Expectation-Maximization based approach to 3D reconstruction from single-waveform multispectral Lidar data

Quentin Legros, Sylvain Meignen, Stephen McLaughlin, Yoann Altmann

***Abstract*—In this paper, we present a novel Bayesian approach for estimating spectral and range profiles from single-photon Lidar waveforms associated with single surfaces in the photon-limited regime. In contrast to classical multispectral Lidar signals, we consider a single Lidar waveform per pixel, whereby a single detector is used to acquire information simultaneously at multiple wavelengths. A new observation model based on a mixture of distributions is developed. It relates the unknown parameters of interest to the observed waveforms containing information from multiple wavelengths. Adopting a Bayesian approach, several prior models are investigated and a stochastic Expectation-Maximization algorithm is proposed to estimate the spectral and depth profiles. The reconstruction performance and computational complexity of our approach are assessed, for different prior models, through a series of experiments using synthetic and real data under different observation scenarios. The results obtained demonstrate a significant speed-up (up to 100 times faster for four bands) without significant degradation of the reconstruction performance when compared to existing methods in the photon-starved regime.**

***Index Terms*—Multispectral imaging, 3D imaging, single-photon Lidar, Bayesian estimation, Expectation-Maximization.**

I. Introduction

Depth measurement from light detection and ranging (Lidar) systems has received an increasing interest over the last decades, as it allows reconstruction of 3D scenes at high resolution [1], [2]. In particular, single-photon Lidar has enabled unprecedented 3D reconstruction results, at longer ranges [3]–[5], through obscurants [6] and underwater [7]–[9], using extremely low illumination powers. This performance improvement thus makes single-photon Lidar particularly attractive for automotive, robotics and defense applications. Single band Lidar (SBL), i.e., using a single illumination wavelength, allows the extraction of spatial structures from 3D scenes, using a pulsed illumination and analysis of the time-of-arrival (ToA) of detected photons, for each spatial location or pixel. More precisely, time correlated single-photon counting (TCSPC) correlates the photon ToAs with the time of emission of the pulses to obtain the time-of-flights (ToFs) of the recorded photons that are gathered to form a histogram of photon detection events. Neglecting light scattering in the medium, the detected photons originally emitted by the pulsed source are temporally clustered and form a peak in the histogram, whose amplitude informs on the reflectivity of the surface. Conversely, the additional detection events caused by ambient illumination and detector dark counts are classically uniformly distributed across the histogram bins. While allowing the reconstruction of high-resolution 3D scenes, SBL only provides one reflectivity value (corresponding to the wavelength of the laser source) per surface and thus does not provide spectral information about the scene of interest. When such information is also required, data fusion strategies can be adopted, as in [10], [11] for instance.

Multispectral Lidar (MSL) overcomes data registration and fusion problems, as spectral and spatial features are collected using a single imaging system [12]. This emerging modality has already led to promising results for underwater imaging [13], object detection in urban area [14] and forest canopy monitoring [15]. Using laser sources at different wavelengths, histograms associated with different spectral bands are usually recorded for each pixel of the scenes, using a single device. The acquisition of MSL signals can be performed either sequentially, i.e., one wavelength at a time, or in parallel. Sequential acquisition requires an overall longer acquisition time which depends on the number of wavelengths sensed. While parallel acquisition allows faster acquisition, it requires a more complex and potentially more expensive imaging system [16]–[18], e.g., using fibers optic [19] or optical volume gratings [20] to separate the spectral components.

To reduce the acquisition time, limit the complexity of the imaging system and the volume of data to be stored and processed, a single-waveform MSL (SW-MSL) approach has recently been developed [21], whereby a single histogram containing spectral information from multiple wavelengths is recorded per pixel via wavelength-time coding (see Fig 1). Using a single waveform per pixel, the acquisition time is of the same order as when using SBL in the photon-starved regime. As discussed in [21], the distribution of the photons ToA is shifted in time with a wavelength-dependent temporal delay. Although these delays are fixed, the amplitude of the different peaks varies depending on the spectral signatures of the scene, which changes the waveform shape.

The study of SW-MSL data reduces to estimating a 3D profile and to quantifying the proportion of detected photons associated with each wavelength. In a similar fashion to the analysis of MSL data, the joint estimation of the spectral

This work was supported by the Royal Academy of Engineering under the Research Fellowship scheme RF201617/16/31 and by the Engineering and Physical Sciences Research Council (EPSRC) (grant EP/S000631/1) and the MOD University Defence Research Collaboration (UDRC) in Signal Processing. and the IDEX of University Grenoble Alpes.

Quentin Legros, Yoann Altmann and Stephen McLaughlin are with the School of Engineering anf Physical Sciences, Heriot Watt University, Edinburgh, United Kingdom). Sylvain Meignen is with the Laboratoire Jean Kuntzmann, Grenoble, France.

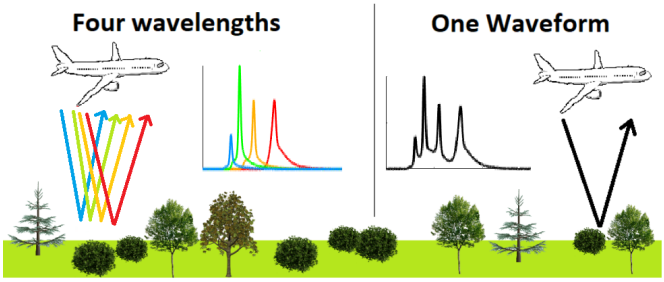

Fig. 1. Examples of acquisition of four MSL waveforms (left), using one wavelength per waveform, and SW-MSL (right) capturing simultaneously information from four wavelengths. In this example, the vegetated scene is mostly green and the Lidar return associated with this wavelength range (second peak from the left) is larger than the other returns.

and range profiles is a challenging problem, mainly due to the multimodal nature of the observation model. To overcome this difficulty, Bayesian approaches have been proposed in the MSL case, e.g., [22]–[25] for their ability to explore multimodal distributions and to quantify the uncertainty associated with the estimates (e.g., range profiles).

To analyse SW-MSL data, the Bayesian method proposed in [21] relies also on Monte Carlo sampling to estimate the spectral and depth profiles. Although it provides promising reconstruction results, this method suffers from a prohibitive computational time (about 15 hours to reconstruct a single 3D scene of 200×200 pixels and 4 wavelengths), which motivates the development of new and more scalable methods based on optimization schemes, as intended here.

In this work, we introduce a new SW-MSL model within a Bayesian framework and an associated inference procedure which is significantly faster than that proposed by Ren et al. [21]. The classical observation model based on Poisson noise as in [23], [26]–[29], is replaced by an equivalent model as in [30] (in the low-flux regime) which is more suited for the proposed inference strategy. More precisely, the distribution of the photon detection events is represented by a mixture model, where one component of the mixture relates to the ambient illumination (background) and where the remaining components are associated with the different illumination wavelengths (signal contributions). Similarly to most 3D reconstruction methods from single-photon data, such as [22], [23], [28], [30], we assume the photons emitted by the imaging system are reflected onto a single target. We then assign prior distributions to the unknown model parameters. To overcome convergence issues (multimodal posterior distribution, slow convergence of the MCMC method in [21]) induced by the joint estimation of the depth and spectral profile, we estimate the parameters sequentially. More precisely, we use an algorithm based on *Expectation-Maximization* (EM) to first estimate the mixture weights in each pixel. These weights are then used to estimate spectral signature in each pixels and the range profile.

The main contributions of the paper are:

- A new SW-MSL observation model whose formulation is well suited for inference using EM-based algorithms.
- An EM-based algorithm for analysis of SW-MSL data which extends the model proposed in [30] for SBL data.
- In contrast to the method proposed in [30], we provide an additional step to estimate reflectivity parameters from the estimated ratios of informative photons detected.
- A comparison of several prior models for the unknown spectral profile in terms of reconstruction performance and associated computational complexity. This study demonstrates that the proposed strategy yields faster reconstruction than the method in [21].

The remainder of this paper is organized as follows. Section II introduces the new SW-MSL observation model and the different prior distributions associated with the unknown model parameters are discussed in Section III. Section IV describes the estimation strategy based on EM. A comparison of the different prior models using synthetic SBL data is conducted in Section V in order to identify the most interesting options in terms of estimation performance and computational cost. This study is then extended in the remainder of Section V, which investigates real SW-MSL data analysis and comparison with the method proposed in [21]. The limitations of the method and future work are discussed in Section VI and conclusions are reported in Section VII.

II. Problem Formulation

Consider a 3D observation array $\boldsymbol{Z}$ of MSL waveforms $\boldsymbol{z}_{n,l} = [\boldsymbol{Z}]_{n,l,:} = [z_{n,l,1},\ldots,z_{n,l,T}]^\top$ of size $N\times L\times T$, where N is the number of pixels or spatial locations, L is the number of spectral bands and T is the length of the ToF histograms. The spectral response of the object in the nth pixel is denoted by $\boldsymbol{r}_n = [r_{n,1},\ldots,r_{n,L}]^\top \in \mathbb{R}_+^L$, and $b_{n,l}$ is the positive, constant over time, background level of that pixel, at the lth wavelength. The classical MSL observation model relies on Poisson noise [22], [26] such that

$$z_{n,l,t}|\left(r_{n,l},t_n,b_{n,l}\right) \sim \mathcal{P}\left(r_{n,l}g_l(t-t_n)+b_{n,l}\right), \qquad (1)$$

where $\{g_l(.)\}_l$ are the instrumental/impulse response functions (IRFs) associated with the L wavelengths. These IRFs are assumed to be known as they are generally estimated during the system calibration. Moreover, t_n is the characteristic ToF of photons emitted by the laser source and reflected by an object at distance d_n of the sensor. Note that d_n and t_n are linearly related when the light speed is constant in the medium between the imaging device and the scene.

