Bayesian 3D Reconstruction of Subsampled Multispectral Single-Photon Lidar Signals

Julián Tachella , *Student Member, IEEE*, Yoann Altmann , *Member, IEEE*,
Miguel Márquez , *Student Member, IEEE*, Henry Arguello-Fuentes , *Senior Member, IEEE*,
Jean-Yves Tourneret, *Fellow, IEEE*, and Stephen McLaughlin , *Fellow, IEEE*

Abstract—Light detection and ranging (Lidar) single-photon devices capture range and intensity information from a three-dimensional (3-D) scene. This modality enables long range 3-D reconstruction with high range precision and low laser power. A multispectral single-photon Lidar system provides additional spectral diversity, allowing the discrimination of different materials. However, the main drawback of such systems can be the long acquisition time needed to collect enough photons in each spectral band. In this work, we tackle this problem in two ways: first, we propose a Bayesian 3-D reconstruction algorithm that is able to find multiple surfaces per pixel, using few photons, i.e., shorter acquisitions. In contrast to previous algorithms, the novel method processes jointly all the spectral bands, obtaining better reconstructions using less photon detections. The proposed model promotes spatial correlation between neighbouring points within a given surface using spatial point processes. Secondly, we account for different spatial and spectral subsampling schemes, which reduce the total number of measurements, without significant degradation of the reconstruction performance. In this way, the total acquisition time, memory requirements and computational time can be significantly reduced. The experiments performed using both synthetic and real single-photon Lidar data demonstrate the advantages of tailored sampling schemes over random alternatives. Furthermore, the proposed algorithm yields better estimates than other existing methods for multi-surface reconstruction using multispectral Lidar data.

Index Terms—Bayesian inference, 3-D reconstruction, Markov chain Monte Carlo, Lidar, multispectral imaging, low-photon imaging, poisson noise.

J. Tachella, Y. Altmann, and S. McLaughlin are with the School of Engineering, Heriot-Watt University, EH14 4AS Edinburgh, U.K. (e-mail: jat3@hw.ac.uk; y.altmann@hw.ac.uk; s.mclaughlin@hw.ac.uk).

M. Márquez is with the Department of Physics, Universidad Industrial de Santander, Bucaramanga 6344000, Colombia (e-mail: hds.marquez@gmail.com).

H. Arguello-Fuentes is with the Department of Systems Engineering, Universidad Industrial de Santander, Bucaramanga 6344000, Colombia (e-mail: henarfu@uis.edu.co).

J.-Y. Tourneret is with the Signal and Image Department, University of Toulouse, 31013 Toulouse, France (e-mail: jean-yves.tourneret@enseeiht.fr).

Digital Object Identifier 10.1109/TCI.2019.2945204

I. INTRODUCTION

SINGLE-PHOTON Lidar devices provide accurate range information by constructing, for each pixel, a histogram of time delays between emitted light pulses and photon detections. Using time correlated single-photon counting (TCSPC) technology, Lidar systems are able to provide accurate depth information (of the order of centimetres) over very long ranges, while allowing the use of laser sources of low power [1]. The acquired range information (3D structure) has many important applications, such as self-driving cars [2], the study of structures behind dense forests [3] and environmental monitoring [4]. Multispectral Lidar (MSL) systems gather measurements at many spectral bands, making it possible to distinguish distinct materials, as illustrated in Fig. 1. For example, spectral diversity was used in [5] to differentiate leaves from trunks and in [6] to estimate plant area indices and abundance profiles. The MSL modality consists of constructing one histogram of time delays per wavelength, as shown in Fig. 2. The spectral diversity can be obtained either using a supercontinuum laser source [6], [7] or multiple lasers [8]. The detector generally consists of a spatial form of wavelength routing to demultiplex the channels [6]–[8] or wavelength-to-time codification [9]. Recovering spatial and spectral information from MSL data is a challenging task, specially in scenes with strong ambient illumination (i.e., multiple spurious detections) or when the acquisition time is very low (i.e., very few photon detections per histogram). Moreover, in a general setting, it is possible to find more than one object per pixel. This phenomenon occurs in scenes where the light goes through semi-transparent materials (e.g., glass), or when the laser footprint is such that multiple surfaces appear in the field of view. Thus, several signal processing algorithms have been proposed to address these challenges: while many algorithms are available for single-wavelength Lidar, either assuming a single surface per pixel [10]–[12] or multiple surfaces per pixel [13]–[15], to the best of our knowledge, only the single-surface-per-pixel case was studied in the multispectral case [7], [16], [17]. This single-depth assumption greatly simplifies the reconstruction problem, as it significantly reduces memory requirements and overall complexity. Datasets containing dozens of wavelengths can be prohibitively large for practical 3D reconstruction algorithms, both in terms of memory and computing requirements. For example, a typical MSL hypercube with 32 wavelengths has more than 10^9 data

Fig. 1. An airborne MSL system can capture multiple objects per pixel and discriminate their materials. The multidepth capability enables the recovery of information from photons reflected off different branches of the trees, ground, pedestrians or even from the interior of a car (i.e., photons which propagated across the windshield) at an intra-pixel level. Moreover, the multispectral information allows us to discriminate different properties of the materials of each 3-D object (e.g., the leaves and trunk of a tree).

voxels. To alleviate this problem, some compressive acquisition strategies have been proposed. While TCSPC technology hinders compressive techniques along the depth axis,[1] reducing the number of measurements can be achieved by integrating multiple wavelengths in a single histogram [9] or measuring fewer histograms (i.e., subsampling) [16]. The wavelength-to-time approach proposed by Ren *et al.* [9] is not well-suited in the presence of multiple surfaces per pixel. Indeed, this method compresses L histograms (associated with L wavelengths) into a single waveform by shifting in time the photon detections according to each measured wavelength. While significantly reducing the data size, the resulting likelihood becomes highly multimodal and extremely difficult to handle in the presence of multiple surfaces. Different random subsampling schemes were studied in [16] without obtaining any significant differences in terms of reconstruction quality in the low-photon count regime.

In this work, we investigate a new pseudo-random subsampling scheme for low-photon count MSL data based on ideas from coded aperture design [19], [20]. By choosing the pixels measured for each wavelength in a more principled way, we achieve better results than the completely random schemes of Altmann *et al.* [16]. Furthermore, the proposed subsampling strategy can be easily implemented in many Lidar systems, reducing the total number of measurements, i.e., the time to acquire an MSL frame. Raster-scan systems using a laser super-continuum source [6], [7] can be easily modified to measure only a subset of pixels per wavelength. More interestingly, single-wavelength array technology [21] can be combined with coded apertures [19], which acquire different wavelengths at each pixel.

Furthermore, we propose a new method to perform 3D reconstruction from subsampled MSL data, which is capable of finding multiple surfaces per pixel. The novel method draws ideas from a recently published algorithm named ManiPoP [15], provided state-of-the-art reconstructions in the multiple surface, single-wavelength case. However, due to the significantly larger dimensionality of multispectral data, we propose to modify the

Bayesian model and estimation strategy of ManiPoP. Adopting a Bayesian framework similar to [15], we assign a spatial point process prior to promote spatial correlation and a Gaussian Markov random field prior to regularize the spectral reflectivity within surfaces/objects. Inference using the resulting posterior distribution is performed using a reversible jump Markov chain Monte Carlo algorithm (RJ-MCMC), coupled with a multiresolution approach that improves the convergence speed and reduces the total computing time of the algorithm. We introduce new RJ-MCMC proposals, which take into account the additional spectral dimension and improve the acceptance ratio, compared to the ones proposed in [15]. Moreover, we propose an empirical Bayes approach to build the prior distribution associated with the background detections, which further improves the convergence of the RJ-MCMC sampler. Contrary to multi-depth methods that require storage of dense volumetric estimates (e.g., [13], [14] in the single-wavelength case and [22] in the multi-temporal case), the memory requirements of the proposed method are minimal (just the 3D points and spectral signatures are stored in memory), enabling the acquisition and processing of very large datasets (dozens of wavelengths and hundreds of pixels).

The main contributions of this work are
- a new Bayesian 3D reconstruction algorithm for subsampled MSL data with multiple surfaces per pixel
- the analysis of an appropriate subsampling scheme based on coded aperture design, which provides better results than completely random alternatives.

The new algorithm is referred to as MuSaPoP, as it models MultiSpectrAl Lidar signals using POint Processes. The remainder of the paper is organized as follows. Section II recalls the classical observation model for single-photon MSL data. Sections III and IV present the Bayesian 3D reconstruction algorithm and the associated RJ-MCMC sampler. Section V details the principled subsampling strategy. Experiments performed with synthetic and real Lidar data are introduced and discussed in Section VI. Conclusions and future work are finally reported in Section VII.

II. SINGLE-PHOTON MULTISPECTRAL LIDAR

A full multispectral Lidar data hypercube $\boldsymbol{Z} \in \mathbb{Z}_+^{N_r \times N_c \times L \times T}$ consists of discrete photon count measurements, where $\mathbb{Z}_+ = \{0, 1, \dots\}$ is the set of positive integers, N_r and N_c are the numbers of vertical and horizontal pixels respectively, L is the number of acquired wavelengths and T is the histogram length (i.e., the number of time bins). The 3D reconstruction algorithm estimates a point cloud, referred to as $\boldsymbol{\Phi}$, from the data hypercube $\boldsymbol{Z}$. The 3D point cloud is represented by an unordered set of points, that is

$$\boldsymbol{\Phi} = \{(\boldsymbol{c}_n, \boldsymbol{r}_n), \ n = 1, \dots, N_\Phi\} \quad (1)$$

where N_Φ is the total number of points, and $\boldsymbol{c}_n = [x_n, y_n, t_n]^T \in [1, N_r] \times [1, N_c] \times [1, T] \subset \mathbb{Z}_+^3$ and $\boldsymbol{r}_n = [r_{n,1}, \dots, r_{n,L}]^T \in \mathbb{R}_+^L$ are the coordinate vector and the spectral response of the nth point, respectively. The observed photon count at pixel (i, j), bin t and spectral band ℓ follows a Poisson distribution [16], whose intensity is a mixture of the background level, denoted by $b_{i,j,\ell}$, and the responses of the surfaces present in
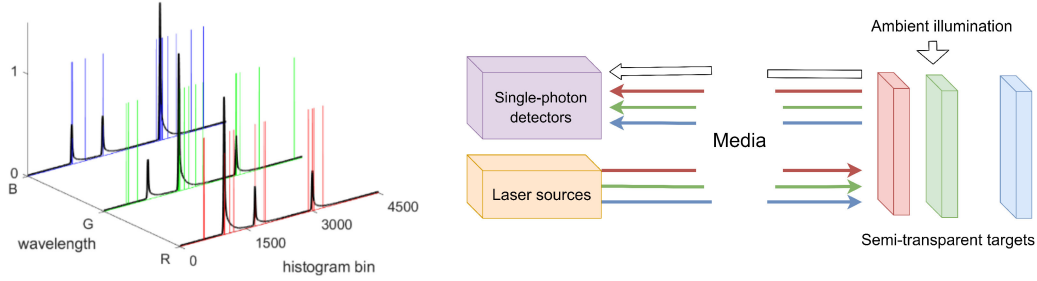
Fig. 2. Example of an MSL system with three different wavelengths (red, green and blue). On the right, a schematic shows the working principles of a single-photon multispectral Lidar device. The red, green and blue arrows illustrate the laser pulses sent by the laser sources and reflected by the target onto the single-photon detectors. The white arrow depicts the background photons emitted by an ambient illumination source that reach the detectors at random times. The figure on the left shows the collected histograms for a given pixel: the discrete measurements are depicted in red, green and blue, while the underlying Poisson intensity (i.e., the parameters to estimate from the data) is shown in black.

