TCSPC histogram data modeling

Valentin Poisson, William Guicquero, Gilles Sicard

TCSPC histogram data modeling

A custom EM algorithm dedicated to a mixture of truncated-shifted Erlangs

Valentin Poisson, William Guicquero, and Gilles Sicard.

***Abstract*—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. State-of-the-art (SoA) works relying on Bayesian approaches [1], [2] or Deep Learning (DL) [3], [4] usually study the restrictive case of low-photon counts mode of operation. On the contrary, this paper aims to consider the high photon counts, synchronous operating mode, where the SPAD "Dead Time" is spread over a large number of bins of the Time-of-Flight (ToF) histogram (*i.e.*, 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 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-GEM) 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-luminance reconstructions.**

I. Introduction

A SPAD-TCSPC system consists in measuring single photon arrival times with Time-to-Digital Converters (TDC). From the system point of view, a transmitter (laser) emits short light pulses, which, once reflected on an object, can be detected by the associated SPAD receiver. This allows the round trip time of the laser pulse to be measured. From there, physical quantities information, such as depth and luminance, can be inferred by analyzing the statistical distributions of the photon arrival times. 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.

For a deep system-level analysis, the TCSPC SPAD-based imaging system is considered as a queueing loss model [5]. Where photon arrivals are the customers, generated from the Poisson process property of photons emitted by a light source [6]; 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. 1). 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. 1. 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. 3)–.

Valentin Poisson, William Guicquero, and Gilles Sicard are with the CEA-LETI, Univ. Grenoble Alpes, 38000 France (mail: firstname.name@cea.fr).

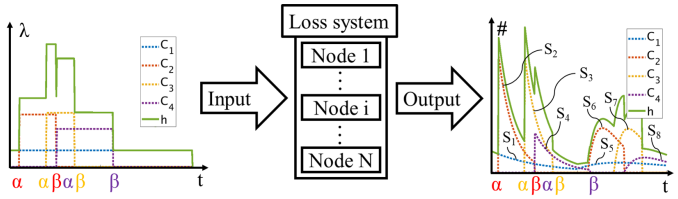


Fig. 1: Loss system overview.

This paper aims at properly modeling the unknown "stored photons" process, in order to develop a dedicated algorithm to estimate latent physical quantities. Based on the Poisson process stationary and independent increments property [7] and the superposition property of independent Poisson (sub)processes [8], [9], 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. 1), 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 [10] (output of the loss system illustrated in Fig. 1). 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 [11], as written in Eq. 1.

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

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. 1 to the Eq. 2 and implying a non-uniform time distribution of events if summing the Erlang-k phases from $k = 0$ to $k \to \infty$.

$$\begin{aligned}\mathbb{P}(Z_k = z_k) &= \mathbb{P}\left[\sum_{j=1}^{k} ((k-1)\tau + X_j)\right] \\ &= \int_0^{z_k-(k-1)\tau}\int_\tau^{x_1+\tau}\dots\int_{(k-1)\tau}^{x_{k-1}+\tau} \lambda\,\mathrm{e}^{-\lambda(z_k-x_1-(k-1)\tau)} \\ &\quad \dots\ \lambda\,\mathrm{e}^{-\lambda(x_{k-1}-x_k+\tau)}\ \lambda\,\mathrm{e}^{-\lambda(x_k-(k-1)\tau)}\,dx_k\ \dots\ dx_1 \\ &= \frac{\lambda^k (z_k-\alpha)^{k-1}\,\mathrm{e}^{-\lambda(z_k-\alpha)}}{(k-1)\,!},\ \text{where } \alpha = (k-1)\tau\end{aligned} \tag{2}$$

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 ($\mathscr{E}$).

Once the "stored photons" process distribution is properly defined, the rest of this paper focuses on a custom optimization method for the estimation of $\mathscr{E}$ parameters. The core algorithmic inner loop implements a set of tools such as Maximum Likelihood Estimation [12], Moments Matching [13], Parametric Imputation [14] and Variable Neighborhood Search [15]. While a Genetic search [16] based on a statistical test, drives an outer loop aiming at inferring the number of model components.

II. 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

Algorithm 1: 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 II-A)**
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 II-B)**
for $s \leftarrow 1$ **to** S **do**
  $\hat{\lambda}_s \leftarrow \frac{k_s \hat{\pi}_s}{\hat{\mu}_s^{(obs)} + \hat{\mu}_s^{(cens)}},\ \hat{\pi}_s \leftarrow \frac{\sum_{b=1}^{\hat{\beta}_s} \hat{\gamma}_{b,s}}{\mathscr{E}_c(\hat{\boldsymbol{\theta}}_s)}$
  $\hat{\alpha}_s \leftarrow \begin{cases} \hat{\alpha}_s - \epsilon_s, \text{ if } D_{KL}(\hat{\mathbf{h}}\|\mathbf{h}) \geq D_{KL}(\hat{\mathbf{h}}_{\epsilon_s}\|\mathbf{h}) \\ \hat{\mu}_s - \sqrt{k_s \hat{v}_s}, \text{ otherwise.} \end{cases}$
  $\hat{\beta}_s \leftarrow \hat{\alpha}_s - \frac{W_{-1}\left(\left(\frac{1}{e}\right)^{(k_s-1)}(\rho_s-1)\right)+(k_s-1)}{\hat{\lambda}_s}$
end