In a similar manner, the SW-MSL observation model based on Poisson noise can be obtained by summing the elements of $\boldsymbol{Z}$ over the spectral dimension (see also [21]), that is, by defining $y_{n,t} = \sum_{l=1}^L z_{n,l,t}$, resulting in an $N\times T$ observation matrix $\boldsymbol{Y}$ with $[\boldsymbol{Y}]_{n,t} = y_{n,t}$. For convenience, we also define $\boldsymbol{y}_n = [y_{n,1},\ldots,y_{n,T}]^\top$, the waveform associated with the nth pixel, such that $\boldsymbol{Y} = [\boldsymbol{y}_1,\ldots,\boldsymbol{y}_N]^\top$. The resulting SW-MSL observation model becomes

$$y_{n,t}|\left(\boldsymbol{r}_n,t_n,b_n\right) \sim \mathcal{P}\left(b_n+\sum_{l=1}^L r_{n,l}g_l(t-t_n)\right), \qquad (2)$$

where b_n gathers all the background contributions in the nth pixel. The main challenge with the model in Eq. (2) is the joint estimation of $\boldsymbol{t} = \{t_n\}_n$ and $(\boldsymbol{R},\boldsymbol{b}) = (\{\boldsymbol{r}_n\}_n,\{b_n\}_n)$,

especially for high background levels that lead to a highly multimodal likelihood in Eq. (2).

Our approach uses a model reformulation similar to that in [30], as it allows the use of EM-based algorithms for faster parameter estimation. Instead of using the waveforms $\boldsymbol{y}_n$ as observations, it uses lists of photon ToFs. We denote by $\boldsymbol{s}_n = \{s_n^p\}_{p=1}^{\bar{y}_n}$ the set of photon ToFs in the nth pixel, with $\bar{y}_n = \sum_{t=1}^{T} y_{n,t}$. From Eq. (2), it can be shown that the observation model for a given s_n^p can be expressed as

$$p(s_n^p|\boldsymbol{w}_n, t_n) = \frac{1}{T}\left(1 - \sum_{l=1}^{L} w_{l,n}\right) + \sum_{l=1}^{L} w_{l,n}\frac{g_l(s_n^p - t_n)}{G_l}, \quad (3)$$

with $G_l = \sum_{t=1}^{T} g_l(t - t_n)$ the integral of the IRF g_l, which is assumed constant over the admissible values of t_n (such that the IRFs are not cropped). In (3), the weights $\boldsymbol{w}_n = [w_{1,n}, \ldots, w_{L,n}]^\top$ represent the probabilities of each detected photon to be a signal photons associated with a given wavelength and are given by

$$w_{l,n} = \frac{r_{n,l}G_l}{\sum_{l=1}^{L} r_{n,l}G_l + Tb_n}. \quad (4)$$

Note that $(1 - \sum_{l=1}^{L} w_{l,n})$ is the probability of a detected photon to be a background photon. The weights are gathered in the $L \times N$ matrix $\boldsymbol{W} = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_N]$, whose rows are denoted by $\boldsymbol{w}_{l,:}^\top$ (as column vectors). By construction, the vectors $\boldsymbol{w}_n$ belong to the set

$$S_L := \{\boldsymbol{x} \in \mathbb{R}^L | \forall l \in \{1, \ldots, L\}, x_l \geq 0, \sum_{l=1}^{L} x_l \leq 1\}, \quad (5)$$

and we denote by $S_L^N = \{\boldsymbol{W} | \forall n \in \{1, \ldots, N\}, \boldsymbol{w}_n \in S_L\}$ the domain of definition of $\boldsymbol{W}$.

Assuming that the ToAs are mutually independent, conditioned on the value of $(\boldsymbol{w}_n, t_n)$, the joint likelihood is

$$p(\boldsymbol{S}|\boldsymbol{W}, \boldsymbol{t}) = \prod_n \prod_{p=1}^{\bar{y}_n} p(s_n^p|\boldsymbol{w}_n, t_n), \quad (6)$$

where the set of ToAs are gathered in $\boldsymbol{S} = \{\boldsymbol{s}_n\}_{n=1}^N$. The next section discusses different prior models associated with the unknown parameters $(\boldsymbol{W}, \boldsymbol{t})$ in Eq. (6).

III. Prior Models

In this work, we address the recovery of $\boldsymbol{W}$ (allowing subsequent estimation of $\boldsymbol{R}$) and $\boldsymbol{t}$ from $\boldsymbol{Y}$ using a Bayesian approach and thus assign prior distributions to these parameters, to incorporate additional information available and regularize the inference problem. Although the depth and reflectivity profiles are expected to be correlated, modeling this correlation is difficult and usually complicates the inference process. To keep the inference process tractable, as in [21], [22], [30], [31] we assume prior independence between the reflectivity and range profiles. Since the estimation quality of $\boldsymbol{t}$ mainly depends on the illumination level (only weakly on the prior model $p(\boldsymbol{t})$, as long as it is informative), a single prior model is used to describe our prior knowledge about $\boldsymbol{t}$, while we consider several prior models for $\boldsymbol{W}$.

A. Prior Models for $\boldsymbol{W}$

In this work, we investigate several prior models for $\boldsymbol{W}$, as such models can have a significant impact on the reconstruction performance and on the computational cost of the estimation strategy. More precisely, we consider models that ensure the positivity and sum-to-one constraints of the elements of $\boldsymbol{W}$, while also promoting correlation among pixels. We consider only proper prior models (in the Bayesian sense) and leave more complex models (e.g., using trained networks) to future work (see Conclusion).

1) TV-based prior model: A classical choice for promoting spatial correlation between parameters in neighboring pixels are Markov random fields (MRFs) [31]–[33]. The first MRF investigated here is based on a total-variation (TV) regularization [34], [35] strategy, which promotes piece-wise constant images, and can be expressed as

$$p(\boldsymbol{W}|\lambda) \propto \exp\left[-\lambda \sum_{l=1}^{L} \|\boldsymbol{w}_{l,:}^\top\|_{TV}\right] \mathbf{1}_{S_L^N}(\boldsymbol{W}), \quad (7)$$

where $\mathbf{1}_{S_L^N}(\cdot)$ is the indicator function defined on S_L^N and λ is a user-defined hyper-parameter controlling the amount of prior smoothness of the rows of $\boldsymbol{W}$. The ℓ_1-based total variation (TV) regularization of a vectorized image $\boldsymbol{x}$ is defined as $\|\boldsymbol{x}\|_{TV} = \|\nabla_v \boldsymbol{x}\|_1 + \|\nabla_h \boldsymbol{x}\|_1$, where ∇_v (resp. ∇_h) is the linear operator computing the discretized vertical (resp. horizontal) finite difference of the image [36]. Note that this prior model is log-concave but not differentiable and that although the rows of $\boldsymbol{W}$ seem a priori independent, they are correlated through the indicator function in Eq. (7). This prior model will be referred as "TV" in this paper.

2) Curvature-based prior model: Another classical MRF is based on the Laplacian operator, which imposes an ℓ_2-norm penalization for the mean (spatial) curvature of images. The resulting model can be expressed as

$$f(\boldsymbol{W}|\lambda) \propto \exp\left[-\frac{\lambda}{2} \sum_{l=1}^{L} \|\boldsymbol{L}_a \boldsymbol{w}_{l,:}^\top\|_2^2\right] \mathbf{1}_{S_L^N}(\boldsymbol{W}), \quad (8)$$

where $\boldsymbol{L}_a$ is the $(N \times N)$ Laplacian operator [37], [38]. This model is log-concave and differentiable on S_L^N. We will refer to this model as "Lap".

3) Dirichlet-Based Prior Models: The two MRFs presented above promote local spatial correlation but are global models correlating all the image pixels. We also investigate three increasingly informative, yet simpler, prior models that can significantly reduce the computational complexity of the inference process (as will be discussed in Section V). These models are based on the Dirichlet distribution, which stands as a natural choice to satisfy the constraints induced by S_L^N.

The first and second models are based on

$$f(\boldsymbol{W}|\boldsymbol{\beta}) = \prod_n \text{Dir}(\boldsymbol{w}_n|\boldsymbol{\beta}), \quad (9)$$

where $\text{Dir}(\cdot|\boldsymbol{\beta})$ is the probability density function of the Dirichlet distribution with parameters $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_{L+1}]^\top$. Here these parameters are assumed to take values in $(1, \infty)$ to ensure the strict log-concavity of Eq. (9) with respect to $\mathbf{W}$. Using Eq. (9), all the vectors $\boldsymbol{w}_n$ share the same hyper-parameters, and are a priori independent, conditioned on $\boldsymbol{\beta}$. Moreover the model in Eq. (9) does not promote any spatial correlation and seems consequently well suited when the data is informative enough or when no strong regularization is required. Its main advantage is that it allows the vectors $\boldsymbol{w}_n$ to be updated independently (see Section V). Our first model based on Eq. (9) uses a fixed and constant set of hyper-parameters, i.e., $\boldsymbol{\beta} = \kappa \mathbf{1}_{L+1}$, where $\mathbf{1}_{L+1}$ is the $(L+1)$ column vector of ones. In practice κ is set to $\kappa = 1.01$ to obtain a log-concave, weakly informative prior model. This first model is referred to as "W-Dirichlet" (weak Dirichlet-based model) in the remainder of the paper.

The second prior model embeds Eq. (9) in a hierarchical model, i.e., considers $\boldsymbol{\beta}$ is an unknown parameter vector which is assigned as prior model a product of $(L+1)$ independent and identical truncated exponential distributions with fixed hyper-parameter θ, that is

$$f(\boldsymbol{\beta}|\theta) \propto \prod_{l=1}^{L+1} \theta e^{-\theta\beta_l} \mathbf{1}_{(1,\infty)}(\boldsymbol{\beta}). \tag{10}$$

In this case, $\boldsymbol{\beta}$ is estimated together with $\boldsymbol{W}$ and the truncation is introduced to ensure the log-concavity of (9) with respect to $\boldsymbol{W}$. In all the results presented in Section V we fixed $\theta = 1/4$. This model is referred to as "G-Dirichlet" (global Dirichlet-based model with unknown $\boldsymbol{\beta}$).