that pixel, i.e.,

$$z_{i,j,\ell,t}|\mathbf{\Phi}, b_{i,j,\ell} \sim \mathcal{P}\left(g_{i,j,\ell}\left(\sum_{\mathcal{N}_{i,j}} r_{n,\ell} h_\ell(t - t_n) + b_{i,j,\ell}\right)\right)$$

where $\mathcal{P}(\cdot)$ denotes the Poisson distribution, $g_{i,j,\ell} \in \{0,1\}$ is a known binary variable that indicates whether wavelength ℓ at pixel (i,j) has been acquired/observed or not, $\mathcal{N}_{i,j} = \{n : (x_n, y_n) = (i,j)\}$ is the set of points corresponding to pixel (i,j) and wavelength ℓ, and $h_\ell(t)$ is the impulse response of the Lidar device at wavelength ℓ, which can be measured using a spectralon panel during a calibration step. Note that to lighten notations, we assume that the impulse responses are only wavelength-dependent but the algorithm can be easily adapted when the responses vary among pixels. Assuming mutual independence between noise realizations in different bins, pixels and spectral bands [1], the likelihood of the proposed model can be written as

$$p(\mathbf{Z}|\mathbf{\Phi}, \mathbf{B}) = \prod_{i=1}^{N_c}\prod_{j=1}^{N_r}\prod_{\ell=1}^{L}\prod_{t=1}^{T} p(z_{i,j,\ell,t}|\mathbf{\Phi}, b_{i,j,\ell}). \quad (2)$$

For clarity in the notation, we will also denote the set of point coordinates as $\mathbf{\Phi}_c = \{\boldsymbol{c}_n, n = 1, \ldots, N_\Phi\}$ and the set of spectral responses as $\mathbf{\Phi}_r = \{\boldsymbol{r}_n, n = 1, \ldots, N_\Phi\}$. The set of all background levels is denoted by $\mathbf{B} = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_L] \in \mathbb{R}_+^{N_r \times N_c \times L}$, which is the concatenation of L images $\boldsymbol{b}_\ell$, one for each wavelength. The cube of binary measurements is designated by $\boldsymbol{G} \in \{0,1\}^{N_r \times N_c \times L}$, where $[\boldsymbol{G}]_{i,j,\ell} = g_{i,j,\ell}$. Note that the model used in the ManiPoP algorithm [15] can be obtained from (2) by setting all the binary variables to 1, and considering only one band, i.e., $L = 1$.

III. MULTIPLE-RETURN MULTIPLE-WAVELENGTH 3-D RECONSTRUCTION

Recovering the position of the objects $\boldsymbol{c}_n$, their spectral signatures $\boldsymbol{r}_n$ and background levels $\boldsymbol{B}$ from the subsampled MSL data $\boldsymbol{Z}$ is an ill-posed inverse problem, as many solutions can explain the observed photon counts. Thus, prior regularization is necessary to promote reconstructions in a set of feasible 3D point clouds. In this work, we adopt a Bayesian framework, which allows us to include prior knowledge about the scene through tailored prior distributions assigned to the parameters of interest.

A. Prior Distributions

The proposed model considers prior regularization for the point positions and reflectivity. As explained in Section III-A3, an empirical Bayes prior [23] is assigned to the background levels.

1) Spatial Configuration: We adopt the spatial prior distribution of 3D points developed in the ManiPoP algorithm. This distribution is designed to promote spatial correlation between points within one surface and repulsion between points belonging to different surfaces. While only a brief summary is included below, we refer the reader to [15] for a detailed discussion about this prior model. The spatial point process prior for the position of the points is modelled by a density defined with respect to a Poisson point process reference measure [24, Chapter 9], i.e.,

$$f(\mathbf{\Phi}_c) \propto f_1(\mathbf{\Phi}_c) f_2(\mathbf{\Phi}_c|\gamma_a, \lambda_a)$$

where $f_1(\mathbf{\Phi}_c)$ and $f_2(\mathbf{\Phi}_c|\gamma_a, \lambda_a)$ are the Strauss and area interaction [25] processes respectively. The repulsive Strauss process is written as

$$f_1(\mathbf{\Phi}_c) \propto \begin{cases} 0 & \text{if } \exists n \neq n' : x_n = x_{n'}, y_n = y_{n'} \\ & \text{and } |t_n - t_{n'}| < d_{\min} \\ 1 & \text{otherwise} \end{cases}$$

where $d_{\min}$ is the minimum separation between two surfaces in the same pixel. Attraction between points within the same surface is promoted by the area interaction process, that is

$$f_2(\mathbf{\Phi}_c|\gamma_a, \lambda_a) \propto \lambda_a^{N_\Phi} \gamma_a^{-m\left(\bigcup_{n=1}^{N_\Phi} S(\boldsymbol{c}_n)\right)} \quad (3)$$

where $m(\cdot)$ denotes the standard Lebesgue measure, $S(\boldsymbol{c}_n)$ defines a convex set around the point $\boldsymbol{c}_n$, and γ_a and λ_a are two hyperparameters, accounting for the amount of attraction and total number of points, respectively. Both densities define Markovian interactions between points, only correlating points in a local neighbourhood. Moreover, the combination of both processes implicitly defines a connected-surface structure, which is used to model 2D manifolds in a 3D space.

2) Reflectivity: The spectral signatures are related to the materials of the surfaces [17]. Neighbouring points corresponding to a surface composed of a specific material show similar spectral signatures. This prior information is added to the model using a Gaussian Markov random field distribution, where the neighbours are defined by the connected-surface structure of the point
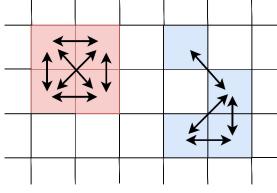
Fig. 3. Connectivity at an inter-pixel level. Two different surfaces are denoted by the colours red and blue, where each square represents one pixel. Pixels without points are represented by white squares. For simplicity, in this example all points are considered to be present at the same depth. Note that each pixel can be connected with at most eight neighbours.
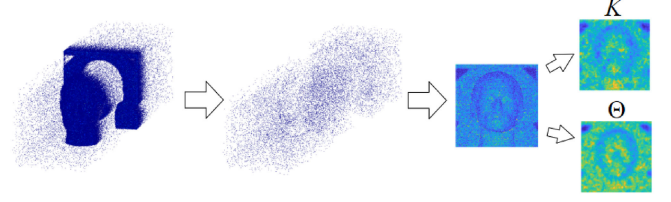

Fig. 4. Computation of the hyperparameters for the priors of the background levels. First, the photons due to the signal are removed from the dataset using a coarse approximation of the point cloud. Secondly, the remaining photons are integrated per pixel, giving a noisy background image. Finally, this image is used to estimate uncertainty about the background levels, computing the gamma hyperparameters $\boldsymbol{K}$ and $\boldsymbol{\Theta}$.

process prior. First, in order to avoid the positivity constraint on the intensity $r_{n,\ell}$, we use the canonical form [26]–[28],

$$m_{n,\ell} = \log(r_{n,\ell}) \tag{4}$$

where $m_{n,\ell} \in \mathbb{R}$ denotes the log-intensity of the nth point at band ℓ. As multispectral devices only acquire dozens of well-separated wavelengths, the spectral measurements within a pixel do not show significant correlation. Hence, although potential correlations between wavelengths could be modelled, we choose here to neglect this correlation to keep the estimation strategy tractable. As a consequence, we consider the following reflectivity prior model

$$p(\boldsymbol{\Phi}_r|\sigma^2, \beta) = \prod_{\ell=1}^{L} p(\boldsymbol{m}_\ell|\boldsymbol{\Phi}_c, \sigma^2, \beta). \tag{5}$$

Spatial correlations between log-intensity values in neighbouring pixels are defined according to the distribution

$$\boldsymbol{m}_\ell|\sigma^2, \beta, \boldsymbol{\Phi}_c \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{P}^{-1}) \tag{6}$$

where σ^2 and β are hyperparameters controlling the level of smoothness. The precision matrix $\boldsymbol{P}$ is very sparse due to the Markovian structure, being defined by

$$[\boldsymbol{P}]_{n,n'} = \begin{cases} \beta + \sum_{\tilde{n} \in \mathcal{M}_{pp}(\boldsymbol{c}_n)} \frac{1}{d(\boldsymbol{c}_n; \boldsymbol{c}_{\tilde{n}})} & \text{if } n = n' \\ -\frac{1}{d(\boldsymbol{c}_n; \boldsymbol{c}_{n'})} & \text{if } \boldsymbol{c}_n \in \mathcal{M}_{pp}(\boldsymbol{c}_{n'}) \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where $\mathcal{M}_{pp}(\boldsymbol{c}_n)$ is the set of neighbours of point $\boldsymbol{c}_n$, which is obtained using the connected-surface structure illustrated in Fig. 3, and $d(\boldsymbol{c}_n; \boldsymbol{c}_{\tilde{n}})$ is the Euclidean distance between two points, normalized according to the camera parameters of the scene to have a physical meaning [29].

3) Background Levels: Background detections are due to detector dark counts and ambient illumination, as explained in [13], [30]. If the transceiver system is mono-static[2] [1], the set of background levels can be interpreted as a multispectral passive image of the scene, as background detections generally come from ambient illumination reflected onto the target and collected by the single-photon detector. In this case, the background levels are strongly spatially correlated within each wavelength. However, in bi-static systems [13], the transmit and receive channels do not share the same objective lens aperture, yielding

---

[2]In mono-static Lidar systems, the laser and detector are coaxial, whereas in bi-static systems, the source and detector do not share the same axis.

weakly or uncorrelated background detections. Although not showing a strong spatial correlation in this second case, all the background levels have similar values, which also serves as prior information.

a) Independent Prior Distributions: In order to simultaneously model potential spatial correlation and ensure the positivity of the background levels, ManiPoP uses a 2D gamma Markov random field, which was introduced by Dikmen *et al.* in [31]. However, this prior is not well suited for MSL data as it introduces an undesired penalization for large background levels, whose negative effects are amplified when considering multiple bands (see Appendix A for details). Other alternatives such as Gaussian Markov random fields [26] cannot be sampled directly in closed form, requiring proposals with a rejection step, whose mixing and convergence scale badly with the dimension of the spectral cube, as shown in [32]. To alleviate these problems, we assign independent gamma priors, i.e.,

$$\begin{cases} p(\boldsymbol{B}|\boldsymbol{K}, \boldsymbol{\Theta}) = \prod_{i=1}^{N_r} \prod_{j=1}^{N_c} \prod_{\ell=1}^{L} p(b_{i,j,\ell}|k_{i,j,\ell}, \theta_{i,j,\ell}) \\ b_{i,j,\ell}|k_{i,j,\ell}, \theta_{i,j\ell} \sim \mathcal{G}(k_{i,j,\ell}, \theta_{i,j\ell}) \end{cases} \tag{8}$$

where $[\boldsymbol{\Theta}]_{i,j,\ell} = \theta_{i,j,\ell}$ and $[\boldsymbol{K}]_{i,j,\ell} = k_{i,j,\ell}$ are the shape and scale hyperparameters of the gamma distributions. Despite using independent priors, we can capture the spatial correlation by setting the hyperparameters $(\boldsymbol{K}, \boldsymbol{\Theta})$ appropriately. More precisely, in a similar fashion to variational Bayes [33] or expectation propagation [34] methods, in order to simplify the estimation of $\boldsymbol{B}$, we specify (8) such that $p(\boldsymbol{B}|\boldsymbol{K}, \boldsymbol{\Theta})$ is similar to another distribution $q(\boldsymbol{B}) = \prod_{\ell=1}^{L} q_\ell(\boldsymbol{b}_\ell)$ which explicitly correlates the background levels in neighbouring pixels and assumes mutual independence between spectral bands. Here, we use as a similarity criterion the Kullback-Leibler divergence

$$(\boldsymbol{K}, \boldsymbol{\Theta}) = \underset{\boldsymbol{K}, \boldsymbol{\Theta}}{\arg\min} \text{KL}(q||p). \tag{9}$$