PCPP is composed of C Poisson intensity levels being the superposition of C independent Poisson arrival sub-processes [17], 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 $m\mathscr{E}$:

$$m\mathscr{E}(z;\boldsymbol{\Theta}) = \sum_{s=1}^{S} \pi_s \frac{\lambda_s^{k_s}(z-\alpha_s)^{k_s-1}\,\mathrm{e}^{-\lambda_s(z-\alpha_s)}}{(k_s-1)\,!}\mathbf{1}_{z\in[\alpha_s,\beta_s]}, \tag{3}$$

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.

A. 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. 3 is thus revised here:

$$\mathscr{E}(b;\boldsymbol{\theta}_s) = \pi_s \frac{\lambda_s^{k_s}(b-\alpha_s)^{k_s-1}\,\mathrm{e}^{-\lambda_s(b-\alpha_s)}}{(k_s-1)\,!}\mathbf{1}_{b\in[\alpha_s,\beta_s]}. \tag{4}$$

Due to its better tractability, the "log-likelihood" optimization trick is preferred (cf. Eq. 5).

$$\begin{aligned} l(\boldsymbol{\Theta}|\mathbf{h}) &= \sum_{b=1}^{B}\sum_{i=1}^{h_b} \ln\left(\sum_{s=1}^{S} \delta_{b,i,s}\pi_s \mathscr{E}(b;\boldsymbol{\theta}_s)\right) \\ &= \sum_{b=1}^{B}\sum_{s=1}^{S} \gamma_{b,s} \ln\left(\pi_s \mathscr{E}(b;\boldsymbol{\theta}_s)\right) \end{aligned} \tag{5}$$

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).

B. Combination of optimization methods (M-step)

Note that, [18], [19] 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 paper presents parameter estimation variants in order to overcome this issue.

1) 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. 6.

$$\hat{\pi}_s = \frac{\sum_{b=1}^{B} \hat{\gamma}_{b,s}}{\sum_{s=1}^{S}\sum_{b=1}^{B} \hat{\gamma}_{b,s}} \tag{6}$$

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. 7 which normalizes the estimated values only over the interval of distribution support, defined as $[\hat{\alpha}_s, \hat{\beta}_s]$.

$$\hat{\pi}_s = \frac{\sum_{b=1}^{B} \hat{\gamma}_{b,s}}{\mathscr{E}_c\left(\hat{\boldsymbol{\theta}}_s\right)} = \frac{\sum_{b=1}^{B} \hat{\gamma}_{b,s}}{1 - \sum_{j=0}^{k_s-1} \frac{\mathrm{e}^{-\hat{\lambda}_s(\hat{\beta}_s-\hat{\alpha}_s)}\left(\hat{\lambda}_s(\hat{\beta}_s-\hat{\alpha}_s)\right)^j}{j\,!}} \tag{7}$$

2) Erlang intensities, $\hat{\lambda}_s$: The standard MLE of the Erlang intensities would be:

$$\hat{\lambda}_s = \frac{k_s\hat{\pi}_s}{\hat{\mu}_s - \hat{\alpha}_s} = \frac{k_s\hat{\pi}_s}{\hat{\mu}_s^{(obs)} + \hat{\mu}_s^{(cens)}} = \frac{k_s\hat{\pi}_s}{\sum_{b=\hat{\alpha}_s+1}^{\infty} \hat{\gamma}_{b,s}(b-\hat{\alpha}_s)}. \tag{8}$$

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. 1-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. 9).

$$\begin{aligned} \hat{\mu}_s^{(cens)} &= \hat{\pi}_s \int_{\hat{\beta}_s}^{\infty} b\mathscr{E}(b;\hat{\boldsymbol{\theta}}_s)db \\ &= \hat{\pi}_s \int_{\hat{\beta}_s}^{\infty} b\frac{\hat{\lambda}_s^{k_s}(b-\hat{\alpha}_s)^{k_s-1}\,\mathrm{e}^{-\hat{\lambda}_s(b-\hat{\alpha}_s)}}{(k_s-1)\,!}db \\ &= \hat{\pi}_s \frac{k_s}{\hat{\lambda}_s} \int_{\hat{\beta}_s}^{\infty} \frac{\hat{\lambda}_s^{k_s+1}(b-\hat{\alpha}_s)^{k_s}\,\mathrm{e}^{-\hat{\lambda}_s(b-\hat{\alpha}_s)}}{k_s\,!}db \\ &= \hat{\pi}_s \frac{k_s}{\hat{\lambda}_s} \sum_{j=0}^{k_s} \frac{\mathrm{e}^{-\hat{\lambda}_s(\hat{\beta}_s-\hat{\alpha}_s)}\left(\hat{\lambda}_s(\hat{\beta}_s-\hat{\alpha}_s)\right)^j}{j\,!} \end{aligned} \tag{9}$$