The third Dirichlet-based model is similar to the G-Dirichlet model but is applied to groups of pixels instead of the whole image [39]. Let us assume that the image pixels are clustered in C distinct and known groups, each group presenting specific spectral profiles. The procedure to form these groups, which are generally unknown will be discussed in Section IV-E. Let I_c be the set of pixel indices in the cth group. The joint prior model for $\boldsymbol{W}$ can be written as

$$f(\boldsymbol{W}|\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_C) = \prod_{c=1}^{C} \prod_{n \in I_c} \text{Dir}(\boldsymbol{w}_n|\boldsymbol{\beta}_c), \tag{11}$$

where $\boldsymbol{\beta}_c$ is a group-dependent vector of hyper-parameters. In a similar fashion to the G-Dirichlet, the vectors $\{\boldsymbol{\beta}_c\}_c$ are assumed unknown, a priori independent and are assigned prior models as in Eq. (10). As for the first two Dirichlet-based models, this model does not explicitly account for local correlation between pixels (the vectors $\boldsymbol{w}_n$ are a priori mutually independent for a fixed clustering and given values of $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_C$) but (local and non-local) correlation can be implicitly introduced when forming the clusters (see Section IV-E). We will refer to this model as "C-Dirichlet" (cluster-based Dirichlet-based model).

In Section IV, we denote by $\boldsymbol{\Phi}$ the set of unknown hyper-parameters involved in the prior model of $\boldsymbol{W}$, i.e., $\boldsymbol{\Phi} = \emptyset$ for Eqs. (7) and (8) and W-Dirichlet, $\boldsymbol{\Phi} = \boldsymbol{\beta}$ for G-Dirichlet and $\boldsymbol{\Phi} = \{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_C\}$ for C-Dirichlet.

B. Prior Model for the Range Profile

Recent single-photon avalanche diode (SPAD) detectors [40] offer a timing resolution of several picoseconds which allows sub-centimeter range resolution. Thus, it makes sense to assume that the elements of $\boldsymbol{t}$ are defined on the discrete grid with the same resolution, as in [27], [41], and take values in $\{t_{min}, \ldots, t_{max}\}$ such that $1 < t_{min} < t_{max} < T$. Using discrete depths also simplifies the estimation of $\boldsymbol{t}$. The bounds (t_{min}, t_{max}) ensure that the integrals $\{G_l\}_l$ are indeed constant, irrespective of the position of the objects. Since real scenes are often composed of distinct surfaces, it is reasonable to define a prior model for $\boldsymbol{t}$ which preserves sharp edges. As in [22], [31], the following TV-based MRF is used in this work

$$p(\boldsymbol{t}|\epsilon) = \exp\left[-\epsilon \|\boldsymbol{t}\|_{TV}\right], \tag{12}$$

where the fixed (user-defined) hyper-parameter ϵ, which mostly depends on the range resolution of the single-photon detector, determines the level of correlation between the depth parameters of neighboring pixels. In Section V, we set $\epsilon = 0.05$ for all the experiments.

IV. Estimation Strategy

In Section III, we defined joint prior models for $(\boldsymbol{W}, \boldsymbol{t})$ or $(\boldsymbol{W}, \boldsymbol{t}, \boldsymbol{\Phi})$ (for the G-Dirichlet and C-Dirichlet models). In this section, we primarily present the estimation strategy to estimate $(\boldsymbol{W}, \boldsymbol{t})$ (and subsequently $\boldsymbol{R}$) for ease of understanding. Cases where $\boldsymbol{\Phi}$ is not empty will be only briefly discussed in Section IV-B as $\boldsymbol{\Phi}$ can be updated together with $\boldsymbol{W}$. Note that the fixed hyper-parameters (e.g., ϵ and λ for the models in Eqs. (7)-(8)) are omitted for brevity in the equations of the remainder of this paper. Using the Bayes rule, the joint posterior distribution of $(\boldsymbol{W}, \boldsymbol{t})$ can be expressed as

$$p(\boldsymbol{W}, \boldsymbol{t}|\boldsymbol{S}, \boldsymbol{\Phi}) \propto p(\boldsymbol{S}|\boldsymbol{W}, \boldsymbol{t}) f(\boldsymbol{W}) p(\boldsymbol{t}). \tag{13}$$

However, as mentioned previously, the joint estimation of $\boldsymbol{W}$ and $\boldsymbol{t}$ is challenging, mainly due to the multimodal nature of $p(\boldsymbol{S}|\boldsymbol{W}, \boldsymbol{t})$. To circumvent this issue, the joint estimation problem is decomposed into two successive problems where we first consider $\boldsymbol{t}$ as a nuisance vector to be marginalized to estimate $\boldsymbol{W}$ via marginal maximum a posteriori (MMAP) estimation. Once $\boldsymbol{W}$ is estimated, the depth profile can then be inferred from the posterior distribution $p(\boldsymbol{t}|\widehat{\boldsymbol{W}}_{MMAP}, \boldsymbol{S})$, e.g., via MMAP estimation, conditioned on the estimated proportions $\widehat{\boldsymbol{W}}_{MMAP}$.

A. Estimation of mixture fractions

The estimation of $\boldsymbol{W}$ is performed using

$$\widehat{\boldsymbol{W}}_{MMAP} = \underset{\boldsymbol{W}}{\text{argmax}} \quad f(\boldsymbol{W}|\boldsymbol{S}), \tag{14}$$

which can be performed using an EM-based algorithm. A typical iteration of the classical EM algorithm [42] is decomposed into two steps. Let $\boldsymbol{W}^{(i)}$ be the current estimate of $\boldsymbol{W}$. The first step consists of computing

$$Q(\boldsymbol{W}|\boldsymbol{W}^{(i)}) = \text{E}_{p(\boldsymbol{t}|\boldsymbol{W}^{(i)}, \boldsymbol{S})}\left[\log p(\boldsymbol{W}, \boldsymbol{t}|\boldsymbol{S})\right], \tag{15}$$

while the second step updates the estimate of $\boldsymbol{W}$ via

$$\boldsymbol{W}^{(i+1)} = \underset{\boldsymbol{W}\in S_L^N}{\operatorname{argmax}} \; Q(\boldsymbol{W}|\boldsymbol{W}^{(i)}). \tag{16}$$

Unfortunately, the first step in (15) cannot be computed directly as the expectation $\mathrm{E}_{p(\boldsymbol{t}|\boldsymbol{W}^{(i)},\boldsymbol{S})}[\boldsymbol{t}]$ is intractable because of the MRF prior distribution assigned to $\boldsymbol{t}$ in (12). In such cases, a classical approach is to resort to stochastic EM (SEM) algorithms [43], [44]. In this work, to keep the overall computational cost of the method low, we adopt the strategy detailed in [45]. More precisely, we replace $p(\boldsymbol{t}|\boldsymbol{W}^{(i)},\boldsymbol{S})$ in (15) by an approximating distribution such that the expectation becomes tractable. In a similar fashion to [46], the approximating distribution is constructed by generating an auxiliary random sample $\tilde{\boldsymbol{t}}$ from $p(\boldsymbol{t}|\boldsymbol{W}^{(i)},\boldsymbol{S})$ and we then use as approximation

$$\tilde{p}(\boldsymbol{t}|\tilde{\boldsymbol{t}},\boldsymbol{W}^{(i)},\boldsymbol{S}) \approx p(\boldsymbol{t}|\boldsymbol{W}^{(i)},\boldsymbol{S}), \tag{17}$$

with

$$\tilde{p}(\boldsymbol{t}|\tilde{\boldsymbol{t}},\boldsymbol{W}^{(i)},\boldsymbol{S}) = \prod_{n=1}^{N} p(t_n|\tilde{\boldsymbol{t}}_{\setminus n},\boldsymbol{W}^{(i)},\boldsymbol{S}), \tag{18}$$

where $\tilde{\boldsymbol{t}}_{\setminus n}$ is the vector $\tilde{\boldsymbol{t}}$ whose nth element has been removed and $p(\cdot|\tilde{\boldsymbol{t}}_{\setminus n},\boldsymbol{W}^{(i)},\boldsymbol{S})$ is the conditional distribution of the nth element of $\tilde{\boldsymbol{t}}$ associated with the joint distribution $p(\tilde{\boldsymbol{t}}|\boldsymbol{W}^{(i)},\boldsymbol{S})$. Note that additional details about the generation of the random sample $\tilde{\boldsymbol{t}}$ and the approximating distribution are provided in Section IV-E. The approximation in Eq. (17) can significantly simplify the computation of expectations as it reduces to a product of N independent distributions and expectations become easier to compute.

Instead of computing (15)-(16), the two EM steps become

$$\widetilde{Q}(\boldsymbol{W}|\boldsymbol{W}^{(i)}) = \mathrm{E}_{\tilde{p}(\boldsymbol{t}|\tilde{\boldsymbol{t}},\boldsymbol{W}^{(i)},\boldsymbol{S})}\left[\log p(\boldsymbol{W},\boldsymbol{t}|\boldsymbol{S})\right], \tag{19}$$

and

$$\boldsymbol{W}^{(i+1)} = \underset{\boldsymbol{W}\in S_L^N}{\operatorname{argmax}} \; \widetilde{Q}(\boldsymbol{W}|\boldsymbol{W}^{(i)}). \tag{20}$$

Defining $\tilde{p}_{n,k} = p(t_n = k|\tilde{\boldsymbol{t}}_{\setminus n},\boldsymbol{W}^{(i)},\boldsymbol{S}), \forall (n,k)$, we obtain

$$\widetilde{Q}(\boldsymbol{W}|\boldsymbol{W}^{(i)}) = \widetilde{Q}_1(\boldsymbol{W}|\boldsymbol{W}^{(i)}) + \widetilde{Q}_2(\boldsymbol{W}|\boldsymbol{W}^{(i)}) + C_1, \tag{21}$$

where C_1 is a constant which does not depend on $\boldsymbol{W}$, $\widetilde{Q}_1(\boldsymbol{W}|\boldsymbol{W}^{(i)}) = \log\left(f(\boldsymbol{W})\right)$ and

$$\widetilde{Q}_2(\boldsymbol{W}|\boldsymbol{W}^{(i)}) = \sum_{n,k} \tilde{p}_{n,k} \sum_{p=1}^{\bar{y}_n} \log\left(p(s_n^p|\boldsymbol{w}_n, t_n = k)\right). \tag{22}$$

The likelihood $p(s_n^p|\boldsymbol{w}_n, t_n)$ in Eq. (3) is concave with respect to (w.r.t.) $\boldsymbol{W}$, which results in $\widetilde{Q}(\boldsymbol{W}|\boldsymbol{W}^{(i)})$ being a concave function w.r.t. $\boldsymbol{W} \in S_L^N$. Thus, the solution in (20) can be obtained using convex optimization methods. This will be discussed further in Section IV-E.