As discussed in Appendix B, solving (9) can be achieved by computing expectations with respect to $q(\boldsymbol{B})$.

b) Empirical Bayes approach: To ensure that the prior model (8) is informative, a suitable distribution $q(\boldsymbol{B})$ should be chosen. Assuming that we have a coarse estimate of the point cloud (this information will be obtained using the multiresolution approach detailed in Section IV-B), we build the distribution $q(\boldsymbol{B})$ following an empirical Bayes approach, as illustrated in Fig. 4. First, one can discard almost all the signal photons in the dataset by removing the photons detected in the compact
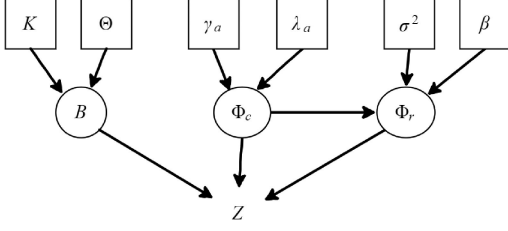
Fig. 5. Directed acyclic graph (DAG) of the proposed hierarchical Bayesian model. The variables inside squares are fixed, whereas the variables inside circles are estimated.

support of $h_\ell(t)$ around each point (see Fig. 4, central subplot). The number of bins that is not excluded in each pixel is referred to as $v_{i,j,\ell}$. Secondly, we integrate the remaining photons of each pixel, obtaining a coarse estimate of the per-pixel background photon levels, denoted by $\bar{z}_{i,j,\ell,b}$. We then define $q_\ell(\boldsymbol{b}_\ell) \propto p(\bar{\boldsymbol{z}}_\ell|\boldsymbol{b}_\ell)p(\boldsymbol{b}_\ell|\alpha_B)$ with

$$
\begin{cases}
\bar{z}_{i,j,\ell,b}|b_{i,j,\ell} \sim \mathcal{P}(g_{i,j,\ell}v_{i,j,\ell}b_{i,j,\ell}) \\
b_{i,j,\ell} = \exp(\tilde{b}_{i,j,\ell} + \mu_\ell) \\
\tilde{\boldsymbol{b}}_\ell|\alpha_B \sim \mathcal{N}(\boldsymbol{0}, \alpha_B \boldsymbol{D}^{-1})
\end{cases} \tag{10}
$$

where $\boldsymbol{b}_\ell$ a vectorized image of background levels at wavelength ℓ, $\boldsymbol{D} \in \mathbb{R}^{NrNc \times NrNc}$ a positive semidefinite matrix, α_B is a fixed hyperparameter controlling the degree of smoothness. In mono-static systems, a two-dimensional Laplacian filter is chosen for $\boldsymbol{D}$ to promote spatial correlation [26], whereas in bi-static systems, $\boldsymbol{D}$ is replaced by the identity matrix, only penalizing large background levels. μ_ℓ is a translation parameter centring $\tilde{b}_{i,j,\ell}$ in the linear part of the exponential function and is defined as $\mu_\ell = \log(\frac{1}{N_r N_c} \sum_i^{N_r} \sum_j^{N_c} \frac{\bar{z}_{i,j,\ell,b} g_{i,j,\ell}}{v_{i,j,\ell}})$. As mentioned above, solving (9) requires the computation of expectations with respect to $q(\boldsymbol{B})$ which are unfortunately not available in closed form. Instead of using additional MCMC sampling to find numerical approximations (detailed in Appendix B), obtaining samples from (10) is more attractive both in terms of convergence and computational complexity than using the original full datacube which includes mixtures of background and signal photons. Indeed, (10) simply involves integrated photon counts (over the range dimension). Moreover, given the independence property of $q(\boldsymbol{B})$ among spectral bands, all the bands can be processed independently in parallel when sampling $\boldsymbol{B}$ (see Section IV-A5).

B. Posterior Distribution

Following Bayes theorem, the joint posterior distribution of the model parameters is given by

$$
p(\boldsymbol{\Phi}_c, \boldsymbol{\Phi}_r, \boldsymbol{B}|\boldsymbol{Z}, \Psi)
$$
$$
\propto p(\boldsymbol{Z}|\boldsymbol{\Phi}_c, \boldsymbol{\Phi}_r, \boldsymbol{B})p(\boldsymbol{\Phi}_r|\boldsymbol{\Phi}_c, \sigma^2, \beta)
$$
$$
\times f_1(\boldsymbol{\Phi}_c|\gamma_a, \lambda_a)f_2(\boldsymbol{\Phi}_c|\gamma_{st})\pi(\boldsymbol{\Phi}_c)p(\boldsymbol{B}|\boldsymbol{K}, \boldsymbol{\Theta}) \tag{11}
$$

where Ψ denotes the set of hyperparameters $\Psi = \{\gamma_a, \lambda_a, \gamma_s, \sigma^2, \beta, \boldsymbol{K}, \boldsymbol{\Theta}\}$ and $\pi(\cdot)$ is the Poisson point process reference measure. Fig. 5 shows the directed acyclic graph associated with the proposed hierarchical Bayesian model.

IV. INFERENCE

In this work, we compute the same posterior statistics as in [15]: the point cloud positions and spectral signatures are estimated using the maximum-a-posteriori (MAP) estimator

$$
\hat{\boldsymbol{\Phi}} = \arg\max_{\boldsymbol{\Phi}} p(\boldsymbol{\Phi}, \boldsymbol{B}|\boldsymbol{Z}, \Psi) \tag{12}
$$

and the minimum mean squared error estimator is considered for $\boldsymbol{B}$

$$
\hat{\boldsymbol{B}} = \mathbb{E}\{\boldsymbol{B}|\boldsymbol{Z}, \Psi\}. \tag{13}
$$

As this expectation cannot be derived analytically, we propose to investigate Markov Chain Monte Carlo (MCMC) simulation methods. As in [15], we use a reversible jump MCMC algorithm that can handle the varying dimension nature of the spatial point process. This sampler generates N_m samples of $\boldsymbol{\Phi}$ and $\boldsymbol{B}$ from the posterior distribution (11) denoted as

$$
\{\boldsymbol{\Phi}^{(s)}, \boldsymbol{B}^{(s)} \quad \forall s = 0, 1, \ldots, N_m - 1\}. \tag{14}
$$

These samples are then used to approximate the statistics of interest, i.e.,

$$
\hat{\boldsymbol{\Phi}} \approx \arg\max_{s=0,\ldots,N_m-1} p(\boldsymbol{\Phi}^{(s)}, \boldsymbol{B}^{(s)}|\boldsymbol{Z}, \Psi)
$$

$$
\hat{\boldsymbol{B}} \approx \frac{1}{N_m - N_{\text{bi}}} \sum_{s=N_{\text{bi}}+1}^{N_m} \boldsymbol{B}^{(s)}
$$

where N_{bi} is the number of burn-in iterations.

A. Reversible Jump MCMC Moves

While other stochastic simulation algorithms for varying dimensions can be used [24], we choose an RJ-MCMC sampler, as it allows us to build proposals tailored for the MSL reconstruction problem. RJ-MCMC can be interpreted as a natural extension of the Metropolis-Hastings algorithm for problems with an a priori unknown dimensionality. Given the actual state of the chain $\boldsymbol{\theta} = \{\boldsymbol{\Phi}, \boldsymbol{B}\}$ of model order N_Φ, a random vector of auxiliary variables $\boldsymbol{u}$ is generated to create a new state $\boldsymbol{\theta}' = \{\boldsymbol{\Phi},' \boldsymbol{B}'\}$ of model order $N_{\Phi'}$, according to an appropriate deterministic function $\boldsymbol{\theta}' = g(\boldsymbol{\theta}, \boldsymbol{u})$. To ensure reversibility, an inverse mapping with auxiliary random variables $\boldsymbol{u}'$ has to exist such that $\boldsymbol{\theta} = g^{-1}(\boldsymbol{\theta},' \boldsymbol{u}')$. The move $\boldsymbol{\theta} \to \boldsymbol{\theta}'$ is accepted or rejected with probability $\rho = \min\{1, r(\boldsymbol{\theta}, \boldsymbol{\theta}')\}$, where $r(\cdot, \cdot)$ is defined as

$$
r(\boldsymbol{\theta}, \boldsymbol{\theta}') = \frac{p(\boldsymbol{\theta}'|\boldsymbol{Z}, \Psi)K(\boldsymbol{\theta}|\boldsymbol{\theta}')p(\boldsymbol{u}')}{p(\boldsymbol{\theta}|\boldsymbol{Z}, \Psi)K(\boldsymbol{\theta}'|\boldsymbol{\theta})p(\boldsymbol{u})} \left|\frac{\partial g(\boldsymbol{\theta}, \boldsymbol{u})}{\partial(\boldsymbol{\theta}, \boldsymbol{u})}\right| \tag{15}
$$

and $K(\boldsymbol{\theta}'|\boldsymbol{\theta})$ is the probability of proposing the move $\boldsymbol{\theta} \to \boldsymbol{\theta}'$, $p(\boldsymbol{u})$ is the probability distribution of the random vector $\boldsymbol{u}$, and $\left|\frac{\partial g(\boldsymbol{\theta}, \boldsymbol{u})}{\partial(\boldsymbol{\theta}, \boldsymbol{u})}\right|$ is the Jacobian of the mapping $g(\cdot)$. The RJ-MCMC algorithm performs birth/death, dilation/erosion, spatial and mark shifts, and split/merge moves with probabilities p_{birth}, $1 - p_{\text{birth}}$, p_{dilation}, $1 - p_{\text{dilation}}$, p_{shift}, p_{mark}, p_{split} and $1 - p_{\text{split}}$ respectively. Due to the Markovian nature of the prior distributions, all the proposed moves are *local*, having a complexity proportional to the size of the neighbourhood. These moves are detailed in the following subsections. For ease of presentation, we summarize the main aspects of each move, inviting the reader to consult Appendix C for more details.

1) Birth and Death Moves: The birth move proposes a new point $\boldsymbol{\theta}' = (\boldsymbol{c}_{N_\Phi+1}, \boldsymbol{m}_{N_\Phi+1})$ uniformly at random in the 3D cube. The spectral signature of the new point is proposed by extracting a fraction $(1 - u_\ell)$ from the current value of the background level $b_{i,j,\ell}$ according to the signal-to-background-ratio (SBR) [11] w_ℓ, that is for each wavelength ℓ

$$\begin{cases} u_\ell \sim \mathcal{U}(0,1), b'_{i,j,\ell} = u_\ell b_{i,j,\ell} \\ e^{m_{N_\Phi+1,\ell}} = w_\ell(1 - u_\ell)b_{i,j,\ell} \dfrac{T}{\sum_{t=1}^T h_\ell(t)} \end{cases}, \quad (16)$$

where $\mathcal{U}(0,1)$ denotes the uniform distribution on the interval $(0, 1)$. The death move proposes the removal of a point. In contrast to the birth move, we modify the background level according to

$$b'_{i,j,\ell} = b_{i,j,\ell} + e^{m_{N_\Phi+1,\ell}} \frac{\sum_{t=1}^T h_\ell(t)}{w_\ell T} \quad \forall \ell = 1, \ldots, L. \quad (17)$$

2) Dilation and Erosion Moves: Birth moves have low acceptance ratio, as the probability of randomly proposing a point within or close to the surfaces of interest is very low. However, this problem can be overcome by using the current estimation of the surface to propose in regions of high probability. The dilation move proposes a point inside the neighbourhood of an existing surface with uniform probability across all possible neighbouring positions where a point can be added. Contrary to [15], where the intensity samples are generated according to the prior distribution, the spectral signature is sampled in the same way as the birth move (16). The complementary move (named erosion), proposes to remove a point $\boldsymbol{c}_n$ with one or more neighbours. In this case, the background is updated in the same way as the death move.