3) Erlang shift, $\hat{\alpha}_s$: A Moment Method (MM) [20] combined with a custom Variable Neighborhood Search (VNS) [21] 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 $\hat{\alpha}_s$,

$$\hat{\alpha}_s = \hat{\mu}_s - \sqrt{k_s\left(\hat{v}_s^{(obs)} + \hat{v}_s^{(cens)}\right)}. \tag{10}$$

The empirical estimation of the variance for the observed data is then obtained using the König-Huygens theorem [7]:

$$\hat{v}_s^{(obs)} = \sum_{b=\hat{\alpha}_s+1}^{\hat{\beta}_s} \hat{\gamma}_{b,s}(b-\hat{\alpha}_s)^2 - \left(\hat{\mu}_s^{(obs)}\right)^2. \tag{11}$$

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

$$\hat{v}_s^{(cens)} = \eta - \left(\hat{\mu}_s^{(cens)}\right)^2, \tag{12}$$

where η is computed using Eq. 13:

$$\begin{aligned} \eta &= \hat{\pi}_s \int_{\hat{\beta}_s}^{\infty} b^2\mathscr{E}(b;\hat{\boldsymbol{\theta}}_s)db \\ &= \hat{\pi}_s \int_{\hat{\beta}_s}^{\infty} b^2\frac{\hat{\lambda}_s^{k_s}(b-\hat{\alpha}_s)^{k_s-1}\,\mathrm{e}^{-\hat{\lambda}_s(b-\hat{\alpha}_s)}}{(k_s-1)\,!}db \\ &= \hat{\pi}_s \frac{k_s(k_s+1)}{\hat{\lambda}_s^2} \int_{\hat{\beta}_s}^{\infty} \frac{\hat{\lambda}_s^{k_s+2}(b-\hat{\alpha}_s)^{k_s+1}\,\mathrm{e}^{-\hat{\lambda}_s(b-\hat{\alpha}_s)}}{(k_s+1)\,!}db \\ &= \hat{\pi}_s \frac{(k_s+1)k_s}{\hat{\lambda}_s^2} \sum_{j=0}^{k_s+1} \frac{\mathrm{e}^{-\hat{\lambda}_s(\hat{\beta}_s-\hat{\alpha}_s)}\left(\hat{\lambda}_s(\hat{\beta}_s-\hat{\alpha}_s)\right)^j}{j\,!}. \end{aligned} \tag{13}$$

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. 14). The drawn value giving the lowest distance using the Kullback–Leibler divergence, $D_{KL}()$ [22] is then selected, cf., Alg. 1-line 6.

$$\epsilon_s \sim \text{Exp}\left(\left|\hat{\mu}_s^{(obs)} + \hat{\mu}_s^{(cens)} - \sqrt{k_s \hat{v}_s}\right|\right) \tag{14}$$

4) Erlang right-truncation, $\hat{\beta}_s$: Instead of using a likelihood function of the truncated data which is not tractable in our case [23], 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:

$$1 - \sum_{j=0}^{k_s-1} \frac{\mathrm{e}^{-\hat{\lambda}_s(\hat{\beta}_s - \hat{\alpha}_s)} \left(\hat{\lambda}_s(\hat{\beta}_s - \hat{\alpha}_s)\right)^j}{j\,!} = \frac{\sum_{b=1}^{B} \hat{\gamma}_{b,s}}{\hat{\pi}_s}. \tag{15}$$

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:

$$\rho_s = \frac{\sum_{b=1}^{B} \hat{\gamma}_{b,s} \left(1 + (\hat{h}_b - h_b) \frac{1}{\sqrt{(B-\hat{\beta}_s)}} \mathrm{e}^{-\frac{1}{\sqrt{(B-\hat{\beta}_s)}}|b-\hat{\beta}_s|}\right)}{\hat{\pi}_s}. \tag{16}$$

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

$$1 - \sum_{j=0}^{k_s-1} \frac{\mathrm{e}^{-\hat{\lambda}_s(\hat{\beta}_s - \hat{\alpha}_s)} \left(\hat{\lambda}_s(\hat{\beta}_s - \hat{\alpha}_s)\right)^j}{j\,!} = \rho_s \tag{17}$$

Solving Eq. 17 for $\hat{\beta}_s$ thus leads to the Eq. of Alg. 1-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 [24].

III. 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 knowledge 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. 2.