B. Estimation of the hyper-parameters

The method proposed can be easily extended to estimate the hyper-parameters in $\boldsymbol{\Phi}$ associated with $\boldsymbol{W}$. In that case, Eqs. (19) and (20) become

$$\widetilde{Q}(\boldsymbol{W},\boldsymbol{\Phi}|\boldsymbol{W}^{(i)},\boldsymbol{\Phi}^{(i)}) = \mathrm{E}_{\tilde{p}(\boldsymbol{t}|\tilde{\boldsymbol{t}},\boldsymbol{W}^{(i)},\boldsymbol{S},\boldsymbol{\Phi}^{(i)})}\left[\log p(\boldsymbol{W},\boldsymbol{t},\boldsymbol{\Phi}|\boldsymbol{S})\right], \tag{23}$$

and

$$(\boldsymbol{W}^{(i+1)},\boldsymbol{\Phi}^{(i+1)}) = \underset{\boldsymbol{W}\in S_L^N,\boldsymbol{\Phi}}{\operatorname{argmax}} \; \widetilde{Q}(\boldsymbol{W},\boldsymbol{\Phi}|\boldsymbol{W}^{(i)},\boldsymbol{\Phi}^{(i)}), \tag{24}$$

where

$$p(\boldsymbol{W},\boldsymbol{t},\boldsymbol{\Phi}|\boldsymbol{S}) \propto p(\boldsymbol{S}|\boldsymbol{W},\boldsymbol{t})p(\boldsymbol{t})p(\boldsymbol{W}|\boldsymbol{\Phi})p(\boldsymbol{\Phi}), \tag{25}$$

and $p(\boldsymbol{\Phi})$ consists of a product of truncated exponential distributions (see Section III-A3). Sampling the auxiliary variable $\tilde{\boldsymbol{t}}$ needed in (23) can be achieved as when $\boldsymbol{\Phi} = \emptyset$. Optimizing (24) globally is challenging as the problem is not convex and alternating optimization w.r.t. $\boldsymbol{W}$ and $\boldsymbol{\Phi}$ can be adopted to ensure local convergence. Instead, to reduce the computational complexity, we simply solve

$$\begin{cases} \boldsymbol{W}^{(i+1)} = \underset{\boldsymbol{W}\in S_L^N}{\operatorname{argmax}} \; \widetilde{Q}(\boldsymbol{W},\boldsymbol{\Phi}^{(i)}|\boldsymbol{W}^{(i)},\boldsymbol{\Phi}^{(i)}) \\ \boldsymbol{\Phi}^{(i+1)} = \underset{\boldsymbol{\Phi}}{\operatorname{argmax}} \; \widetilde{Q}(\boldsymbol{W}^{(i+1)},\boldsymbol{\Phi}|\boldsymbol{W}^{(i)},\boldsymbol{\Phi}^{(i)}) \end{cases}, \tag{26}$$

leading to a generalized EM algorithm [43]. Both lines in Eq. (26) can be solved using convex optimization as Eq. (20).

The resulting algorithm is summarized in Algo. 1. Due to the stochastic nature of the EM-based method used, i.e., due to the sampling step (Line 5 in Algo. 1), the proposed iterative procedure does not converge exactly to the MMAP estimator of $\boldsymbol{W}$ but after some iterations, referred to as burn-in iterations, the generated $\boldsymbol{W}^{(i)}$ oscillate around it. The number of burn-in iterations N_{bi} and the total number of iterations N_{iter} are user-defined parameters. For all the simulations results presented in Section V however, the convergence of the algorithm is fast (less than 5-10 iterations) and the oscillations are quite small. This can be monitored using the relative error between successive $\boldsymbol{W}^{(i)}$. Thus, we stop the burn-in when the error becomes smaller than a fixed threshold $d_\epsilon = 10^{-10}$. Since the oscillations are small in practice, we used $N_{iter} = N_{bi} + 5$. After N_{iter} iterations, we finally use the average of the last $N_{\text{iter}} - N_{\text{bi}}$ generated values of $\boldsymbol{W}$ as our estimate $\widehat{\boldsymbol{W}}_{MMAP}$ (Line 14 in Algo. 1).

ALGORITHM 1

EM-based estimation of $\boldsymbol{W}$

1: Fixed input parameters: ϵ, λ, number of burn-in iterations N_{bi}, total number of iterations $N_{\text{iter}} > N_{\text{bi}}$.
2: Initialization ($k = 0$)
3: Set $\boldsymbol{W}^{(0)}$ (and $\boldsymbol{\Phi}^{(0)}$)
4: **for** $i = 0, \ldots N_{\text{iter}}$ **do**
5: Sample $\tilde{\boldsymbol{t}} \sim p(\boldsymbol{t}|\boldsymbol{W}^{(i)},\boldsymbol{S},\boldsymbol{\Phi}^{(i)})$.
6: Compute $\tilde{p}_{n,k} = p(t_n = k|\tilde{\boldsymbol{t}}_{\setminus n},\boldsymbol{W}^{(i)},\boldsymbol{S},\boldsymbol{\Phi}^{(i)}), \forall (n,k)$
7: Compute $\boldsymbol{W}^{(i+1)}$ using Eq. (20) (or (26)).
8: **if** $\boldsymbol{\Phi}^{(i)} \neq \emptyset$ **then**
9: Compute $\boldsymbol{\Phi}^{(i+1)}$ using Eq. (26).
10: **else**

11: Set $\boldsymbol{\Phi}^{(i+1)} = \boldsymbol{\Phi}^{(i)}$.
12: **end if**
13: **end for**
14: Set $\widehat{\boldsymbol{W}}_{MMAP} = 1/(N_{\text{iter}} - N_{\text{bi}}) \sum_{i=N_{\text{bi}}+1}^{N_{\text{iter}}} \boldsymbol{W}^{(i)}$

C. Estimation of the spectral responses

The algorithm presented in the previous section estimates $\boldsymbol{W}$ but does not provide directly an estimate of $\boldsymbol{R}$. It can be seen from (1) that

$$E_{p(\boldsymbol{y}_n|\boldsymbol{r}_n,t_n,b_n)}[\bar{y}_n] = \sum_{l=1}^{L} r_{n,l} G_l + T b_n, \tag{27}$$

i.e., that $r_{n,l} = \frac{w_{n,l} E[\boldsymbol{y}_n]}{G_l}$ using (4). While it is possible to estimate $r_{n,l}$ using $\hat{w}_{n,l}\bar{y}_n/G_l$, this estimate, which assumes $\bar{y}_n \approx E[\bar{y}_n]$, is in practice too noisy, especially in the photon-starved regime. Thus, we propose to denoise the image $\bar{\boldsymbol{y}} = \{\bar{y}_n\}_n$ to get a better estimate $\hat{\boldsymbol{y}} = \{\hat{y}_n\}_n$, leading to

$$\hat{r}_{n,l} = \frac{\hat{w}_{n,l}\hat{y}_n}{G_l}. \tag{28}$$

Using the fact that $\bar{\boldsymbol{y}}$ is corrupted by Poisson noise, $\hat{\boldsymbol{y}}$ is obtained by denoising the image $\bar{\boldsymbol{y}}$ using a variant of the BM3D algorithm [47], i.e., the method developed in [48] to account for Poisson noise. While other denoising algorithms could be used, this denoiser provides state-of-the-art reconstruction results with a reduced computational time and limited user supervision. Here, we used the default parameter settings of the code available at http://www.cs.tut.fi/~foi/invansc/.

D. Estimation of the depth profile

Following the computation of $\widehat{\boldsymbol{W}}_{MMAP}$ via the EM-based algorithm, the depth profile is finally estimated using

$$\hat{t}_{n,MMAP} = \underset{t_n}{\text{argmax}} \sum_{\boldsymbol{t}_{\setminus n}} p(\boldsymbol{t}|\widehat{\boldsymbol{W}}_{MMAP}, \boldsymbol{S}), \forall n \tag{29}$$

These estimates are obtained via Monte Carlo simulation from $p(\boldsymbol{t}|\widehat{\boldsymbol{W}}_{MMAP}, \boldsymbol{S})$, which can be performed efficiently using a 2-step Gibbs sampler, exploiting the MRF structure of the TV-based prior model (12), as in [41]. In all the experiments carried out and presented in Section V, the Gibbs sampler used to estimate the depth profile was stopped after 300 iterations, with a burn-in period of 50 iterations.