3) Mark and shift moves: As in ManiPoP, the mark move updates the log-intensity of a randomly chosen point $\boldsymbol{c}_n$. Each wavelength is updated independently using a Gaussian proposal with variance δ_m as a proposal (also known as Metropolis Gaussian random walk), that is

$$m'_{n,\ell} \sim \mathcal{N}(m_{n,\ell}, \delta_m) \quad \forall \ell = 1, \ldots, L. \quad (18)$$

Similarly, the shift move updates the position of a uniformly chosen point using a Gaussian proposal with variance δ_t

$$t'_n \sim \mathcal{N}(t_n, \delta_t) \quad (19)$$

The values of δ_m and δ_t are adjusted by cross-validation[3] to yield an acceptance ratio close to 41% for each move, which is the optimal value for a one dimensional Metropolis random walk, as explained in [24, Chapter 4].

4) Split and Merge Moves: Some pixels might present two points with overlapping impulse responses in depth. In such cases, a death move followed by two birth moves would happen with very low probability. Hence, as in ManiPoP, we propose a split move, which randomly picks a point $(\boldsymbol{c}_n, \boldsymbol{m}_n)$ and proposes two new points, $(\boldsymbol{c}'_{k_1}, \boldsymbol{m}'_{k_1})$ and $(\boldsymbol{c}'_{k_2}, \boldsymbol{m}'_{k_2})$. The log-intensity is proposed for each wavelength following the mapping

$$\begin{cases} u_\ell \sim \mathcal{B}(\eta, \eta) \\ m'_{k_1,\ell} = m_{n,\ell} + \log(u_\ell) \\ m'_{k_2,\ell} = m_{n,\ell} + \log(1 - u_\ell) \end{cases} \quad (20)$$

[3]Intensities are normalized to belong to a fixed interval across datasets. Hence, we can fix the variance of the proposal to achieve similar acceptance ratios.

where $\mathcal{B}(\cdot)$ denotes the beta distribution and η is a fixed parameter. The new positions are determined according to

$$\begin{cases} s_\ell \sim \mathcal{B}e(0.5) \\ \Delta \sim \mathcal{U}(d_{\min}, L_h) \\ t'_{k_1} = t_n + (-1)^{s_\ell}\Delta \dfrac{\sum_{\ell=1}^L (1 - u_\ell)e^{m_{n,\ell}}}{\sum_{\ell=1}^L e^{m_{n,\ell}}} \\ t'_{k_2} = t_n - (-1)^{s_\ell}\Delta \dfrac{\sum_{\ell=1}^L u_\ell e^{m_{n,\ell}}}{\sum_{\ell=1}^L e^{m_{n,\ell}}} \end{cases} \quad (21)$$

where $\mathcal{B}e(\cdot)$ denotes the Bernoulli distribution, L_h is the length in bins of the instrumental response at the wavelength with longest impulse response. The complementary move, named merge move, is performed by randomly choosing two points $\boldsymbol{c}_{k_1}$ and $\boldsymbol{c}_{k_2}$ inside the same pixel ($x_{k_1} = x_{k_2}$ and $y_{k_1} = y_{k_2}$) that satisfy the condition

$$d_{\min} < |t_{k_1} - t_{k_2}| \leq L_h \quad \forall \ell = 1, \ldots, L \quad (22)$$

The merged point $(\boldsymbol{c}'_n, \boldsymbol{r}'_n)$ is obtained by the inverse mapping

$$\begin{cases} e^{m'_{n,\ell}} = e^{m_{k_1,\ell}} + e^{m_{k_2,\ell}} \quad \forall \ell = 1, \ldots, L \\ t'_n = t_{k_1} \dfrac{\sum_{\ell=1}^L e^{m_{k_1,\ell}}}{\sum_{\ell=1}^L e^{m_{k_1,\ell}} + e^{m_{k_2,\ell}}} + t_{k_2}\dfrac{\sum_{\ell=1}^L e^{m_{k_2,\ell}}}{\sum_{\ell=1}^L e^{m_{k_1,\ell}} + e^{m_{k_2,\ell}}} \end{cases}$$

which preserves the total pixel intensity and weights the spatial shift of each peak according to the sum of the intensity values.

5) Sampling the Background: In order to exploit the conjugacy between the Poisson likelihood and gamma priors for the background levels, we use a data augmentation scheme as in [35], which classifies each photon-detection according to the source (target(s) or background), i.e.,

$$z_{i,j,t,\ell} = \sum_{n:(x_n,y_n)=(i,j)} \tilde{z}_{i,j,t,\ell,n} + \tilde{z}_{i,j,t,\ell,b}$$

$$\tilde{z}_{i,j,t,\ell,b} \sim \mathcal{P}(g_{i,j,\ell}b_{i,j,\ell})$$

$$\tilde{z}_{i,j,t,\ell,n} \sim \mathcal{P}(g_{i,j,\ell}r_{n,\ell}h(t - t_n))$$

where $\tilde{z}_{i,j,t,\ell,n}$ are the photons at bin t associated with the nth surface and $\tilde{z}_{i,j,t,\ell,b}$ are the ones associated with the background. In the augmented space defined by $(\tilde{z}_{i,j,t,\ell,n}, \tilde{z}_{i,j,t,\ell,b})$, the background levels are sampled as follows

$$\begin{cases} \tilde{z}_{i,j,t,\ell,b} \sim \mathcal{B}\left(z_{i,j,l,t}, \dfrac{b_{i,j,\ell}}{\sum_{n:(x_n,y_n)=(i,j)} r_{n,\ell}h(t-t_n) + b_{i,j,\ell}}\right) \\ b_{i,j,\ell} \sim \mathcal{G}\left(r_{i,j,\ell} + \sum_{t=1}^T \tilde{z}_{i,j,t,\ell,b}, \dfrac{\theta_{i,j,\ell}}{g_{i,j,\ell}T\theta_{i,j,\ell}+1}\right) \end{cases}$$
$$(23)$$

where $\mathcal{B}(\cdot)$ denotes the Binomial distribution. The transition kernel defined by (23) produces samples of $b_{i,j,\ell}$ distributed according to the marginal posterior distribution of $\boldsymbol{B}$. In practice, we observed that only one iteration of this kernel is sufficient.

B. Full Algorithm

We adopt a multi-resolution approach to speed up the convergence of the RJ-MCMC algorithm, in a fashion similar to [15]. The dataset is downsampled by integrating photon detections in patches of $N_{\text{bin}} \times N_{\text{bin}}$ pixels. Hence, the number of pixels is

---

Algorithm 1: Multiresolution MuSaPoP.

Input: MSL waveforms $\boldsymbol{Z}$, hyperparameters Ψ, window size N_{bin} and number of scales K.

Initialization:

$\boldsymbol{\Phi}_1^{(0)} \leftarrow \emptyset$

$\boldsymbol{B}_1^{(0)} \leftarrow$ sample from (23)

$(\boldsymbol{K}_1, \boldsymbol{\Theta}_1) \leftarrow$ non-informative hyperparameter values

Main loop:

for $k = 1, \dots, K$ **do**

  if $k > 1$ **then**

    $(\boldsymbol{\Phi}_k^{(0)}, \boldsymbol{B}_k^{(0)}) \leftarrow$ upsample$(\hat{\boldsymbol{\Phi}}_{k-1}, \hat{\boldsymbol{B}}_{k-1})$

    Compute hyperparameters $(\boldsymbol{K}_k, \boldsymbol{\Theta}_k)$ and SBR using Section III-A3

  end if

  $(\hat{\boldsymbol{\Phi}}_k, \hat{\boldsymbol{B}}_k) \leftarrow$MuSaPoP$(\boldsymbol{Z}_k, (\boldsymbol{\Phi}_k^{(0)}, \boldsymbol{B}_k^{(0)}), \Psi_k, \text{SBR})$

end for

Output: $(\hat{\boldsymbol{\Phi}}_K, \hat{\boldsymbol{B}}_K)$

---

reduced by a factor of N_{bin}^2, meaning less points and background levels to infer with N_{bin}^2 times more photons per pixel. The estimated point cloud at the coarse scale is upsampled using a simple nearest neighbour algorithm and used as initialization for the next (finer) scale. In all our experiments we repeat the process for $K = 3$ scales. The background hyperpriors $\boldsymbol{K}$ and $\boldsymbol{\Theta}$ are initialized with non-informative values, i.e., $k_{i,j,\ell} = 0.01$ and $\theta_{i,j,\ell} = 100$ for all pixels (i, j) and wavelengths ℓ. In finer scales, these hyperparameters are computed using the algorithm detailed in Section III-A3. The multi-resolution approach is finally summarized in Algorithm 1.

V. SUBSAMPLING STRATEGY

Despite not being able to design compressive measurements along the depth axis, we can still reduce the number of measurements in the two spatial (horizontal and vertical) dimensions and in the spectral dimension [16]. Given the point positions, recovering their reflectivity profile reduces to a multispectral image restoration problem using measured data corrupted by Poisson noise. While many compressive sensing strategies have been proposed for measurements under this noise assumption [36]–[38], MSL datasets have an additional limitation if multiple surfaces per pixel are considered: photon-detections belonging to different wavelengths cannot be integrated into a single histogram, as the reconstruction problem generally becomes highly non-convex, preventing practical reconstruction algorithms.[4] Indeed, summing histograms associated with different wavelengths and including multiple peaks generates histograms with even more peaks (possibly overlapping), which makes the 3D reconstruction and the reflectivity estimation more difficult. As a consequence, we only consider subsampling of depth histograms, which incorporates all of the practical sampling limitations. Following the formulation of the observation model (2), the subsampling strategy consists of choosing the binary

---

[4]As mentioned in the introduction, the system presented in [9] considers the integration of photons belonging to different histograms, but is limited to one surface per pixel.
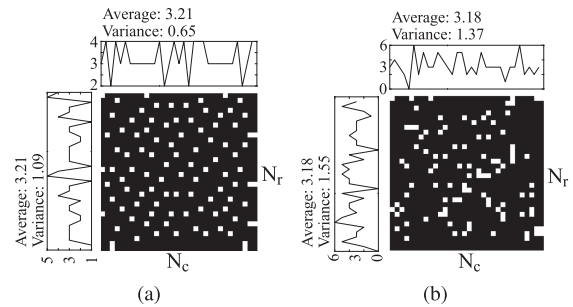

Fig. 6. Subsampling strategies for a Lidar cube with $L = 8$ wavelengths, $N_r = N_c = 32$ pixels and total compression of $1/L$, i.e., one observed band per pixel. The sampled pixels at the first wavelength are shown in white. A completely random strategy [16] is shown in (b), whereas the one proposed here is shown in (a).

coefficients $\boldsymbol{G}$ for a given compression level W/L, with W the average number of observed band per pixel. Several subsampling strategies have been proposed for different applications, such as halftoning and stippling [39], rendering, compressive spectral imaging [40]–[42], compressive computed tomography [43], [44], geometry processing [45], amongst others [46]–[48]. These approaches exploit the sampling geometry of the sensing devices to design a set of criteria. Similarly to coded aperture snapshot spectral imaging systems, the distribution of reflectivity profiles of 3D surfaces in real natural scenes suggests uniform sampling in the row, column and spectral axes. Following the design in Correa *et al.* [20], the coefficients are chosen according to the spatiotemporal characteristics of blue noise, which distributes the measurements in spectral and spatial dimensions as homogeneously as possible [49]. The binary cube $\boldsymbol{G}$ is obtained by minimizing the variance of (weighted) measurements per local neighbourhood, i.e.,

$$\arg\min_{\boldsymbol{G}} \text{VAR}\left\{ \sum_{\ell=1}^{L} \sum_{(i,'j') \in \mathcal{M}_{ss}^{(i,j)}} w_{i,'j'} g_{i,'j',\ell} \right\}$$

$$\text{subject to } \sum_{\ell=1}^{L} g_{i,j,\ell} = W \quad \forall(i,j)$$

where $\text{VAR}\{\cdot\}$ denotes the variance operator, $\mathcal{M}_{ss}^{(i,j)}$ denotes the set of pixels in a local neighbourhood of (i, j) and $w_{i,'j'}$ are the weights. The minimization is simplified by dividing the data in slices of L/W contiguous bands and running the algorithm introduced in [20] per slice. As shown in Fig. 6, the proposed strategy distributes the measurements uniformly in space, while other random strategies [16] tend to exhibit clusters, leaving some regions without measurements.