Algorithm 2: Genetic based c-EM (c-GEM) algorithm to fit a mixture of truncated Shifted Erlang distribution

Data: $\mathbf{h} \in \mathbb{N}^B$
Input: $P, C, \epsilon_d, \epsilon_{KL}, \chi^2, \hat{\chi}^2$
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 $\Delta\boldsymbol{\Theta}^{(1)} > \epsilon_d, \quad D_{KL}^{(1)} > \epsilon_{KL}$ **do**
    for $p \leftarrow 1$ **to** P **do**
      // c-EM (Sec. II):
      $\hat{\boldsymbol{\alpha}}^{(p)}, \hat{\boldsymbol{\beta}}^{(p)}, \hat{\boldsymbol{\pi}}^{(p)}, \hat{\boldsymbol{\lambda}}^{(p)} \leftarrow$ Alg. 1
      // Regularization (Sec. III-A):
      $\hat{\boldsymbol{\lambda}}_c^{(p)} \leftarrow \zeta \hat{\boldsymbol{\lambda}}_c^{(p)} (\boldsymbol{\Delta}^\top \boldsymbol{\Delta} - \zeta \mathbf{I}_d)^{-1}$
      $\hat{\boldsymbol{\alpha}}_c^{(p)} \leftarrow (\boldsymbol{\Delta}^\top \boldsymbol{\Delta} - \zeta \mathbf{I}_d)^{-1} (\boldsymbol{\Delta}^\top \mathbf{D_T} + \zeta \hat{\boldsymbol{\alpha}}_c^{(p)})$
    end
    // Genetic search (Sec. III-B)
    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. III-B)
  $\hat{\chi}^2 = \sum_{b=1}^{B} \frac{(\hat{h}_b^{(1)} - h_b)^2}{h_b}$
end

A. 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. 18). These equality constraints are relaxed by using $l2$ regularization [25] 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. 18, 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 $\hat{\boldsymbol{\alpha}}_c^\top$.

$$\hat{\boldsymbol{\lambda}}_c^\top = \arg\min_{\hat{\boldsymbol{\lambda}}_c} \left(\zeta \|\hat{\boldsymbol{\lambda}}_c^\top - \hat{\boldsymbol{\lambda}}_c^{\top(prev)}\|_2^2 + \|\boldsymbol{\Delta}\hat{\boldsymbol{\lambda}}_c^\top\|_2^2\right)$$
$$\hat{\boldsymbol{\alpha}}_c^\top = \arg\min_{\hat{\boldsymbol{\alpha}}_c} \left(\zeta \|\hat{\boldsymbol{\alpha}}_c^\top - \hat{\boldsymbol{\alpha}}_c^{\top(prev)}\|_2^2 + \|\boldsymbol{\Delta}\hat{\boldsymbol{\alpha}}_c^\top - \mathbf{D_T}\|_2^2\right) \quad (18)$$

B. 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– of the parameter vectors values below $\epsilon_d = 10^{-7}$, and the relative difference of the Kullback-Leibler [22] below $\epsilon_{KL} = 10^{-7}$ as written in Alg. 2). 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 [26] (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. 2) 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 [27] (cf., Alg. 2). The mutation consists in a random draw of $\hat{\alpha}_s$, $\forall s \in S$ from the positive residual support function (*i.e.*, $\hat{\alpha}_s \overset{iid}{\sim} Supp(max\{\hat{\mathbf{h}} - \mathbf{h}, 0\})$).

IV. Experimental results

A. 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. 2 that even include groups with overlapping distribution supports.

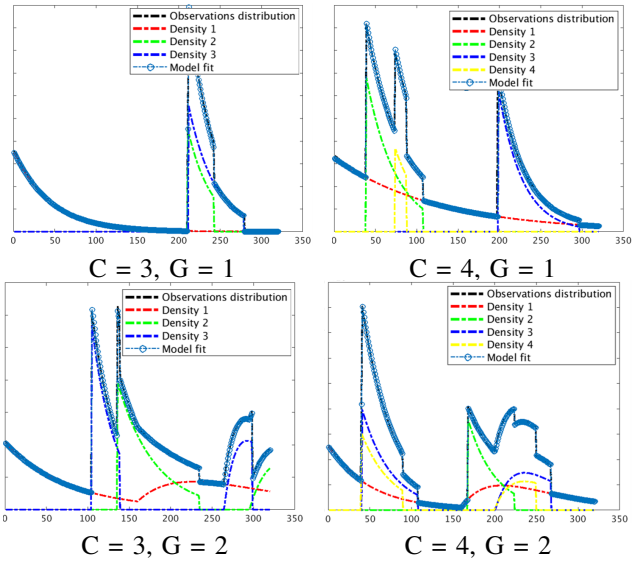


Fig. 2: c-EM estimations on synthetic histogram data.

	$\mathbf{h}$ (%)	$\boldsymbol{\lambda}$ (%)	$\boldsymbol{\alpha}$ (%)	$\boldsymbol{\pi}$ (%)	$\boldsymbol{\beta}$ (%)
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

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

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

B. Real traffic traces, simulated SPAD data

A SPAD TCSPC simulation platform has been derived from the model described in [28], [29] to provide "pseudo-realistic" pixel-wise histogram data from depth-luminance scenes of the Middlebury dataset [30] as illustrated in Fig. 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. 5 and 6 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 [31]. In addition, 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. 5 and 6).

