyas S Vasanawala, Greg Zaharchuk, Lei Xing, and John M Pauly. Deep generative adversarial neural networks for compressive sensing mri. *IEEE transactions on medical imaging*, 38(1):167–179, 2018.

[285] Valentin Poisson, William Guicquero, David Coriat, and Gilles Sicard. Histogram compressive sensing using shuffled cellular automata: the tcspc sensor use case. In *2022 20th IEEE Interregional NEWCAS Conference (NEWCAS)*, pages 124–128, 2022.

[286] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[287] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

[288] Daniel J. McDonald and Cosma Rohilla Shalizi. Rademacher complexity of stationary sequences. *arXiv:1106.0730 [cs, stat]*, June 2011. arXiv: 1106.0730.

[289] Li Tian, Guorui Li, and Cong Wang. A data reconstruction algorithm based on neural network for compressed sensing. In *2017 Fifth International Conference on Advanced Cloud and Big Data (CBD)*, pages 291–295. IEEE, 2017.

[290] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007.

[291] Tak-Wai Hui, Chen Change Loy, and Xiaoou Tang. Depth Map Super-Resolution by Deep Multi-Scale Guidance. In *Computer Vision – ECCV 2016*, Lecture Notes in Computer Science, pages 353–369. Springer, Cham, October 2016.

[292] 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*, pages 993–1000, December 2013.

[293] Alice Ruget, Stephen McLaughlin, Robert K. Henderson, Istvan Gyongy, Abderrahim Halimi, and Jonathan Leach. Robust super-resolution depth imaging via a multi-feature fusion deep network. *arXiv.org*, November 2020.

[294] E. J. Candes and M. B. Wakin. An Introduction To Compressive Sampling. *IEEE Signal Processing Magazine*, 25(2):21–30, March 2008.

[295] 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*, pages 96–102, June 2012. ISSN: 1063-6919.

[296] Gregory A. Howland, Daniel J. Lum, Matthew R. Ware, and John C. Howell. Photon counting compressive depth mapping. *Optics Express*, 21(20):23822–23837, October 2013.

[297] A. Farina, A. Farina, A. Candeo, A. Dalla 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*, 27(22):31889–31899, October 2019.

[298] Zhanghao Sun, David B. Lindell, Olav Solgaard, and Gordon Wetzstein. SPADnet: deep RGB-SPAD sensor fusion assisted by monocular depth estimation. *Optics Express*, 28(10):14948–14962, May 2020.

[299] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[300] 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*, 55(7):1960–1970, July 2020.

[301] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor Segmentation and Support Inference from RGBD Images. In *Computer Vision – ECCV 2012*, Lecture Notes in Computer Science, pages 746–760. Springer, Berlin, Heidelberg, October 2012.

[302] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. *arXiv.org*, December 2014.

[303] C. Callenberg, A. Lyons, D. den 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*, 11(1):1–8, January 2021.

[304] 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)*, pages 1944–1953, June 2019.

[305] Rajkumar L. Biradar and Vinayadatt V. Kohir. A novel image inpainting technique based on median diffusion. *Sadhana*, 38(4):621–644, August 2013.

[306] 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)*, pages 658–661, August 2012.

[307] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[308] V. Fernández 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)*, pages 4978–4988, June 2020.

List of Figures

List of Tables

From algorithm to hardware co-design for multimodal information retrieval of TCSPC data

Algorithme et co-conception matérielle pour l'analyse multimodale de données TCSPC

Résumé

Depuis l'invention des premiers capteurs de comptage de photons, plusieurs avancées technologiques ont été réalisées dans le domaine des systèmes d'imagerie à comptage de photons notamment avec l'invention des capteurs à avalanche à photon unique (SPAD pour single photon avalanche diode). Bien que le SPAD soit devenu une technologie de pointe pour beaucoup d'utilisation, dans le cadre applicatif spécifique de l'imagerie 3D, l'attractivité du capteur SPAD à temps de vol direct (D-ToF pour direct time-of-flight) est moindre de part sa faible résolution spatiale et son niveau de bruit élevé dans des conditions de fonctionnement extérieures. Un des leviers pour pallier ces limitations serait de résoudre le problème lié au format des données brut du capteur SPAD qui sont sous la forme de comptage de photon corrélé dans le temps (TCSPC pour Time Correlated Single Photon Counting). Deux grands axes de recherches ont été approfondis dans ce sens. Un premier axe porte sur le développement de méthodes d'acquisitions et de reconstructions alternatives basées sur des approches Bayésiennes, dans le but de n'extraire que les informations latentes du modèle statistique des données TCSPC. Tandis que le second axe porte sur le domaine de l'acquisition comprimée (CS pour compressive sensing) combinée à des méthodes de régression personnalisée ou à des algorithmes d'apprentissage profond pour la reconstruction.

Mots-clés : Capteur 3D, Comptage de photons uniques corrélés dans le temps, Photodiodes à avalanche à photon unique, Analyse statistique, Acquisition comprimée, Apprentissage profond

Abstract

Since the invention of the first photon counting sensors, several technological breakthroughs have been made in the field of photon counting imaging systems, notably with the invention of single photon avalanche diode (SPAD). Even though SPAD sensors became a key technology, the direct time-of-flight (D-ToF) SPAD sensor is not really interesting for the 3D imaging application due to its low spatial resolution and high noise level in outdoor operating conditions. These limitations are mainly based on the format of the raw SPAD data being in the form of time correlated single photon counts (TCSPC). Therefore, two research axes have been pursued in this thesis in order to rethink the format of the raw SPAD data. The first axis consists in developing alternative acquisition and reconstruction methods based on variational Bayesian approaches, with the aim of extracting only latent information from the statistical model of TCSPC data. While the second axis focuses on the compressive sensing (CS) domain in combination with custom regression methods or deep learning algorithms for reconstruction.

Keywords : 3D imaging, Time correlated single photon counting, Single photon avalanche diode, Statistical signal processing, Compressive sensing, Deep Learning