VI. EXPERIMENTS

To illustrate the efficacy of the proposed method, the new reconstruction algorithm is compared to other alternatives (based on the work conducted in [6]) using a synthetic dataset. Subsequently, the new subsampling scheme is compared with other random subsampling choices for a real MSL dataset. In all the experiments, the performance was measured using the following summary statistics:
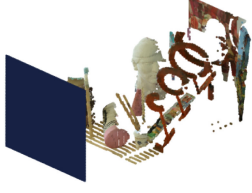
Fig. 7. Synthetic "Art" scene from Middlebury dataset with an additional semitransparent surface (blue plane).

Fig. 8. From left to right: $F_{\text{true}}(\tau)$, $F_{\text{false}}(\tau)$ and IAE(τ) for the proposed method and the two alternatives.

- True detections $F_{\text{true}}(\tau)$: Probability of true detection as a function of the distance τ, considering an estimated point as a true detection if there is another point in the ground truth point cloud in the same pixel ($x_n^{\text{true}} = x_{n'}^{\text{est}}$ and $y_n^{\text{true}} = y_{n'}^{\text{est}}$) such that $|t_n^{\text{true}} - t_{n'}^{\text{est}}| \leq \tau$.
- False detections $F_{\text{false}}(\tau)$: Number of estimated points that cannot be assigned to a ground truth point at a distance τ.
- Mean intensity absolute error at distance τ (IAE): Mean across all the points of the intensity absolute error $\sum_{\ell=1}^{L} |r_{n,\ell}^{\text{true}} - r_{n,'\ell}^{\text{est}}|$, normalized with respect to the total number of ground truth points. The ground truth and estimated points are coupled using the probability of detection $F_{\text{true}}(\tau)$. Note that if a point was falsely estimated or a ground truth point was not found, then they are considered to have resulted in an error of $\sum_{\ell=1}^{L} |r_{n,\ell}|$. The comparison is done with normalized intensity values, that is $\sum_{t=1}^{T} h_\ell(t) = 1$ for $\ell = 1, \dots, L$.
- Background normalized mean squared error NMSE$_B$: Mean of the normalized squared error of the estimated background at each wavelength, i.e., $\frac{1}{L} \sum_{\ell=1}^{L}$ $\frac{\sum_{i=1}^{N_r} \sum_{j=1}^{N_r} (b_{i,j,\ell}^{\text{true}} - \hat{b}_{i,j,\ell})^2}{\sum_{i=1}^{N_r} \sum_{j=1}^{N_r} (b_{i,j,\ell}^{\text{true}})^2}$.

A. Synthetic Data

We first assessed the performance of the proposed algorithm using a synthetic dataset created from the "Art" scene of the Middlebury dataset [50], as shown in Fig. 7. The measurements were obtained by simulating the single-photon multispectral Lidar system of [17], whose bin width is 0.3 mm. The generated dataset has $N_r = 283$ and $N_c = 231$ pixels, $T = 4500$ histogram bins and $L = 4$ wavelengths (red, green, blue and yellow), where only $W = 2$ wavelengths out of 4 were sampled per pixel using the coded aperture introduced in Section V. The mean number of photons per wavelength per pixel is 10, where approximately 3.4 photons are due to the background illumination. As in mono-static Lidar systems, the background levels are generated as a passive image of the scene (see Fig. 9).

We compared the proposed method to a two-stage algorithm that first estimates the point positions using ManiPoP [15] (single-wavelength multiple-return state-of-the-art algorithm) by integrating the photons across wavelengths and then infers the spectral signatures with fixed point positions, similarly to the procedure suggested by Wallace *et al.* [6]. The resulting method, referred to as ManiPoP #1, is summarized in Algorithm 2. We also compared with ManiPoP in the strict single-wavelength setting, by choosing the most powerful wavelength and using
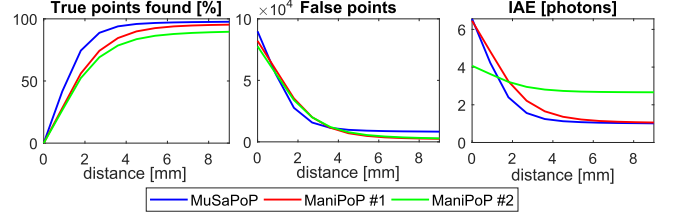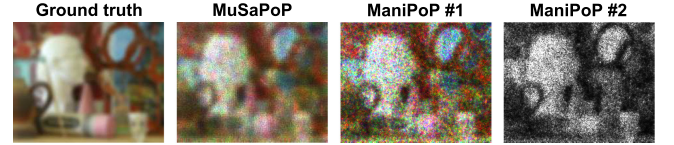

Fig. 9. From left to right: Ground truth background levels, estimates obtained by MuSaPoP and the two ManiPoP alternatives. Only the red, blue and green channels were used to generate these images. The proposed method provides smooth estimates due to the empirical prior distribution described in Section III-A3. ManiPoP #2 only estimates one wavelength, which is shown in grayscale.

Algorithm 2: ManiPoP #1 [6], [15].

Input: MSL waveforms Z
Depth estimation: Accumulate photons across wavelengths $z'_{i,j,t} = \sum_\ell z_{i,j,\ell,t}$ for all pixels (i,j)
$(\hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{B}}) \leftarrow$ ManiPoP($\boldsymbol{Z}'$)
for $\ell = 1, \dots, L$ and $\ell \neq w$ **do**
    Update $(\hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{B}})$ using ManiPoP($\boldsymbol{Z}_\ell$) in a fixed dimensional setting (only using background and reflectivity moves)
end for

---

the same total acquisition time per pixel than in the multispectral case (i.e., a per-pixel acquisition time $W = 2$ longer than the one considered for each wavelength in MuSaPoP). This second alternative is referred to as ManiPoP #2.

Fig. 8 illustrates $F_{\text{true}}(\tau)$, $F_{\text{false}}(\tau)$ and IAE for both methods. The proposed algorithm performs better than the other alternatives, as it finds 97.7% of the true points, whereas ManiPoP #1 and #2 only recover 95.34% and 89.6% respectively. ManiPoP #1 relies on an approximate impulse response $\tilde{h}(t) = \sum_{\ell=1}^{L} h_\ell(t)$, which biases the depth estimates, as the true accumulated response varies across points depending on their spectral signature.[5] The bias degrades the performance in terms of average depth absolute error (computed for true detections within a distance of 9 mm from the ground truth point). The proposed method obtains an average error of 3.9 mm, whereas the estimates by ManiPoP #1 present an average error of 5.7 mm. Despite having double acquisition time for the single-wavelength, ManiPoP #2 fails to find points corresponding to materials that have very low reflectivity in the blue wavelength

---

[5]Note that this bias can be arbitrarily large depending on the variations of $h_\ell(t)$ across wavelengths.
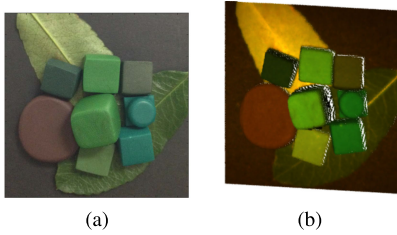
Fig. 10. (a) is an RGB image of the target and (b) shows the 3-D reconstructed scene (the colors were generated according to the CIE 1931 RGB color space).

(e.g., the red helicoidal structure shown in Fig. 7). The proposed MuSaPoP algorithm performs slightly worse in terms of false detections, finding 3 times more false points than the competitor methods. In terms of intensity estimation, MuSaPoP obtains better results, having an asymptotic IAE of 1 photon, whereas the alternatives #1 and #2 provide IAE equal to 1.1 and 2.7. The estimated background levels are shown in Fig. 9. The proposed method yields a better background NMSE (0.04) than alternatives #1 (0.14) and #2 (0.79). The improvement in background estimation over the ManiPoP alternatives can be attributed to the use of an empirical Bayes prior instead of a gamma Markov random field. The total execution time was 811 s for MuSaPoP and 294 and 348 s for alternatives #1 and #2.

B. Real MSL Data

The proposed subsampling scheme was evaluated on a real MSL dataset [17]. The scene consists of $L = 32$ wavelengths sampled at regular intervals of 10 nm from 500 nm to 810 nm, $N_r = N_c = 198$ pixels and $T = 4500$ histogram bins. The target is composed by a series of blocks of different types of clay and two leaves. Fig. 10 shows an RGB image of the scene and the 3D reconstruction using acquisition times up to 10 ms per wavelength per pixel. We compare the blue noise codes mentioned in Section V with the random schemes introduced in [16], all yielding the same total number of measurements and acquisition time:

1) Random sampling without overlap: W out of L bands per pixel are sampled without replacement (i.e., for a given pixel, each wavelength is measured at most once).
2) Random sampling without overlap: For each wavelength, $W/L\%$ of the pixels are sampled without replacement.
3) Proposed sampling method: For each wavelength, $W/L\%$ of the pixels are chosen following the scheme presented in Section V.

The codes were evaluated for $W = 1, 2, 4, 8, 16$ bands per pixel and acquisition times of 0.1, 1 and 10 ms per measurement (i.e., the histogram of one wavelength), using as ground-truth the reconstruction obtained with all the measurements and an acquisition time of 10 ms. Fig. 11 shows the percentage of true detections, IAE and background NMSEs for all codes, acquisition times and numbers of sensed bands per pixel W. All the evaluated compressive strategies yield good results, where a small improvement can be obtained by the use of blue noise codes. In terms of total number of estimated points, the blue
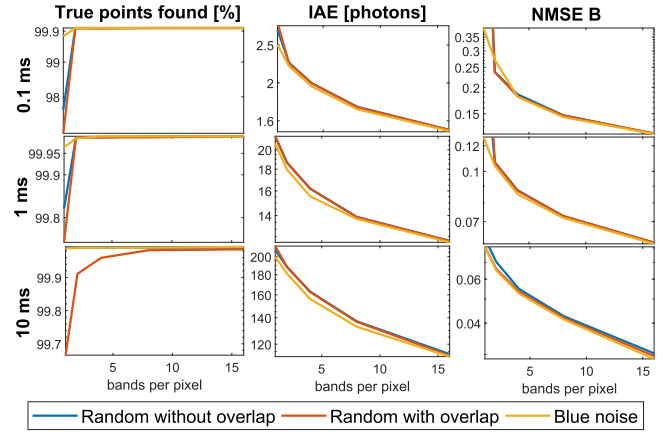

Fig. 11. $F_{\text{true}}(\tau)$, $F_{\text{false}}(\tau)$ and $\text{IAE}(\tau)$ obtained with the proposed reconstruction method for different acquisition times and sensed bands per pixel.
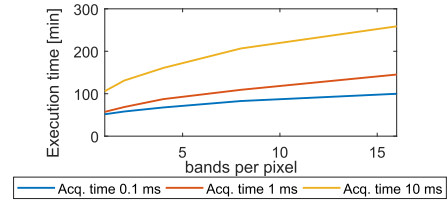

Fig. 12. Total execution time for different number of sensed bands per pixel and acquisition times of 0.1, 1 and 10 ms.

noise codes achieve better performance in high compression scenarios $W = 1, 2$ and low acquisition times (0.1 and 1 ms). For example, for an acquisition time of 1 ms, almost all points are reconstructed using blue noise codes, whereas the random codes only yield around 97% of the ground-truth points. The choice of blue noise codes has a stronger impact in terms of IAE, achieving smaller IAE for all acquisition times and number of bands per pixel. Fig. 12 shows the execution time for acquisition times of 10, 1 and 0.1 ms and different numbers of sensed bands. The proposed RJ-MCMC sampler has a complexity proportional to the number of photon detections in the support of the impulse response around the 3D point being modified, whereas the background update has a complexity proportional to the total number of active histogram bins in the Lidar scene. The background extraction step required around 15% of the total execution time, which could be significantly reduced if all the bands were processed in parallel instead of sequentially as it is done in the current implementation. All the experiments were performed using a Matlab R2018a implementation on a i7-3.0 GHz desktop computer (16 GB RAM).