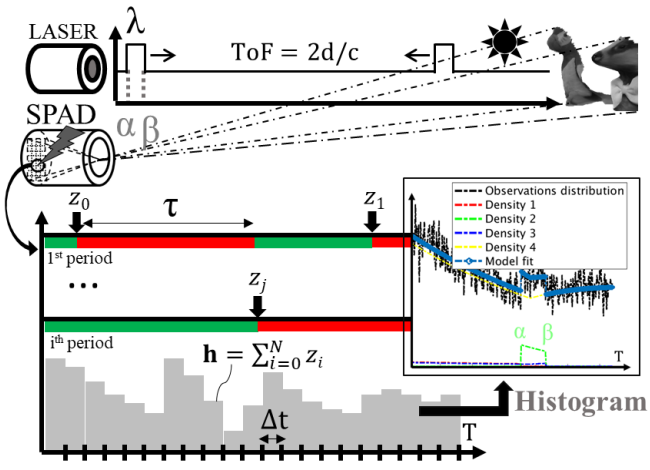
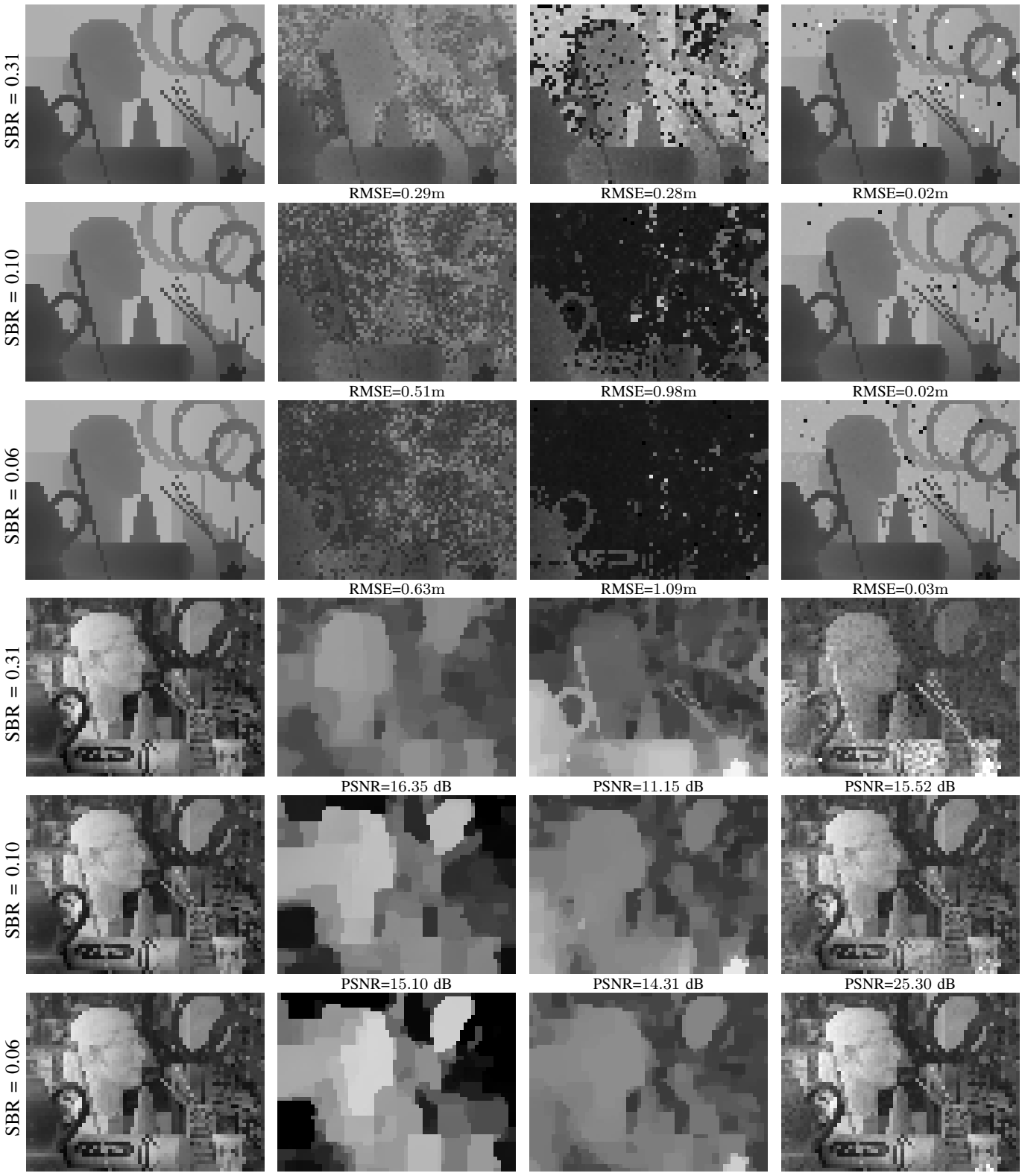


Fig. 3: SPAD system overview.



(a) Ground Truth (b) Shin [1] (c) Rapp [2] (e) Our

Fig. 4: 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.*, $5 \cdot 10^3$ photons ToF samples size).