E. Computational considerations

Mazimization step: although a thorough study of algorithms to solve (20) is out of scope of this paper, we discuss here two different approaches that have been used depending on the prior model for $\boldsymbol{W}$. With the prior models (7) or (8), we used the classical alternating direction method of multipliers (ADMM) [49] as it can easily handle the constraints on $\boldsymbol{W}$ as well as smooth or non-smooth prior models. The same ADMM stopping criterion as in [49] has been used in all our experiments. Since the splitting schemes (e.g., the number of splitting variables) varies, the different ADMM schemes implemented are not further detailed here. Note however that some steps (e.g., involving $\widetilde{Q}_1(\boldsymbol{W}|\boldsymbol{W}^{(i)})$) in the ADMM sequential approach cannot be computed analytically and require optimization sub-routines. In this case, where the cost function is differentiable, we adopted a line search approach. Since a variable-splitting approach is used, the overall computational cost of the method can be significant, as it requires storing several matrices (the splitting variables) that have the same size as $\boldsymbol{W}$. Using the Dirichlet-based models, the optimization in (20) can be performed pixel-wise and $\widetilde{Q}(\boldsymbol{W}|\boldsymbol{W}^{(i)})$ is twice differentiable and strictly concave. Thus we can use second-order methods, e.g., a simple Newton-Raphson (NR) algorithm to solve (20) efficiently. Note that the complexity of the method depends on the number of wavelengths considered, as Hessian matrices of size $L+1$ have to be inverted. When $\boldsymbol{\Phi} \neq \emptyset$, the second line of Eq. (26) is solved iteratively using a component wise NR algorithm, where the elements of $\boldsymbol{\Phi}$ are updated sequentially until convergence. As will be seen in Section V, this makes Dirichlet-based models computationally more attractive than MRF-based models.

Sampling the auxiliary variable $\boldsymbol{t}$: the sampling step in Line 5 of Algo. 1 requires sampling from a discrete MRF, which can be done in a similar fashion to the final estimation of $\boldsymbol{t}$ in Section IV-D. However, using a Gibbs sampling approach with a random initialization would yield a long burn-in period to obtain a sample from $p(\boldsymbol{t}|\boldsymbol{W}^{(i)}, \boldsymbol{S}, \boldsymbol{\Phi}^{(i)})$. Instead, we use the value of $\tilde{\boldsymbol{t}}$ generated at the previous iteration to hot-start the Gibbs sampler and only iterate a reduced number of times (1 to 2 in practice) to generate a sample.

MRF hyper-parameter λ: the dynamic range of $\boldsymbol{W}$ decreases in the presence of high background and λ in (7) and Eq. (8) should be adjusted accordingly if prior information about the signal to background ratio is available. In Section V however, we do not use such information and set $\lambda = 10$ in (7) and Eq. (8) for all the experiments as this value leads to satisfactory results on average, irrespective of the level of background in the range considered. Note that the level of user-supervision required for fine-tuning λ is one of the motivations for using the C-Dirichlet model, whose parameters are more robust to illumination conditions.

Pixel clustering for the C-Dirichlet model: the C-Dirichlet model defined in Section III relies on a user-defined clustering of the pixels, which is difficult to set beforehand. To define the C groups of pixels, we first run Algo. 1 for a few iterations using the W-Dirichlet model to get a coarse initial guess $\bar{\boldsymbol{W}}_0$ of $\boldsymbol{W}$ which is used to create the clusters. While direct clustering of the N L-dimensional vectors in $\bar{\boldsymbol{W}}_0$ is possible, it is in practice prone to noise and patch-based clustering is preferred here. More precisely, we first construct from $\bar{\boldsymbol{W}}_0$ a set of N 3D patches of size $N_p \times N_p \times L$ (one patch per pixel). The patches are then vectorized and clustered into C groups via k-means clustering with the ℓ_2 distance as similarity measure between patches [50]. In all the results presented in Section V, we used $C = 7$ classes of (3×3) pixels patches. Note however that different clustering strategies, e.g., [51]–[54] could be used instead of k-means.

V. Results

We first assess the performance of the proposed approach with the different prior models for $\boldsymbol{W}$ using synthetic SBL data, i.e., with $L = 1$ in Section V-A and then apply our approach on real SW-MSL data in Section V-B with a reduced number of prior models.

A. Analysis of synthetic SBL data

In this section, we generated synthetic data using the $N = 200 \times 200$ pixels depth and reflectivity profiles, depicted in Fig. 2 (top), which are actual depth and reflectivity (at 532nm) profiles obtained in [21] and [22]. The object is approximately 42 mm tall and 30 mm wide and the interested reader is invited to consult [21], [22] for a comprehensive description of the experimental setup. The associated IRF $g(.)$ shown in Fig. 2 (bottom) in orange (532nm) was obtained during the calibration of the imaging system. The number of temporal bins was set to $T = 1500$ (bin width of 2ps) and a spatially constant background $\boldsymbol{b} = \{b_n\}_n$ was added to all of the waveforms. For all of the experiments presented in this paper, we set the admissible temporal positions to $[t_{min}, t_{max}] = [301, T - 600]$ to ensure the integrals of the IRFs remain constant over the admissible object range.

To control the quality of the generated data, we introduce as in [30], two parameters (α and γ) controlling the mean signal detection and the background levels. More precisely, datasets have been generated using Eq. (2), such that the average number of signal and background photons in the nth pixel are $\alpha r_n G_1$ and $b_n = \alpha\gamma T$, respectively. The parameter α is chosen in $\{25, 100\}$ and γ takes values in $\{0.01, 0.05, 0.1, 0.3, 0.5, 1, 5\}/T$. This leads to average counts per pixel from 10.3 (for $(\alpha, \gamma) = (25, 0.01/T)$) to 542.7 (for $(\alpha, \gamma) = (100, 5/T)$). In the latter case, about 500 of these photons arise from the background. With this parametrization, and average signal to background ratio (SBR)

$$\text{SBR} = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{Tb_n}\sum_{l=1}^{L} r_{n,l}G_l,$$

can be expressed as $\text{SBR} = \frac{1}{NT\gamma}\sum_{n=1}^{N}\sum_{l=1}^{L} r_{n,l}G_l$, and it ranges from 0.086 (for $\gamma = 5/T$) to 42 (for $\gamma = 0.01/T$).

The proposed approach is compared in the SBL case with two approaches designed to reconstruct 3D surfaces assuming exactly one visible surface per pixel:

Cross-correlation approach: denoted as "Xcorr", it corresponds to the classical matched-filter [55]. The depth profile is estimated by maximizing the correlation between the IRF and each histogram $\boldsymbol{y}_n$. These estimates are then used in Eq. (3) to estimate $\boldsymbol{W}$ via maximum likelihood, and then $\boldsymbol{R}$ using the method described in Section IV-C. This strategy yields better reflectivity estimates than the standard estimation of $(\boldsymbol{R}, \boldsymbol{b})$ from Eq. (3), conditioned on the estimated depth profile.

Photon unmixing algorithm [27]: Pixel patches of size 5×5 have been used to locally improve the SBR and the probability threshold has been set to $\tau_{FA} = 0.05$ to censor background photons before parameter estimation (see [27] for additional details). These values have been optimized manually to obtain the best results on average for each (α, γ), according to the metric in (30). This method assumes that the SBR is known.

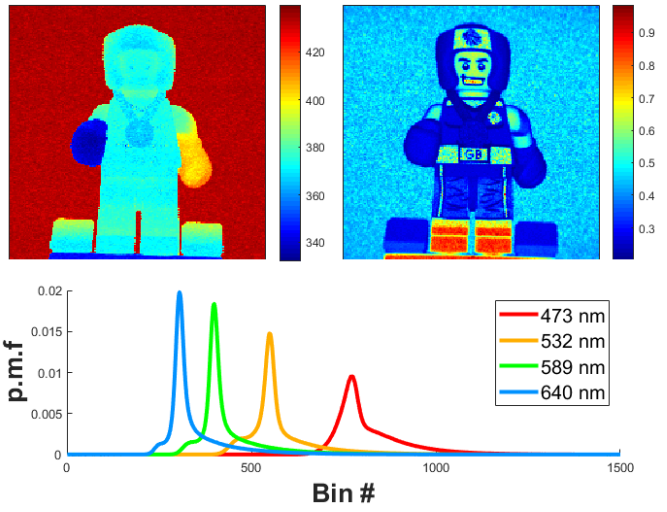

Fig. 2. Top: depth (left) and reflectivity profiles used to simulate SBL data. Bottom: normalised IRFs from [22], corresponding to the probability mass functions (p.m.f) of the photons ToAs associated with the wavelengths $[473, 532, 589, 640]$nm. The orange IRF is that used in Section V-A.

Note that the method proposed in [30] coincides with our Lap model for $L = 1$, i.e., when the Dirichlet distribution becomes a beta distribution, and this method is thus not included above.

The reflectivity estimation is assessed using the mean squared error

$$\text{MSE} = \frac{1}{N}\sum_{n=1}^{N}\|\boldsymbol{r}_n - \hat{\boldsymbol{r}}_n\|_2^2, \tag{30}$$

where $\boldsymbol{r}_n$ (resp. $\hat{\boldsymbol{r}}_n$) is the actual (resp. estimated) spectral response of the nth pixel. The MSEs obtained with the different methods for $\alpha = 25$ and $\alpha = 100$ as a function of γ are shown in Fig. 3. In the low illumination scenario ($\alpha = 25$), this figure shows that the TV, Lap and C-Dirichlet models perform similarly well and that the C-Dirichlet model provides the most accurate estimate for high values of γ. The W-Dirichlet and G-Dirichlet models provide the least accurate

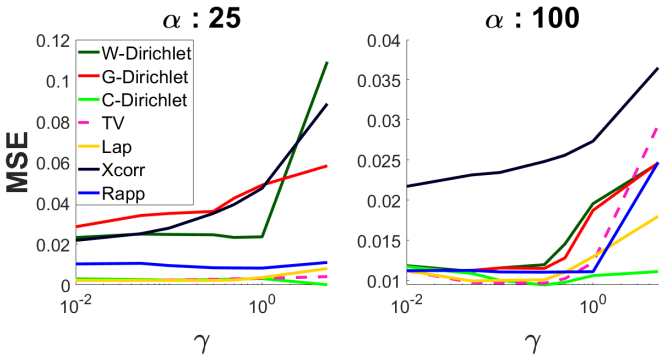

Fig. 3. MSE of the reflectivity (averaged over 3 realizations) obtained with the different competing methods for $L = 1$, as function of γ, with $\alpha = 25$ (left) and $\alpha = 100$ (right).

reflectivity reconstruction. The results obtained with the Xcorr model are close to that of W-Dirichlet, explained by the lack of strong regularization. We obtained similar results in the high illumination scenario ($\alpha = 100$), where only the TV, Lap and C-Dirichlet models allow more accurate reconstruction than the method from [27] for low background levels ($\gamma < 0.5$), and only the C-Dirichlet provides better reconstruction in term of the MSE than the method of Rapp [27].