Finally, we compared the performance of MuSaPoP with the single-depth multiple-wavelength algorithm by Altmann *et al.* [16]. The algorithm is referred to as Depth TV and considers total variation regularizations for the background, reflectivity and depth images. Note that this method requires a (global) depth interval where all signal photons are found, which is given manually by the user. We also considered a target detection scenario (i.e., some pixels without surfaces), by removing the
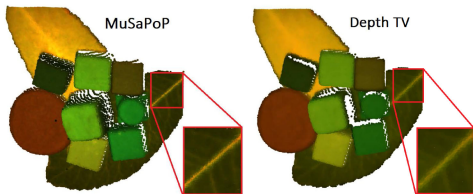
Fig. 13. From left to right: 3-D reconstructions obtained by the proposed method and Depth TV for an acquisition time of 10 ms. Note that Depth TV tends to smooth out fine scale details (zoom for better visualization). Moreover, the thresholding step used in Depth TV removes some low intensity points in the borders of each 3-D object.
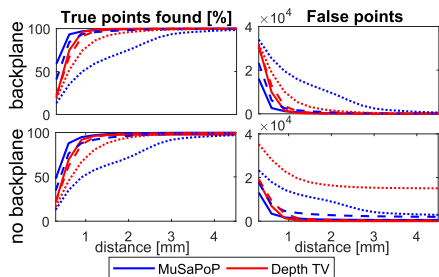

Fig. 14. True and false point detections for MuSaPoP and Depth TV for the real MSL dataset with (top row) and without (bottom row) backplane. Solid, dashed and dotted lines represent the datasets with acquisition times of 1, 0.1 and 0.01 ms respectively.

TABLE I
HYPERPARAMETERS VALUES

Hyperparameter	Coarse scale	Fine scale
γ_a	e^2	e^3
λ_a	$(N_r N_r / N_{\text{bin}}^2)^{1.5}$	$(N_r N_r)^{1.5}$
N_b	$8 N_{\text{bin}} \Delta_p / \Delta_{\text{bin}}$	$8 \Delta_p / \Delta_{\text{bin}}$
$d_{\min}$	$2N_b + 1$	$2N_b + 1$
σ^2	0.6^2	$0.6^2 / N_{\text{bin}}$
β	$\sigma^2 / 100$	$\sigma^2 / 100$

backplane of the scene and keeping only photons associated with background levels. In this case, we post-process the Depth TV estimates, removing points with a mean normalized intensity below 10%, which gave the best results across the evaluated datasets. In both experiments, we used the blue noise codes with $W = 8$ wavelengths per pixel out of $L = 32$. Fig. 13 shows the 3-D reconstructions obtained by MuSaPoP and Depth TV using an acquisition time of 10 ms. Fig. 14 shows the performance of both algorithms in terms of true and false detections. MuSaPoP performs better in the 1 and 0.1 ms cases, whereas Depth TV obtains better depth estimates in the lowest acquisition time case (0.01 ms). However, in the 0.01 ms case without backplane, the intensity thresholding step does not remove backplane points, hence obtaining a very large number of false detections. This result illustrates the inefficiency of simple thresholding in target detection scenarios, whereas MuSaPoP includes these cases within its general formulation. Table II shows the performance of both algorithms in terms of IAE, background NMSE and execution time. The proposed method yields a better IAE than Depth TV (approximately half), as the latter tends to smooth out details within the blocks and leaves, as shown in Fig. 13. Moreover, in terms of background NMSE, Depth TV fails to

TABLE II
IAE, BACKGROUND NMSE AND EXECUTION TIME OF DEPTH TV AND THE PROPOSED METHOD FOR THE BLOCKS AND LEAVES DATASET WITH AND WITHOUT THE BACKPLANE

Backplane		Yes			No		
Acq. time [ms]		1	0.1	0.01	1	0.1	0.01
IAE [photons]	Depth TV	36	3.7	0.5	45	4.9	0.8
	MuSaPoP	**14**	**1.7**	**0.3**	**19**	**2.4**	**0.4**
Bkg. NMSE	Depth TV	0.12	>1	>1	0.12	>1	>1
	MuSaPoP	**0.04**	**0.11**	**0.36**	**0.04**	**0.11**	**0.26**
Execution time [h]	Depth TV	19.8	17.9	17.7	19.3	17.9	17.7
	MuSaPoP	**2.8**	**1.4**	**1.0**	**1.5**	**1**	**0.8**

provide good estimates in the low-photon cases, as it only considers photon counts within the global interval without signal returns. The execution time of Depth TV was significantly higher than MuSaPoP.

VII. CONCLUSIONS

This paper has studied a new 3D reconstruction algorithm referred to as MuSaPoP using multispectral Lidar data, which is able to find multiple surfaces per pixel. The proposed method leads to better reconstruction quality than other alternatives, as it considers all measured wavelengths in a single observation model. While based on some ideas initially investigated in ManiPoP [15], MuSaPoP also relies on new strategies to deal with the very high dimensionality of the multispectral problem. The first novelty is the use of an empirical Bayes prior for the background levels, which speeds up significantly the RJ-MCMC algorithm. A second improvement is the adapted dilation/erosion and split/merge moves for the multispectral case, profiting from SBR estimates to increase the acceptance rate. Finally, the subsampling strategy further reduces both the algorithm's complexity and total number of measurements, leading to faster acquisitions and reconstructions. The sparse point cloud representation of the proposed method speeds up the computations proportionally to the number of measurements, whereas other dense models [13], [14] would not achieve similar improvements.

Further work will be devoted to the design of compressive systems that are not limited to subsampling the spectral cube. Moreover, the compression can also be extended to depth information, for example using a range-gated camera [18].

APPENDIX A
GAMMA MARKOV RANDOM FIELD

Most classical prior distributions (often referred to as regularization terms in the convex optimization literature) for images, such as Laplacian [26] and total variation [51], only penalize variations between neighbouring pixels, ignoring the mean intensity of the image. However, the gamma Markov random field [31], includes a penalization on the mean intensity, which promotes pixels with smaller values. This can be shown by inspecting the marginal distribution (defined as [32]),

$$p(\boldsymbol{B} | \alpha_B) \propto \prod_{\ell=1}^{L} \prod_{i=1}^{N_c} \prod_{j=1}^{N_r} \frac{b_{i,j,\ell}^{\alpha_B - 1}}{\left(\sum_{(i,'j') \in \mathcal{M}_B} b_{i,'j',\ell} \right)^{\alpha_B}} \quad (24)$$

where α_B is a hyperparameter controlling the degree of smoothness and $\mathcal{M}_B$ denotes the set of pixels in the neighbourhood of pixel (i, j). For an image of constant intensity c, we have $b_{i',j',\ell} = b_{i,j,\ell} = c$ for all pixels and spectral bands, yielding the density

$$p(\boldsymbol{B}|\alpha_B) \propto \prod_{\ell=1}^{L} \prod_{i=1}^{N_c} \prod_{j=1}^{N_r} c^{-1} = \prod_{i=1}^{N_c} \prod_{j=1}^{N_r} c^{-L}. \tag{25}$$

This dependency on the mean promotes reconstructions with lower background levels, decreasing the acceptance ratio of death and erosion moves (that propose to increase the background levels). In the case of ManiPoP, only one band is considered ($L = 1$). Thus, the bias towards smaller background levels does not impact the overall reconstruction significantly. However, in the MSL case ($L \gg 1$), the reconstruction quality is reduced, hindering the use of gamma Markov random fields.

APPENDIX B
EMPIRICAL PRIOR FOR THE BACKGROUND LEVELS

The prior for the background levels is chosen to minimize the Kullback-Leibler divergence in (9), where the correlated model q is given by (10). The minimization of (9) can be written as

$$(\boldsymbol{K}, \boldsymbol{\Theta}) = \underset{(\boldsymbol{K}, \boldsymbol{\Theta})}{\arg \min} -\mathbb{E}_q\{\log p(\boldsymbol{B}|\boldsymbol{K}, \boldsymbol{\Theta})\}. \tag{26}$$

Considering the product of independent gamma distributions in (8), the problem reduces to

$$
\begin{aligned}
(k_{i,j,\ell}, \theta_{i,j,\ell}) = & \underset{(k_{i,j,\ell}, \theta_{i,j,\ell})}{\arg \min} \frac{\mathbb{E}_q\{b_{i,j,\ell}\}}{\theta_{i,j,\ell}} \\
& - k_{i,j,\ell} \left(\mathbb{E}_q\{\log b_{i,j,\ell}\} - \log \theta_{i,j,\ell}\right) \\
& + \log \Gamma(k_{i,j,\ell})
\end{aligned} \tag{27}
$$

for all pixels $i = 1, \ldots, N_r$ and $j = 1, \ldots, N_c$ and wavelengths $\ell = 1, \ldots, L$. The expected values $\mathbb{E}_q\{b_{i,j,\ell}\}$ and $\mathbb{E}_q\{\log b_{i,j,\ell}\}$ cannot be obtained in closed form for the Poisson-Gaussian model of (10). Thus, we approximate them numerically by obtaining MCMC samples of $\tilde{b}_{i,j,\ell}$. As explained in [32], off-the-shelf sampling strategies (e.g., Hamiltonian Monte Carlo [24], Chapter 9) do not scale well with the dimension of the problem, being inefficient when applied to large multispectral Lidar datasets. Hence, we consider proposals from a Gaussian approximation of (10) (as detailed in [26]) using the perturbation optimization algorithm [52], accepting or rejecting them according to the Metropolis-Hastings rule [24], [26]. We generate 10^3 samples $\{\tilde{b}_{i,j,\ell}^{(s)}, s = 1, \ldots, 10^3\}$ and compute the desired expected values as

$$\mathbb{E}_q\{b_{i,j,\ell}\} = \sum_s \exp \tilde{b}_{i,j,\ell}^{(s)} \tag{28}$$

$$\mathbb{E}_q\{\log b_{i,j,\ell}\} = \sum_s \tilde{b}_{i,j,\ell}^{(s)}. \tag{29}$$

Finally, the values of the hyperparameters are obtained by setting $\theta_{i,j,\ell} = \mathbb{E}_q\{b_{i,j,\ell}\}$ and minimizing (27) with a one-dimensional Newton method.