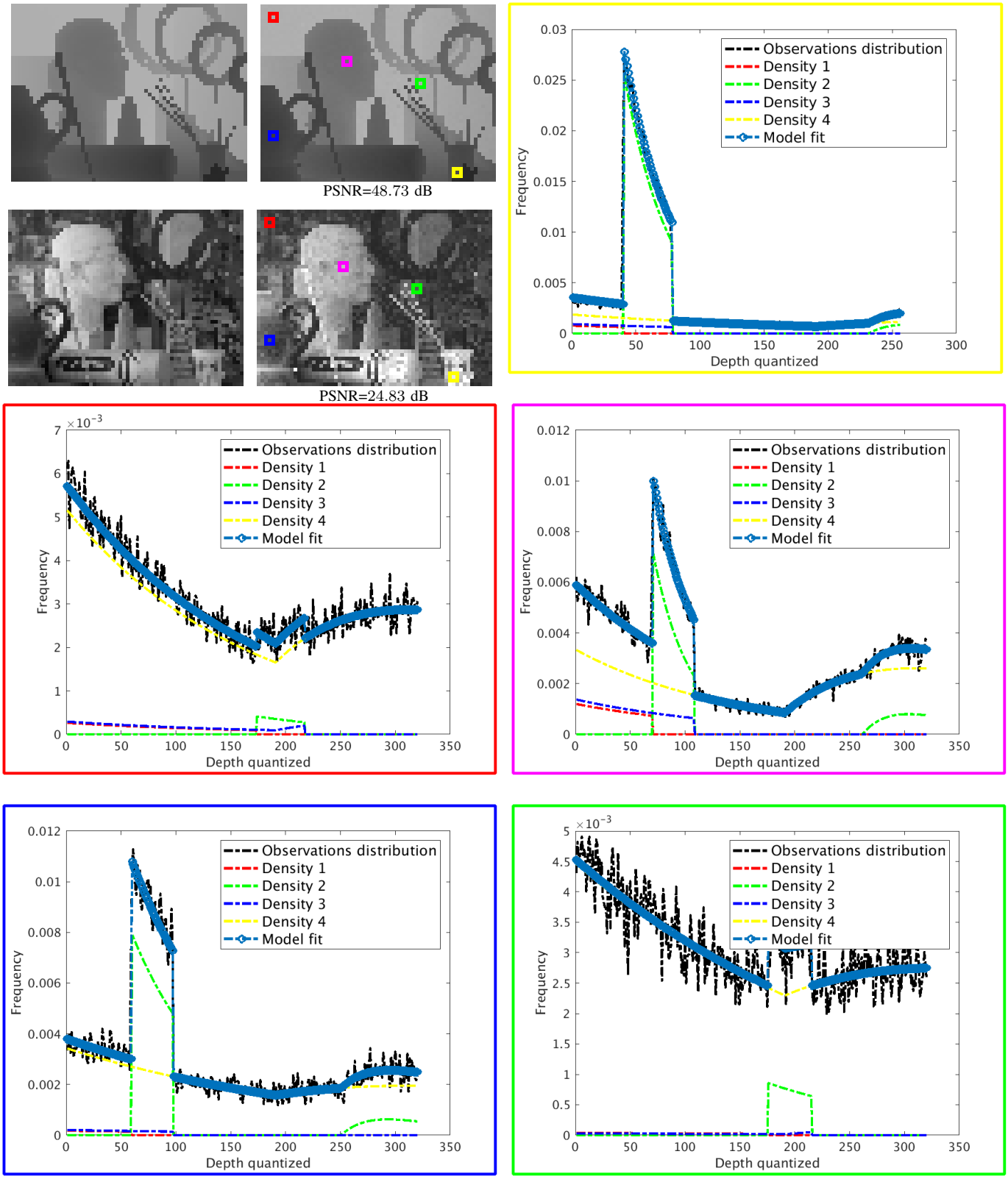


Fig. 5: Depth-Luminance estimation from $2 \cdot 10^4$ photons ToF samples size. Note that third and forth image rows correspond to the pixels histograms surrounded in color in the upper image reconstructions.