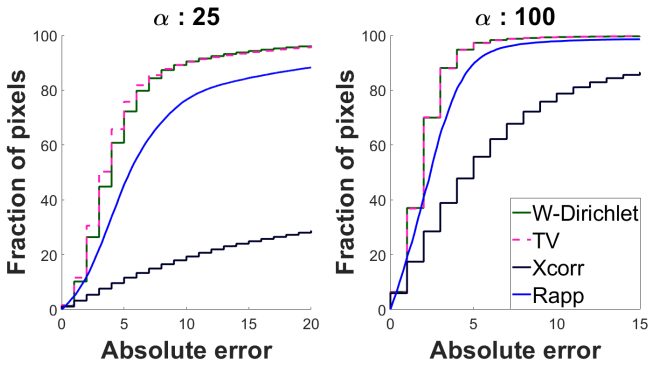

Fig. 4. CDFs of the depth absolute error in bin obtained using the competing methods for data generated with $L = 1$, $\gamma = 5$ and $\alpha = 25$ (left) and $\alpha = 100$ (right).

The ranging performance is quantified using the cumulative density function (CDF) of the depth absolute error $h_n = |t_n - \hat{t}_n|, \forall n$, where $\hat{t}_n$ is the estimate of the actual depth parameter t_n in the nth pixel. The depth error CDFs obtained with $\alpha \in \{25, 100\}$ are depicted in Fig. 4. Note that TV and Lap perform similarly and that the Dirichlet-based priors also perform similarly. Thus the results obtained with Lap, G-Dirichlet and C-Dirichlet are omitted here. In the low illumination scenario, Xcorr provides the poorest estimation, whereas TV, similarly to W-Dirichlet, achieves the most accurate. Note that among the prior models we proposed, W-Dirichlet and TV provide respectively the least and most accurate estimates and both perform better than the method from [27] for $\alpha = 25$. With $\alpha = 100$, all the methods (except Xcorr) perform similarly. Interestingly, the choice of the prior model for $\boldsymbol{W}$ has in general a limited impact on the quality of $\hat{\boldsymbol{t}}$, which is more affected by the (lack of) depth regularization.

Table I illustrates the computational cost of our method with SBL data, for the different models considered. All the experiments reported in this work have been obtained using Matlab R2019a running on a workstation with an Intel i7-8700k CPU @ 3.70GHz. The time corresponds to that the execution of Algo. 1 and does not include the estimation of $\mathbf{t}$ which has a fixed complexity (for fixed (N, L, T)), that is 106s here. It however includes the estimation of the reflectivity profile from $\boldsymbol{W}$, which takes 0.25s and is negligible compared to the other steps. Since Lap (resp. G-Dirichlet) has a complexity similar to that of TV (resp. W-Dirichlet), the corresponding results are omitted. While the convergence speed is not significantly affected by the SBR or number of photons, the iterations become slower as the photon counts increase, mostly due to the cost of the likelihood evaluation. All the proposed methods present comparable computational costs, although W-Dirichlet (resp. TV) is generally the fastest (resp. slowest) method with SBL data. The method of Rapp [27] was originally developed for starved photon regimes. It allows a fast estimation of the model parameters in the low-illumination scenarios, but it also the most computationally expensive method when the histograms are denser.

(α, γ)	Counts / SBR	Method	N_{iter}	Time / iter	Total time
(25,0.01)	10.3/42	W-Dirichlet	12	7s	93s
		C-Dirichlet	12	8s	95s
		TV	12	10s	127s
		Rapp [27]	NA	NA	20s
(25,5)	135.6/0.086	W-Dirichlet	12	12s	155s
		C-Dirichlet	11	11s	128s
		TV	12	14s	157s
		Rapp [27]	NA	NA	125s
(100,0.01)	43.8/42	W-Dirichlet	11	17s	190s
		C-Dirichlet	12	16s	197s
		TV	12	19s	230s
		Rapp [27]	NA	NA	21s
(100,5)	542.7/0.086	W-Dirichlet	12	27s	300s
		C-Dirichlet	11	30s	336s
		TV	11	36s	392s
		Rapp [27]	NA	NA	601s

TABLE I
COMPUTATIONAL COMPLEXITY OF THE PROPOSED APPROACHES USING SBL DATA, I.E., WITH $L = 1$ (TIMINGS IN SECONDS).

Based on these first results, we only consider C-Dirichlet, Lap and TV, as the most promising models, for the analysis of SW-MSL data in the next section. The W-Dirichlet model is also considered to illustrate the impact of the lack of informative prior models on the inference process.

B. Analysis of real SW-MSL data

Here, we compare our approach to that developed in [21], using the real SW-MSL data acquired in that work, composed of $L = 4$ wavelengths ($473, 532, 589$ and 640nm), whose IRFs are shown in Fig. 2 (see [21] for additional details about the imaging setup). While perfect ground truth reflectivity and range profiles are not available, we use as reference profiles the results obtained with the method developed in [21], applied to data acquired with a long illumination time and a negligible background contribution. We consider datasets with 1.1, 5.7, 11.4 and 114.3 signal photons per pixel on average. Two scenarios are considered. One set of data was acquired with negligible background illumination while a second data set was acquired with spatially varying background (see [21]), leading to an average SBR of 1.4.

The reflectivity MSEs obtained by the different competing methods are shown in Fig. 5 for different acquisition scenarios, i.e., number of signal photons. Irrespective of the background level, C-Dirichlet is the method whose performance is the closest to the method from [21]. This figure also highlights the importance of prior information, especially in the presence of strong background. Note that TV and Lap perform generally worse than C-Dirichlet, especially when the photon counts are large. This is mainly due to the value of λ, which has been fixed for all the illumination conditions. Although C-Dirichlet does not outperform the method of Ren [21], the performance degradation is not significant and is balanced by

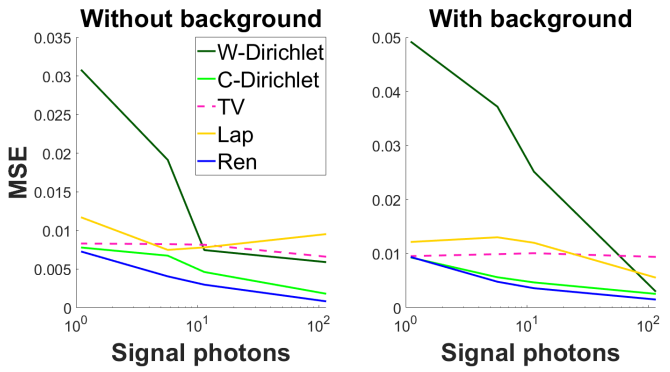

Fig. 5. MSE of $\boldsymbol{R}$ obtained with real data ($L = 4$) and the different competing methods as function of the number of photon counts, with (left) and without background illumination (right).

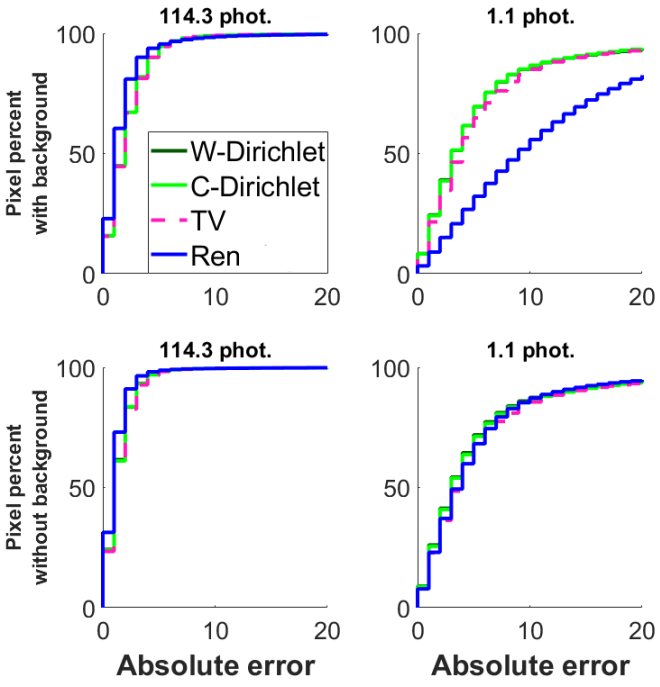

Fig. 6. CDFs of the depth absolute error in bin, with high background (SBR $=$ 1.4) (top) and with negligible background (bottom). The first (resp. second) column corresponds to 114.3 (resp. 1.1) signal photons per pixel on average.

the significantly lower computational cost, as will be shown at the end of this section.