APPENDIX C
COMPLETE EXPRESSIONS OF THE RJ-MCMC MOVES

The birth move of point $(\boldsymbol{c}_{N_\Phi+1}, \boldsymbol{r}_{N_\Phi+1})$ has an acceptance ratio given by $\rho = \min\{1, r(\boldsymbol{\theta}, \boldsymbol{\theta}')\}$ with

$$
r(\boldsymbol{\theta}, \boldsymbol{\theta}') = \begin{cases} C_1 & \text{if } |t_{N_\Phi+1} - t_n| > d_{\min} \forall n \neq N_\Phi + 1: \\ & x_n = x_{N_\Phi+1} \text{ and } y_n = y_{N_\Phi+1} \\ 0 & \text{otherwise} \end{cases}
$$

where C_1 is defined as

$$
\begin{aligned}
C_1 = & \prod_{\ell=1}^{L} \prod_{t=1}^{T} \left(\frac{\sum_{\substack{n:x_n=i \\ y_n=j}} r'_{n,\ell} h_\ell(t - t'_n) + b'_{i,j,\ell}}{\sum_{\substack{n:x_n=i \\ y_n=j}} r_{n,\ell} h_\ell(t - t_n) + b_{i,j,\ell}} \right)^{z_{i,j,t,\ell}} \frac{p_{\text{death}}}{p_{\text{birth}}} \\
& \times \lambda_a \gamma_a^{-m\left(S(\boldsymbol{c}_{N_\Phi+1}) \backslash \bigcup_{n' \in \mathcal{M}_{pp}(\boldsymbol{c}_{N_\Phi+1})} S(\boldsymbol{c}_{n'})\right)} \\
& \times \frac{1}{N_\Phi + 1} \left(\frac{|\boldsymbol{P}'|}{|\boldsymbol{P}|} \frac{1}{2\pi\sigma^2} \right)^{\frac{L}{2}} \prod_{\ell=1}^{L} \exp \\
& \times \left(-\sum_{n' \in \mathcal{M}_{pp}(\boldsymbol{c}_n)} \frac{(m_{N_\Phi+1,\ell} - m_{n',\ell})^2}{2\sigma^2 d(\boldsymbol{c}_{N_\Phi+1}; \boldsymbol{c}_{n'})} - \frac{m_{N_\Phi+1,\ell}^2 \beta}{2\sigma^2} \right) \\
& \times (1 - u)^{-L} \prod_{\ell=1}^{L} \exp \\
& \times \left(g_{i,j,\ell} r_{N_\Phi+1,\ell}(1 - w_\ell^{-1}) \left(\sum_{t=1}^{T} h_\ell(t) \right) \right) \\
& \times \prod_{(i,j) \in \mathcal{M}_B(b_{i,j})} \prod_{\ell=1}^{L} \left(\frac{b'_{i,j,\ell}}{b_{i,j,\ell}} \right)^{r_{i,j,\ell}-1} \exp \left(\frac{b_{i,j,\ell} - b'_{i,j,\ell}}{\theta_{i,j,\ell}} \right).
\end{aligned}
$$

Similarly, the death move is accepted with probability $\rho = \min\{1, C_1^{-1}\}$, where the term $\frac{1}{N_\Phi+1}$ in the second line is replaced by $\frac{1}{N_\Phi}$. The dilation move of point $(\boldsymbol{c}_{N_\Phi+1}, r_{N_\Phi+1})$ is accepted with probability $\rho = \min\{1, r(\boldsymbol{\theta}, \boldsymbol{\theta}')\}$ with

$$
r(\boldsymbol{\theta}, \boldsymbol{\theta}') = \begin{cases} C_2 & \text{if } |t_{N_\Phi+1} - t_n| > d_{\min} \quad \forall n \neq N_\Phi + 1 \\ & x_n = x_{N_\Phi+1} \text{ and } y_n = y_{N_\Phi+1} \\ 0 & \text{otherwise} \end{cases}
$$

where C_2 is defined as

$$
\begin{aligned}
C_2 = & \prod_{\ell=1}^{L} \prod_{t=1}^{T} \left(\frac{\sum_{\substack{n:x_n=i \\ y_n=j}} r'_{n,\ell} h_\ell(t - t'_n) + b'_{i,j,\ell}}{\sum_{\substack{n:x_n=i \\ y_n=j}} r_{n,\ell} h_\ell(t - t_n) + b_{i,j,\ell}} \right)^{z_{i,j,t,\ell}} \\
& \times \frac{p_{\text{erosion}}}{p_{\text{dilation}}} \lambda_a \gamma_a^{-m\left(S(\boldsymbol{c}_{N_\Phi+1}) \backslash \bigcup_{n' \in \mathcal{M}_{pp}(\boldsymbol{c}_{N_\Phi+1})} S(\boldsymbol{c}_{n'})\right)} \\
& \times \left(\frac{|\boldsymbol{P}'|}{|\boldsymbol{P}|} \frac{1}{2\pi\sigma^2} \right)^{\frac{L}{2}} \frac{N_\Phi(2N_b + 1)}{\sum_{m \in \mathcal{M}_{pp}(\boldsymbol{c}_{N_\Phi+1})} \# \mathcal{M}_{pp}(\boldsymbol{c}_m)} \\
& \times \frac{1}{\sum_{m=1}^{N_\Phi+1} \mathbb{1}_{\mathbb{Z}_+}(\# \mathcal{M}_{pp}(\boldsymbol{c}_m))} \prod_{\ell=1}^{L} \exp
\end{aligned}
$$

$$\times \left(-\sum_{n' \in \mathcal{M}_{pp}(\boldsymbol{c}_n)} \frac{(m_{N_\Phi+1,\ell} - m_{n',\ell})^2}{2\sigma^2 d(\boldsymbol{c}_{N_\Phi+1}; \boldsymbol{c}_{n'})} - \frac{m_{N_\Phi+1,\ell}{}^2 \beta}{2\sigma^2}\right)$$

$$\times (1-u)^{-L} \prod_{\ell=1}^{L} \exp\left(g_{i,j,\ell} r_{N_\Phi+1,\ell}(1 - w_\ell^{-1})\right)$$

$$\times \left(\sum_{t=1}^{T} h_\ell(t)\right) \prod_{(i,j) \in \mathcal{M}_B(b_{i,j})} \prod_{\ell=1}^{L} \left(\frac{b'_{i,j,\ell}}{b_{i,j,\ell}}\right)^{r_{i,j,\ell}-1}$$

$$\times \exp\left(\frac{b_{i,j,\ell} - b'_{i,j,\ell}}{\theta_{i,j,\ell}}\right).$$

A shift of the point $(\boldsymbol{c}_n, \boldsymbol{r}_n)$ to the new position $\boldsymbol{c}'_n = [x_n, y_n, t'_n]^T$ has an acceptance probability of $\rho = \min\{1, r(\boldsymbol{\theta}, \boldsymbol{\theta}')\}$ with

$$r(\boldsymbol{\theta}, \boldsymbol{\theta}') = \begin{cases} C_3 & \text{if } |t'_n - t_m| > d_{\min} \quad \forall n \neq m : \\ & x_m = x_n \text{ and } y_m = y_n \\ 0 & \text{otherwise} \end{cases}$$

where

$$C_3 = \prod_{\ell=1}^{L} \prod_{t=1}^{T} \left(\frac{\sum_{\substack{n:x_n=i \\ y_n=j}} r'_{n,\ell} h_\ell(t - t'_n) + b'_{i,j,\ell}}{\sum_{\substack{n:x_n=i \\ y_n=j}} r_{n,\ell} h_\ell(t - t_n) + b_{i,j,\ell}}\right)^{z_{i,j,t,\ell}}$$

$$\times \prod_{\ell=1}^{L} \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{n' \in \mathcal{M}_{pp}(\boldsymbol{c}'_n)} \frac{(m_{n,\ell} - m_{n',\ell})^2}{d(\boldsymbol{c}'_n; \boldsymbol{c}_{n'})}\right)\right)$$

$$\times \left(\frac{|\boldsymbol{P}'|}{|\boldsymbol{P}|}\right)^{\frac{L}{2}} \prod_{\ell=1}^{L} \exp\left(\frac{1}{2\sigma^2}\left(\sum_{n' \in \mathcal{M}_{pp}(\boldsymbol{c}_n)} \frac{(m_{n,\ell} - m_{n',\ell})^2}{d(\boldsymbol{c}_n; \boldsymbol{c}_{n'})}\right)\right)$$

$$\times \gamma_a^{-m\left(S(\boldsymbol{c}'_n)\backslash \bigcup_{n' \in \mathcal{M}_{pp}(\boldsymbol{c}'_n)} S(\boldsymbol{c}_{n'})\right) + m\left(S(\boldsymbol{c}_n)\backslash \bigcup_{n' \in \mathcal{M}_{pp}(\boldsymbol{c}_n)} S(\boldsymbol{c}_{n'})\right)}.$$

A mark update randomly picks a point $(\boldsymbol{c}_n, \boldsymbol{r}_n)$ and proposes a new spectral signature $\boldsymbol{r}'_n = \log(\boldsymbol{m}'_n)$. Each spectral log-intensity is accepted independently with probability $\rho = \min\{1, C_4\}$, where

$$C_4 = \prod_{\ell=1}^{L} \prod_{t=1}^{T} \left(\frac{\sum_{\substack{n:x_n=i \\ y_n=j}} r'_{n,\ell} h_\ell(t - t'_n) + b'_{i,j,\ell}}{\sum_{\substack{n:x_n=i \\ y_n=j}} r_{n,\ell} h_\ell(t - t_n) + b_{i,j,\ell}}\right)^{z_{i,j,t,\ell}}$$

$$\times \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{n' \in \mathcal{M}_{pp}(\boldsymbol{c}'_n)} \frac{(m'_{n,\ell} - m_{n',\ell})^2}{d(\boldsymbol{c}'_n; \boldsymbol{c}_{n'})} + m'_{n,\ell}{}^2 \beta\right)\right)$$

$$\times \exp\left(\frac{1}{2\sigma^2}\left(\sum_{n' \in \mathcal{M}_{pp}(\boldsymbol{c}_n)} \frac{(m_{n,\ell} - m_{n',\ell})^2}{d(\boldsymbol{c}_n; \boldsymbol{c}_{n'})} + m_{n,\ell}^2 \beta\right)\right)$$

$$\times \prod_{\ell=1}^{L} \exp\left(g_{i,j,\ell}(r_{n,\ell} - r'_{n,\ell})(1 - w_\ell^{-1})\right)\left(\sum_{t=1}^{T} h_\ell(t)\right).$$

The split move from $(\boldsymbol{c}_n = [x_n, y_n, t_n]^T, \boldsymbol{r}_n)$ to $(\boldsymbol{c}'_{k_1} = [x_n, y_n, t'_{k_1}]^T, \boldsymbol{r}'_{k_1})$ and $(\boldsymbol{c}'_{k_2} = [x_n, y_n, t'_{k_2}]^T, \boldsymbol{r}'_{k_2})$ is accepted

with probability $\rho = \min\{1, r(\boldsymbol{\theta}, \boldsymbol{\theta}')\}$, where

$$r(\boldsymbol{\theta}, \boldsymbol{\theta}') = \begin{cases} C_5 & \text{if } |t'_n - t_m| > d_{\min} \quad \forall n \neq m : \\ & x_m = x_n \text{ and } y_m = y_n \\ 0 & \text{otherwise} \end{cases}$$

and

$$C_5 = \prod_{\ell=1}^{L} \prod_{t=1}^{T} \left(\frac{\sum_{\substack{n:x_n=i \\ y_n=j}} r'_{n,\ell} h_\ell(t - t'_n) + b'_{i,j,\ell}}{\sum_{\substack{n:x_n=i \\ y_n=j}} r_{n,\ell} h_\ell(t - t_n) + b_{i,j,\ell}}\right)^{z_{i,j,t,\ell}}$$

$$(u(1-u))^{-L} N_\Phi (\# \text{ points in } \boldsymbol{\Phi} \text{ that verify (22)})^{-1} \left(\frac{|\boldsymbol{P}'|}{|\boldsymbol{P}|}\right)^{\frac{L}{2}}$$

$$\prod_{\ell=1}^{L} \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{n' \in \mathcal{M}_{pp}(\boldsymbol{c}'_{k_1})} \frac{(m_{k_1,\ell} - m_{n',\ell})^2}{d(\boldsymbol{c}'_{k_1}; \boldsymbol{c}_{n'})}\right)\right)$$

$$\prod_{\ell=1}^{L} \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{n' \in \mathcal{M}_{pp}(\boldsymbol{c}'_{k2})} \frac{(m_{k_1,\ell} - m_{n',\ell})^2}{d(\boldsymbol{c}'_{k_2}; \boldsymbol{c}_{n'})}\right)\right)$$

$$\prod_{\ell=1}^{L} \exp\left(\frac{1}{2\sigma^2}\left(\sum_{n' \in \mathcal{M}_{pp}(\boldsymbol{c}_n)} \frac{(m_{n,\ell} - m_{n',\ell})^2}{d(\boldsymbol{c}_n; \boldsymbol{c}_{n'})}\right)\right)$$

$$\gamma_a^{-m\left(S(\boldsymbol{c}'_{k_1})\backslash \bigcup_{n' \in \mathcal{M}_{pp}(\boldsymbol{c}'_{k_1})} S(\boldsymbol{c}_{n'})\right) + m\left(S(\boldsymbol{c}_n)\backslash \bigcup_{n' \in \mathcal{M}_{pp}(\boldsymbol{c}_n)} S(\boldsymbol{c}_{n'})\right)}$$

$$\lambda_a \gamma_a^{-m\left(S(\boldsymbol{c}'_{k_2})\backslash \bigcup_{n' \in \mathcal{M}_{pp}(\boldsymbol{c}'_{k_2})} S(\boldsymbol{c}_{n'})\right)} 2(d_{\min} + L_h) \frac{p_{\text{merge}}}{p_{\text{split}}}.$$

Finally, the merge move is accepted with probability $\rho = \min\{1, C_5^{-1}\}$.