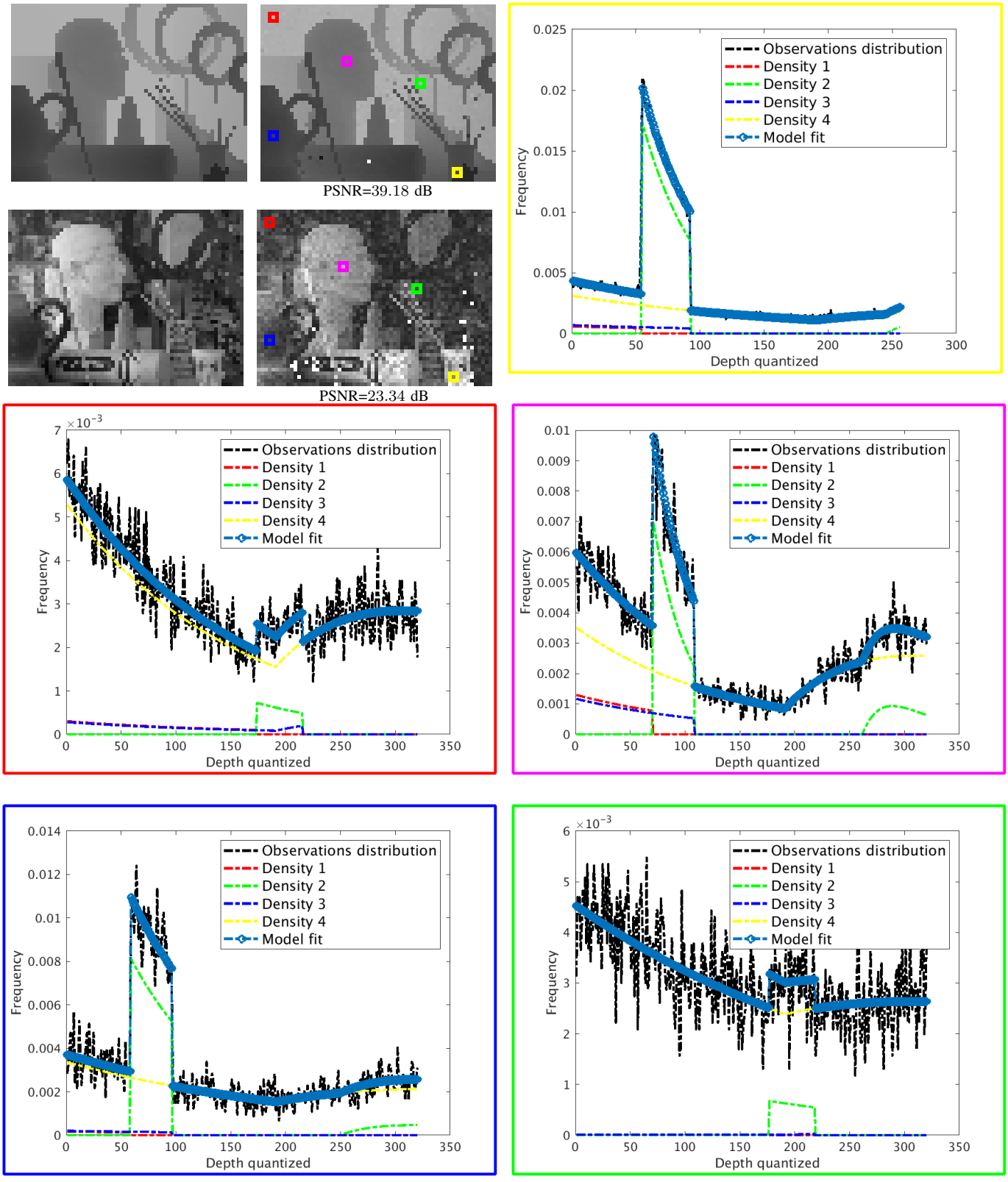


Fig. 6: Depth-Luminance estimation from $5 \cdot 10^3$ photons ToF samples size. Note that third and forth image rows correspond to the pixels histograms surrounded in color in the upper image reconstructions.

The joint reconstructions using our proposed algorithm are shown in Fig. 4. Depth-luminance reconstruction accuracy for the considered scene has been respectively estimated to $0.03\mathrm{m}$ root mean square error (RMSE) and 24dB peak signal to noise ratio (PSNR). Tab. II reports a comparison with respect to prior works of [1] and [2] that respectively exhibit a $0.9\mathrm{m}$ and $1.28\mathrm{m}$ RMSE (depth) and a 16.2dB PSNR and 13.2dB PSNR (luminance) under 0.10 SBR. [1], [2] 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 [1], [2] 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). 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 error in the estimation of the mean as observed in Fig. 4, for example for the estimation of the teapot depth. Moreover, [1], [2] reconstruction algorithms based on the uniform-Gaussian 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 [32], [33] 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 [1].

V. Conclusion

This paper 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 [12], Moments Matching [13], Parametric Imputation [14] and Variable Neighborhood Search [15]. An outer loop additionally aims at inferring the number of model components thanks to a Genetic search approach [16] based on iterative statistical tests.

SBR	Intensity (PSNR in dB)			Depth (RMSE in m)		
	Shin [1]	Rapp [2]	Our	Shin [1]	Rapp [2]	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

TABLE II: Quantitative comparisons of average depth-luminance reconstructions of the Middlebury dataset [30] under various SBR and from the optimal photon regime of each methods (*i.e.*, low photon counts for [1], [2] and high photon counts for our custom c-GEM algorithm.

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 problem statement 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 [34], patient arriving to the hospital [35], portfolio credit risk [36].

References

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

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

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

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

[5] M. F. Neuts, "Renewal processes of phase type," *Naval Research Logistics Quarterly*, vol. 25, no. 3, pp. 445–454, Sep. 1978.

[6] A. C. Sparavigna, "Poissonian distributions in physics: Counting electrons and photons," 2021.

[7] O. C. Ibe, "1 - Basic Concepts in Probability," in *Markov Processes for Stochastic Modeling (Second Edition)*, second edition ed., O. C. Ibe, Ed. Oxford: Elsevier, 2013, pp. 1–27.

[8] E. Cinlar and R. Agnew, "On the superposition of point processes," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 30, no. 3, pp. 576–581, 1968.