The depth error CDFs of the different methods, obtained with 114.3 and 1.1 signal photons per pixel, are shown in Fig. 6. In this figure, Lap is omitted as it provides nearly the same results as TV. As in the SBL case (see Fig. 4), the ranging results obtained with our method do not strongly depend on the choice of the prior model for $\boldsymbol{W}$. With high photon counts (Fig. 6, left column), the method of Ren [21] provides slightly better results, while in the low illumination regime, our method performs similarly or even better (see Fig. 6, top-right subplot), due to the slow convergence of the method [21] in such cases. Conversely, our method which first marginalizes $\boldsymbol{t}$ converges quickly and is more robust in the photon-starved and high background regime. For completeness, the estimated depth

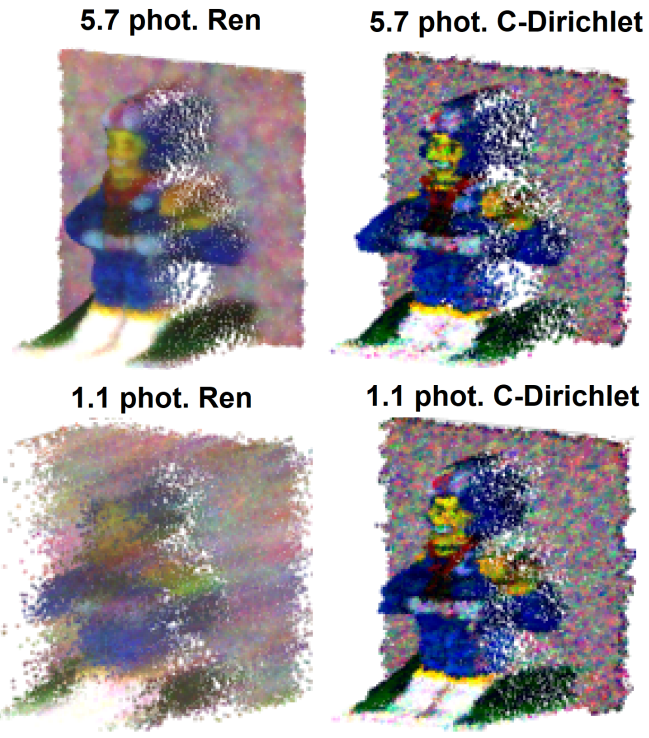

Fig. 7. Depth/RGB reconstruction using real SW-MSL data, with the method of [21] and our approach (C-Dirichlet), with 1.1 (bottom) and 5.7 (top) signal photons per pixel, with non-negligible background (SBR$=$ 1.4).

profiles obtained by Ren [21] and our approach using C-Dirichlet for 1.1 and 11.4 signal photons per pixel, and with strong ambient illumination are shown in Fig. 7. This figure illustrates the overall agreement between the two methods, except in the most extreme scenario where our new method provides a more accurate range profile. It also shows that the C-Dirichlet leads less blurred reflectivity profiles.

Counts / SBR	Method	N_{iter}	Time / iter	Total time
1.1/∞	W-Dirichlet	11	35s	386s
	C-Dirichlet	11	36s	396s
	TV	11	413s	4546s
1.9/1.4	W-Dirichlet	12	34s	433s
	C-Dirichlet	12	37s	445s
	TV	12	381s	4574s
114.3/∞	W-Dirichlet	10	53s	531s
	C-Dirichlet	11	56s	617s
	TV	11	581s	6391s
194/1.4	W-Dirichlet	13	55s	716s
	C-Dirichlet	11	58s	639s
	TV	11	609s	6702s
Any	Ren [21]	NA	NA	$\approx$ 15 hours

TABLE II
COMPUTATIONAL COST OF THE COMPETING APPROACHES FOR REAL SW-MSL DATA ANALYSIS, FOR DIFFERENT TOTAL COUNTS AND SBRS.

Finally, Table II reports the computational time associated with the different prior models and the method from [21]. The results related to the Lap are similar to those of TV and are omitted here. Note that the table does not include the estimation of $\boldsymbol{t}$, which equals 112s, irrespective of the background level and signal photon counts. It does however include the estimation of $\boldsymbol{R}$ from $\boldsymbol{W}$, which lasts 0.25s and does not depend on L. While the two Dirichlet-based models

have almost the same complexity, the TV-based model has a higher computational cost due to the use of an ADMM to solve (20), whose complexity grows with L.

The results confirm the benefits of the C-Dirichlet model, which generally achieves similar or better depth estimation and satisfactory spectral signatures reconstruction, when compared to the existing method. This is however performed at a significantly lower computational cost.

VI. Discussion

We proposed a new method for analysis of SW-MSL data which can be applied for any number of spectral bands L. However, its computational cost grows with the number of unknown parameters. The main bottlenecks are the constraints imposed by S_L^N, which require a certain class of optimization schemes to be used in the maximization step. If a second-order method is used, the inversion of the Hessian matrix can also become computationally intensive for large values of L. However, we believe that practical SW-MSL systems will only consist of up to a dozen of spectral bands, for which our algorithm remains computationally attractive. Indeed, increasing the number of bands (either by increasing the spectral range or the spectral resolution with a fixed range) requires the consideration of longer histograms to ensure all the L peaks can be resolved, i.e., to reduce the overlap between peaks. Longer histograms can increase the overall background level and longer exposures might be required as the signal photon budget has to be shared across the L bands.

While the proposed method does not allow real-time analysis of SW-MSL data, the EM-based framework offers a variety of options to further accelerate the inference process. Thus, we believe that similar results could be obtained at an even lower computational cost, using separable models allowing distributed/parallel implementations. We have shown that using a separable model simplifies the maximization step but other separable models could also be used. For instance, we used a k-means based clustering method to initialize the classes with a low computational overhead. However, more evolved methods could be investigated, e.g., as in [54], [56], provided they remain fast. While we concentrated on proper Bayesian models, it would also be interesting to investigate how our approach could be coupled and improved with tailored data-driven approaches, e.g., [57]. Another limitation of the method is that it assumes exactly one visible surface per pixel. Although empty pixels (where no surface is visible) can be estimated by thresholding the estimated spectral responses, the presence of multiple surfaces is a much more difficult problem. This is true for SBL but it becomes even more challenging in the SW-MSL context since each additional surface will create L peaks, overlapping with those of the first surface.

VII. Conclusion

In this paper, we developed a novel algorithm for spectral and depth profile estimation from SW-MSL waveforms, in the presence of non negligible background and reduced acquisition time (down to 1 signal photon per pixel, on average). The reformulation of the model into a mixture of illumination sources and ambient illumination allowed the use of an EM-based algorithm for sequential parameter estimation, allowing in turn a significant speed-up compared to the existing method for SW-MSL data analysis, without significant degradation of the reconstruction performance. We compared several reflectivity models and demonstrated the benefits of the C-Dirichlet model in terms of quality of estimation, level of user supervision and computational cost. While our new approach benefits from a significantly reduced computational cost, these new results are still incompatible with real-time requirements. Thus, future work should consist of further accelerating the inference process. An interesting route for improvement could be online reconstruction, as in [58].

Acknowledgement

We thank the single-photon group led by Prof. G. S. Buller for providing the real single-photon data used in this work. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan XP GPU used in this research.

References

[1] T. Bosch, "Laser ranging: a critical review of usual techniques for distance measurement," *Opt. Eng.*, vol. 40, no. 1, 2001.

[2] P. Vines, K. Kuzmenko, J. Kirdoda, D. Dumas, M. Mirza, R. Millar, D. Paul, and G. Buller, "High performance planar ge-on-si single-photon avalanche diode detectors," *Nature Communications*, Sept. 2018.

[3] Z.-P. Li, X. Huang, P.-Y. Jiang, Y. Hong, C. Yu, Y. Cao, J. Zhang, F. Xu, and J.-W. Pan, "Super-resolution single-photon imaging at 8.2 kilometers," *Opt. Exp.*, vol. 28, no. 3, pp. 4076–4087, 2020.

[4] A. M. Pawlikowska, A. Halimi, R. A. Lamb, and G. S. Buller, "Single-photon three-dimensional imaging at up to 10 kilometers range," *Opt. Exp.*, vol. 25, no. 10, pp. 11919–11931, 2017.

[5] Z.-P. Li, X. Huang, Y. Cao, B. Wang, Y.-H. Li, W. Jin, C. Yu, J. Zhang, Q. Zhang, C.-Z. Peng, F. Xu, and J.-W. Pan, "Single-photon computational 3D imaging at 45 km," 2019.

[6] R. Tobin, A. Halimi, A. McCarthy, M. Laurenzis, F. Christnacher, and G. S. Buller, "Three-dimensional single-photon imaging through obscurants," *Opt. Exp.*, vol. 27, pp. 4590–4611, Feb. 2019.

[7] A. Maccarone, A. McCarthy, X. Ren, R. E. Warburton, A. M. Wallace, J. Moffat, Y. Petillot, and G. S. Buller, "Underwater depth imaging using time-correlated single-photon counting," *Opt. Exp.*, vol. 23, no. 26, pp. 33911–33926, 2015.

[8] A. Halimi, A. Maccarone, A. McCarthy, S. McLaughlin, and G. S. Buller, "Object depth profile and reflectivity restoration from sparse single-photon data acquired in underwater environments," *IEEE Trans. Comput. Imaging*, vol. 3, no. 3, pp. 472–484, 2017.

[9] A. Maccarone, F. M. Della Rocca, A. McCarthy, R. Henderson, and G. S. Buller, "Three-dimensional imaging of stationary and moving targets in turbid underwater environments using a single-photon detector array," *Opt. Exp.*, vol. 27, no. 20, pp. 28437–28456, 2019.

[10] C. Premebida, J. Carreira, J. Batista, and U. Nunes, "Pedestrian detection combining RGB and dense Lidar data," in *2014 IEEE Int. Conf. on Intel. Robots and Syst.*, pp. 4112–4117, Sept. 2014.

[11] B. Bigdeli, F. Samadzadegan, and P. Reinartz, "A decision fusion method based on multiple support vector machine system for fusion of hyperspectral and Lidar data," *Int. Journal of Image and Data Fusion*, vol. 5, no. 3, pp. 196–209, 2014.

[12] P. S. Chhabra, A. Maccarone, A. McCarthy, A. M. Wallace, and G. S. Buller, "Analysis of foliage penetrating photon counting Lidar data for underwater mine counter measures," in *MTS/IEEE Oceans Conference*, June 2017.

[13] P. Chhabra, A. Maccarone, A. McCarthy, G. Buller, and A. Wallace, "Discriminating underwater Lidar target signatures using sparse multi-spectral depth codes," in *Sensor Signal Processing for Defence (SSPD)*, pp. 1–5, Sept. 2016.

[14] A. Janowski, P. Nierebiński, M. Przyborski, and J. Szulwic, "Object detection from Lidar data on the urban area, on the basis of image analysis methods. experiments on real data," in *1st int. conf. on innovative research and maritime app. of space tech.*, May 2015.