Acknowledgment

References

[1] A. McCarthy *et al.*, "Kilometer-range, high resolution depth imaging via 1560 nm wavelength single-photon detection," *Opt. Express*, vol. 21, no. 7, pp. 8904–8915, Apr. 2013.

[2] J. Hecht, "Lidar for self-driving cars," *Opt. Photon. News*, vol. 29, no. 1, pp. 26–33, 2018.

[3] M. A. Canuto *et al.*, "Ancient lowland Maya complexity as revealed by airborne laser scanning of northern Guatemala," *Science*, vol. 361, no. 6409, pp. 1–17, 2018.

[4] C. Mallet and F. Bretar, "Full-waveform topographic Lidar: State-of-the-art," *ISPRS J. Photogrammetry Remote Sens.*, vol. 64, no. 1, pp. 1–16, 2009.

[5] E. S. Douglas *et al.*, "Finding leaves in the forest: The dual-wavelength echidna Lidar," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 4, pp. 776–780, Apr. 2015.

[6] A. M. Wallace *et al.*, "Design and evaluation of multispectral LiDAR for the recovery of arboreal parameters," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4942–4954, Aug. 2014.

[7] Y. Altmann, A. Wallace, and S. McLaughlin, "Spectral unmixing of multispectral Lidar signals," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5525–5534, Oct. 2015.

[8] 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 J. Photogrammetry Remote Sens.*, vol. 69, pp. 1–9, 2012.

[9] 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. Express*, vol. 26, no. 23, pp. 30 146–30 161, Nov. 2018.

[10] A. Kirmani *et al.*, "First-photon imaging," *Science*, vol. 343, no. 6166, pp. 58–61, 2014.

[11] J. Rapp and V. K. Goyal, "A few photons among many: Unmixing signal and noise for photon-efficient active imaging," *IEEE Trans. Comput. Imag.*, vol. 3, no. 3, pp. 445–459, Sep. 2017.

[12] D. B. Lindell, M. OToole, and G. Wetzstein, "Single-photon 3D imaging with deep sensor fusion," *ACM Trans. Graph.*, vol. 37, no. 4, Aug. 2018, Art. no. 113.

[13] D. Shin, F. Xu, F. N. Wong, J. H. Shapiro, and V. K. Goyal, "Computational multi-depth single-photon imaging," *Opt. Express*, vol. 24, no. 3, pp. 1873–1888, 2016.

[14] A. Halimi *et al.*, "Restoration of intensity and depth images constructed using sparse single-photon data," in *Proc. Eur. Signal Process. Conf.*, Budapest, Hungary, Aug./Sep. 2016, pp. 86–90.

[15] J. Tachella *et al.*, "Bayesian 3D reconstruction of complex scenes from single-photon Lidar data," *SIAM J. Imag. Sci.*, vol. 12, no. 1, pp. 521–550, 2019.

[16] Y. Altmann *et al.*, "Bayesian restoration of reflectivity and range profiles from subsampled single-photon multispectral lidar data," in *Proc. Eur. Signal Process. Conf.*, Kos Island, Greece, Aug. 2017, pp. 1410–1414.

[17] Y. Altmann *et al.*, "Robust spectral unmixing of sparse multispectral lidar waveforms using gamma Markov random fields," *IEEE Trans. Comput. Imag.*, vol. 3, no. 4, pp. 658–670, Dec. 2017.

[18] X. Ren *et al.*, "High-resolution depth profiling using a range-gated CMOS SPAD quanta image sensor," *Opt. Express*, vol. 26, no. 5, pp. 5541–5557, Mar. 2018.

[19] H. Arguello and G. R. Arce, "Colored coded aperture design by concentration of measure in compressive spectral imaging," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1896–1908, Apr. 2014.

[20] C. V. Correa, H. Arguello, and G. R. Arce, "Spatiotemporal blue noise coded aperture design for multi-shot compressive spectral imaging," *JOSA A*, vol. 33, no. 12, pp. 2312–2322, 2016.

[21] D. Shin *et al.*, "Photon-efficient imaging with a single-photon camera," *Nature Commun.*, vol. 7, 2016, Art. no. 12046.

[22] 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. Imag.*, to be published, doi: 10.1109/TCI.2019.2929918.

[23] C. Robert, *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. Berlin, Germany: Springer, 2007.

[24] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng, *Handbook of Markov Chain Monte Carlo*. Boca Raton, FL, USA: CRC Press, 2011.

[25] A. J. Baddeley and M. Van Lieshout, "Area-interaction point processes," *Ann. Inst. Statistical Math.*, vol. 47, no. 4, pp. 601–619, 1995.

[26] H. Rue and L. Held, *Gaussian Markov Random Fields: Theory and Applications*. Boca Raton, FL, USA: CRC Press, 2005.

[27] H. Rue, S. Martino, and N. Chopin, "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations," *J. Roy. Statistical Soc., Series B (Statistical Methodol.)*, vol. 71, no. 2, pp. 319–392, 2009.

[28] J. Salmon, Z. Harmany, C.-A. Deledalle, and R. Willett, "Poisson noise reduction with non-local PCA," *J. Math. Imag. Vis.*, vol. 48, no. 2, pp. 279–294, 2014.

[29] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[30] D. Shin, A. Kirmani, V. K. Goyal, and J. H. Shapiro, "Photon-efficient computational 3D and reflectivity imaging with single-photon detectors," *IEEE Trans. Comput. Imag.*, vol. 1, no. 2, pp. 112–125, Jun. 2015.

[31] O. Dikmen and A. T. Cemgil, "Gamma Markov random fields for audio source modeling," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 3, pp. 589–601, Mar. 2010.

[32] J. Tachella, Y. Altmann, M. Pereyra, S. McLaughlin, and J.-Y. Tourneret, "Bayesian restoration of high-dimensional photon-starved images," in *Proc. Eur. Signal Process. Conf.*, Rome, Italy, Sep. 2018, pp. 747–751.

[33] M. J. Beal, "Variational algorithms for approximate Bayesian inference," Ph.D. dissertation, Gatsby Computational Neuroscience Unit, University College London, London WC1E 6BT, U.K., 2003.

[34] T. P. Minka, "Expectation propagation for approximate Bayesian inference," in *Proc. Conf. Uncertainty Artif. Intell.*, Seattle, WA, USA, Aug. 2001, pp. 362–369.

[35] M. Zhou, L. Hannah, D. B. Dunson, and L. Carin, "Beta-negative binomial process and Poisson factor analysis," in *Proc. 15th Int. Conf. Artif. Intell. Statist.*, Apr. 2012, vol. 22, pp. 1462–1471.

[36] M. Raginsky, R. M. Willett, Z. T. Harmany, and R. F. Marcia, "Compressed sensing performance bounds under poisson noise," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 3990–4002, Aug. 2010.

[37] M. Raginsky, S. Jafarpour, Z. T. Harmany, R. F. Marcia, R. M. Willett, and R. Calderbank, "Performance bounds for expander-based compressed sensing in Poisson noise," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4139–4153, Sep. 2011.

[38] Y. Li and G. Raskutti, "Minimax optimal convex methods for Poisson inverse problems under ℓ_q-ball sparsity," *IEEE Trans. Inf. Theory*, vol. 64, no. 8, pp. 5498–5512, Aug. 2018.

[39] O. Deussen, S. Hiller, C. Van Overveld, and T. Strothotte, "Floating points: A method for computing stipple drawings," in *Proc. Comput. Graph. Forum*, 2000, vol. 19, no. 3, pp. 41–50.

[40] L. Galvis, E. Mojica, H. Arguello, and G. R. Arce, "Shifting colored coded aperture design for spectral imaging," *Appl. Opt.*, vol. 58, no. 7, pp. B28–B38, 2019.

[41] L. Galvis, D. Lau, X. Ma, H. Arguello, and G. R. Arce, "Coded aperture design in compressive spectral imaging based on side information," *Appl. Opt.*, vol. 56, no. 22, pp. 6332–6340, 2017.

[42] K. M. León-López, L. V. G. Carreño, and H. A. Fuentes, "Temporal colored coded aperture design in compressive spectral video sensing," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 253–264, Jan. 2019.

[43] E. Mojica, S. Pertuz, and H. Arguello, "High-resolution coded-aperture design for compressive x-ray tomography using low resolution detectors," *Opt. Commun.*, vol. 404, pp. 103–109, 2017.

[44] K. Choi and D. J. Brady, "Coded aperture computed tomography," *Proc. SPIE,* 2009, vol. 7468, Art. no. 74680B.

[45] C. Hinojosa, J. Bacca, and H. Arguello, "Coded aperture design for compressive spectral subspace clustering," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 6, pp. 1589–1600, Dec. 2018.

[46] V. Surazhsky, P. Alliez, and C. Gotsman, "Isotropic remeshing of surfaces: A local parameterization approach," Research Report RR-4967, INRIA, Rocquencourt, France, 2003.

[47] O. Deussen, P. Hanrahan, B. Lintermann, R. Měch, M. Pharr, and P. Prusinkiewicz, "Realistic modeling and rendering of plant ecosystems," in *Proc. 25th Annu. Conf. Comput. Graph. Interactive Techn.*, 1998, pp. 275–286.

[48] G. Liang, L. Lu, Z. Chen, and C. Yang, "Poisson disk sampling through disk packing," *Comput. Vis. Media*, vol. 1, no. 1, pp. 17–26, 2015.

[49] D. L. Lau, R. Ulichney, and G. R. Arce, "Blue and green noise halftoning models," *IEEE Signal Process. Mag.*, vol. 20, no. 4, pp. 28–38, Jul. 2003.

[50] D. Scharstein and C. Pal, "Learning conditional random fields for stereo," in *Proc. Int. Conf. Comput. Vis. Pattern Recognit.*, Minneapolis, MN, USA, Jun. 2007, pp. 1–8.

[51] A. Chambolle and T. Pock, "An introduction to continuous optimization for imaging," *Acta Numerica*, vol. 25, 2016, Art. no. 161319.

[52] C. Gilavert, S. Moussaoui, and J. Idier, "Efficient Gaussian sampling for solving large-scale inverse problems using MCMC," *IEEE Trans. Signal Process.*, vol. 63, no. 1, pp. 70–80, Jan. 2015.