[9] H. Crane and P. Mccullagh, "Poisson superposition processes," *Journal of Applied Probability*, vol. 52, no. 4, pp. 1013–1027, 2015.

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

[11] O. C, "Fundamentals of applied probability and random processes," *Academic Press*, 2014.

[12] I. J. Myung, "Tutorial on maximum likelihood estimation," *Journal of mathematical Psychology*, vol. 47, no. 1, pp. 90–100, 2003.

[13] M. A. Johnson and M. R. Taaffe, "Matching moments to phase distributions: Mixtures of erlang distributions of common order," *Stochastic Models*, vol. 5, no. 4, pp. 711–743, 1989.

[14] H. Okamura, T. Dohi, and K. S. Trivedi, "Improvement of expectation–maximization algorithm for phase-type distributions with grouped and truncated data," *Applied Stochastic Models in Business and Industry*, vol. 29, no. 2, pp. 141–156, 2013.

[15] P. Hansen and N. Mladenović, "An introduction to variable neighborhood search," in *Meta-heuristics*. Springer, 1999, pp. 433–458.

[16] M. Srinivas and L. M. Patnaik, "Genetic algorithms: A survey," *computer*, vol. 27, no. 6, pp. 17–26, 1994.

[17] M. F. Neuts, "Renewal processes of phase type," *Naval Research Logistics Quarterly*, vol. 25, no. 3, pp. 445–454, 1978.

[18] S. Asmussen, O. Nerman, and M. Olsson, "Fitting phase-type distributions via the em algorithm," *Scandinavian Journal of Statistics*, pp. 419–441, 1996.

[19] 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*, vol. 3, no. 3, pp. 245–258, 2006.

[20] X. Peng, Q. Bai, X. Xia, Z. Huang, K. Saenko, and B. Wang, "Moment Matching for Multi-Source Domain Adaptation," 2019, pp. 1406–1415.

[21] P. Hansen and N. Mladenović, "An Introduction to Variable Neighborhood Search," in *Meta-Heuristics*. Springer, Boston, MA, 1999, pp. 433–458, dOI: 10.1007/978-1-4615-5775-3_30.

[22] F. Bavaud, "Relative Entropy and Statistics," *arXiv:0808.4111 [cs, math, stat]*, Aug. 2008, arXiv: 0808.4111.

[23] R. R. Andridge and R. J. Little, "A review of hot deck imputation for survey non-response," *International statistical review*, vol. 78, no. 1, pp. 40–64, 2010.

[24] 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)*, 2020, pp. 36–39.

[25] Y. Tian and Y. Zhang, "A comprehensive survey on regularization strategies in machine learning," *Inf. Fusion*, vol. 80, no. C, p. 146–166, apr 2022.

[26] M. L. McHugh, "The Chi-square test of independence," *Biochemia Medica*, vol. 23, no. 2, pp. 143–149, Jun. 2013.

[27] S. Huda, J. Yearwood, and R. Togneri, "Hybrid Metaheuristic Approaches to the Expectation Maximization for Estimation of the Hidden Markov Model for Signal Modeling," *IEEE Transactions on Cybernetics*, vol. 44, no. 10, pp. 1962–1977, Oct. 2014.

[28] P. Padmanabhan, C. Zhang, and E. Charbon, "Modeling and Analysis of a Direct Time-of-Flight Sensor Architecture for LiDAR Applications," *Sensors*, vol. 19, no. 24, p. 5464, Jan. 2019.

[29] V. Poisson, W. Guicquero, G. Sicard *et al.*, "Luminance-depth reconstruction from compressed time-of-flight histograms," *IEEE Transactions on Computational Imaging*, vol. 8, pp. 148–161, 2022.

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

[31] P. Gatt, S. Johnson, and T. Nichols, "Geiger-mode avalanche photodiode ladar receiver performance characteristics and detection statistics," *Applied optics*, vol. 48, no. 17, pp. 3261–3276, 2009.

[32] R. Verbelen, L. Gong, K. Antonio, A. Badescu, and S. Lin, "Fitting mixtures of erlangs to censored and truncated data using the em algorithm," *ASTIN Bulletin: The Journal of the IAA*, vol. 45, no. 3, pp. 729–758, 2015.

[33] T. C. Fung, A. L. Badescu, and X. S. Lin, "Fitting censored and truncated regression data using the mixture of experts models," *North American Actuarial Journal*, pp. 1–25, 2021.

[34] L. M. LEEMIS, "TECHNICAL NOTE: Nonparametric estimation and variate generation for a nonhomogeneous Poisson process from event count data," *IIE Transactions*, vol. 36, no. 12, pp. 1155–1160, Dec. 2004.

[35] M. Thomason and J. Whittaker, "Rare failure-state in a Markov chain model for software reliability," in *Proceedings 10th International Symposium on Software Reliability Engineering (Cat. No.PR00443)*, Nov. 1999, pp. 12–19, iSSN: 1071-9458.

[36] A. Herbertsson, "Modelling default contagion using multivariate phase-type distributions," *Review of Derivatives Research*, vol. 14, no. 1, pp. 1–36, Apr. 2011.