[15] A. Wallace, C. Nichol, and I. Woodhouse, "Recovery of forest canopy parameters by inversion of multispectral Lidar data," *Remote Sensing*, vol. 4, pp. 509–531, Feb. 2012.

[16] W. Gong, J. Sun, S. Shi, J. Yang, L. Du, B. Zhu, and S. Song, "Investigating the potential of using the spatial and spectral information of multispectral Lidar for object classification," *Sensors*, vol. 15, pp. 21989–22002, Sept. 2015.

[17] B. Bright, J. A. Hicke, and A. Hudak, "Estimating aboveground carbon stocks of a forest affected by mountain pine beetle in Idaho using Lidar and multispectral imagery," *Remote Sensing of Environment*, vol. 124, pp. 270—281, Sept. 2012.

[18] J. Manzanera, A. Garc, C. Pascual, R. Gimeno, S. Mart, T. Tokola, and R. Valbuena, "Fusion of airborne Lidar and multispectral sensors reveals synergic capabilities in forest structure characterization," *GIScience & Remote Sensing*, vol. 53, Sept. 2016.

[19] G. Wei, S. Shalei, Z. Bo, S. Shuo, L. Faquan, and C. Xuewu, "Multi-wavelength canopy Lidar for remote sensing of vegetation: design and system performance," *ISPRS Phot. Remote Sensing*, vol. 69, pp. 1–9, 2012.

[20] A. Wallace, A. McCarthy, C. Nichol, X. Ren, S. Morak, D. Martinez-Ramirez, I. Woodhouse, and G. Buller, "Design and evaluation of multispectral Lidar for the recovery of arboreal parameters," *IEEE Trans. Geosci. and Remote Sensing*, Aug. 2014.

[21] X. Ren, Y. Altmann, R. Tobin, A. Mccarthy, S. Mclaughlin, and G. S. Buller, "Wavelength-time coding for multispectral 3D imaging using single-photon Lidar," *Opt. Exp.*, vol. 26, pp. 30146–30161, Nov. 2018.

[22] R. Tobin, Y. Altmann, X. Ren, A. McCarthy, R. A. Lamb, S. McLaughlin, and G. S. Buller, "Comparative study of sampling strategies for sparse photon multispectral Lidar imaging: towards mosaic filter arrays," *Journal of Optics*, vol. 19, p. 094006, Aug. 2017.

[23] Y. Altmann, A. Wallace, and S. McLaughlin, "Spectral unmixing of multispectral Lidar signals," *IEEE Trans. Signal Process.*, vol. 63, pp. 5525–5534, Oct. 2015.

[24] J. Tachella, Y. Altmann, M. Márquez, H. Arguello-Fuentes, J.-Y. Tourneret, and S. McLaughlin, "Bayesian 3D reconstruction of subsampled multispectral single-photon Lidar signals," *IEEE Trans. Comput. Imaging*, Aug. 2019.

[25] A. Halimi, R. Tobin, A. McCarthy, J. M. Bioucas-Dias, S. McLaughlin, and G. S. Buller, "Robust restoration of sparse multidimensional single-photon Lidar images," *IEEE Trans. Comput. Imaging*, 2019.

[26] Y. Altmann, R. Tobin, A. Maccarone, X. Ren, A. McCarthy, G. S. Buller, and S. McLaughlin, "Bayesian restoration of reflectivity and range profiles from subsampled single-photon multispectral Lidar data," in *25th European Signal Processing Conference (EUSIPCO)*, (Kos, Greece), pp. 1410–1414, Sept. 2017.

[27] J. Rapp and V. K. Goyal, "A few photons among many: unmixing signal and noise for photon-efficient active imaging," *IEEE Trans. Comput. Imaging*, vol. 3, pp. 445–459, Sept. 2017.

[28] A. M. Wallace, J. Ye, N. J. Krichel, A. McCarthy, R. J. Collins, and G. S. Buller, "Full waveform analysis for long-range 3D imaging laser Radar," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, p. 896708, Aug. 2010.

[29] J. Tachella, Y. Altmann, N. Mellado, A. McCarthy, R. Tobin, G. S. Buller, J.-Y. Tourneret, and S. McLaughlin, "Real-time 3d reconstruction from single-photon lidar data using plug-and-play point cloud denoisers," *Nature Communications*, vol. 10, p. 4984, Nov. 2019.

[30] Y. Altmann and S. McLaughlin, "Range estimation from single-photon Lidar data using a stochastic EM approach," in *26th European Signal Processing Conference (EUSIPCO)*, (Rome, Italy), pp. 1112–1116, Sept. 2018.

[31] Y. Altmann, A. Maccarone, A. McCarthy, G. Newstadt, G. S. Buller, S. McLaughlin, and A. Hero, "Robust spectral unmixing of sparse multispectral Lidar waveforms using gamma Markov random fields," *IEEE Trans. Comput. Imaging*, vol. 3, pp. 658–670, Dec. 2017.

[32] E. Bratsolis, M. Sigelle, and E. Charou, "Building block extraction and classification by means of Markov random fields using aerial imagery and Lidar data," in *SPIE Remote Sensing*, Oct. 2016.

[33] J. Tachella, Y. Altmann, S. McLaughlin, and J.-Y. Tourneret, "3D reconstruction using single-photon Lidar data exploiting the widths of the returns," in *Proc. IEEE ICASSP-19*, pp. 7815–7819, May 2019.

[34] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, 1992.

[35] A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Math. Image and Vision*, vol. 20, pp. 89–97, Jan. 2004.

[36] A. Kadambi and P. T. Boufounos, "Coded aperture compressive 3D Lidar," in *Proc. IEEE ICASSP-15*, pp. 1166–1170, April 2015.

[37] X. Wang, "Laplacian operator-based edge detectors," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 29, pp. 886–90, June 2007.

[38] M. Meyer, M. Desbrun, P. Schroder, and A. H. Barr, "Discrete differential-geometry operators for triangulated 2-manifolds," in *Visualization and Math. III*, pp. 35–57, Springer Berlin Heidelberg, 2003.

[39] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *Int. Conf. on Comput. Vision*, pp. 479–486, Nov. 2011.

[40] D. E. Schwartz, E. Charbon, and K. L. Shepard, "A single photon avalanche diode array for fluorescence lifetime imaging microscopy," *IEEE Journal of Solid-State Circuits*, vol. 43, pp. 2546–2557, Nov. 2008.

[41] Y. Altmann, X. Ren, A. McCarthy, G. S. Buller, and S. McLaughlin, "Lidar waveform-based analysis of depth images constructed using sparse single-photon data," *IEEE Trans. Image Processing*, vol. 25, pp. 1935–1946, May 2016.

[42] 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)*, vol. 39, no. 1, pp. 1–22, 1977.

[43] G. McLachlan and T. Krishnan, *The EM algorithm and extensions*, vol. 382. John Wiley & Sons, 2007.

[44] G. Celeux, D. Chauveau, and J. Diebolt, "Stochastic versions of the EM algorithm: an experimental study in the mixture case," *Journal of Stat. Comp. and Simulation*, vol. 55, no. 4, pp. 287–314, 1996.

[45] G. Celeux, F. Forbes, and N. Peyrard, "EM procedures using mean field-like approximations for Markov model-based image segmentation," *Pattern Recognition*, vol. 36, no. 1, pp. 131 – 144, 2003.

[46] F. Forbes and G. Fort, "Combining Monte-Carlo and mean-field-like methods for inference in hidden Markov random fields," *IEEE Trans. Image Processing*, vol. 16, pp. 824–837, March 2007.

[47] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering," *IEEE Trans. Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[48] L. Azzari and A. Foi, "Variance stabilization for noisy+estimate combination in iterative Poisson denoising," *IEEE Signal Processing letters*, vol. 23, no. 8, pp. 1086–1090, 2016.

[49] S.Boyd, N.Parikh, E.Chu, B.Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

[50] D. Benjamini, M. E. Komlosh, L. A. Holtzclaw, U. Nevo, and P. J. Basser, "White matter microstructure from nonparametric axon diameter distribution mapping," *NeuroImage*, vol. 135, pp. 333–344, 2016.

[51] K.-S. Chuang, H.-L. Tzeng, S. Chen, J. Wu, and T.-J. Chen, "Fuzzy c-means clustering with spatial information for image segmentation," *computerized Medical Imaging and Graphics*, vol. 30, no. 1, pp. 9–15, 2006.

[52] W. Zhao and S. Du, "Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4544–4554, 2016.

[53] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *IEEE Signal Processing Magazine*, vol. 31, pp. 45–54, Jan. 2014.

[54] S. Chen, A. Halimi, X. Ren, A. McCarthy, X. Su, S. McLaughlin, and G. S. Buller, "Learning non-local spatial correlations to restore sparse 3D single-photon data," *IEEE Trans. Image Processing*, vol. 29, pp. 3119–3131, 2019.

[55] A. McCarthy, N. J. Krichel, N. R. Gemmell, X. Ren, M. G. Tanner, S. N. Dorenbos, V. Zwiller, R. H. Hadfield, and G. S. Buller, "Kilometer-range, high resolution depth imaging via 1560 nm wavelength single-photon detection," *Opt. Exp.*, vol. 21, no. 7, pp. 8904–8915, 2013.

[56] N. Bouguila, D. Ziou, and J. Vaillancourt, "Unsupervised learning of a finite mixture model based on the dirichlet distribution and its application," *IEEE Trans. Image Processing*, vol. 13, no. 11, pp. 1533–1543, 2004.

[57] N. Chodosh, C. Wang, and S. Lucey, "Deep convolutional compressed sensing for lidar depth completion," in *Computer Vision – ACCV 2018* (C. V. Jawahar, H. Li, G. Mori, and K. Schindler, eds.), (Cham), pp. 499–513, Springer International Publishing, 2019.

[58] Y. Altmann, S. McLaughlin, and M. E. Davies, "Fast online 3D reconstruction of dynamic scenes from individual single-photon detection events," *IEEE Trans. Image Processing*, vol. 29, pp. 2666–2675, 2020.