This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCI.2021.3111572, IEEE Transactions on Computational Imaging

Robust and Guided Bayesian Reconstruction of Single-Photon 3D Lidar Data: Application to Multispectral and Underwater Imaging

Abderrahim Halimi [1], *Senior Member, IEEE*, Aurora Maccarone [1], Robert A. Lamb [2], Gerald S. Buller [1], Stephen McLaughlin [1], *Fellow, IEEE*

***Abstract*—3D Lidar imaging can be a challenging modality when using multiple wavelengths, or when imaging in high noise environments (e.g., imaging through obscurants). This paper presents a hierarchical Bayesian algorithm for the robust reconstruction of multispectral single-photon Lidar data in such environments. The algorithm exploits multi-scale information to provide robust depth and reflectivity estimates together with their uncertainties to help with decision making. The proposed weight-based strategy allows the use of available guide information that can be obtained by using state-of-the-art learning based algorithms. The proposed Bayesian model and its estimation algorithm are validated on both synthetic and real images showing competitive results regarding the quality of the inferences and the computational complexity when compared to the state-of-the-art algorithms.**

***Index Terms*—3D reconstruction, Lidar, multispectral imaging, obscurants, robust estimation, Poisson noise, Bayesian inference.**

I. Introduction

Three-dimensional (3D) imaging has generated significant interest from the scientific community due to its increasing use in applications such as self-driving autonomous vehicles [1], [2]. Single-photon light detection and ranging (Lidar) is a technology for high resolution 3D imaging, where its high sensitivity and excellent surface-to-surface resolution can provide rich information on the depth profile and reflectivity of observed targets in challenging imaging scenarios. Single-photon Lidar operates by emitting picosecond duration laser pulses and collecting the reflected photons using a single-photon sensitive detector which measures the arrival time of each return photon using a time correlated single-photon counting (TCSPC) system [3]. This results in a collection of X-Y pixels, where a timing histogram of photon counts with respect to their time of flight is constructed for each pixel. In the presence of a target with partially reflective or scattering surfaces, the histogram will contain a peak whose amplitude and location are related to the object reflectivity and distance from the sensor. This process can be repeated using different laser wavelengths to obtain a multispectral 3D image of the scene.

(1) School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, EH14 4AS, United Kingdom

(2) Leonardo MW Ltd, Crewe Road North, Edinburgh, EH5 2XS, UK

This work was supported by the UK Royal Academy of Engineering under the Research Fellowship Schemes (RF/201718/17128, RF/201920/19/190), the EPSRC Grants EP/T00097X/1, EP/N003446/1, and by DSTL DASA project DSTLX1000147844.

Several practical challenges currently limit the use of Lidar in real world conditions. This paper focuses on some of them and provides a principled statistical-based solution to improve performance. Such challenges include the photon sparse regime [4]–[6] often observed for long-range imaging [7]–[9] or rapid imaging based on short acquisition times [10], [11] or adaptive imaging [12], [13]. Lidar is also sensitive to the observation environment when imaging in bright daylight conditions [14], and through obscurants or turbid media, such as underwater [11], [15], or through fog, rain [1], [16]. The latter causes photon scattering which results in the immersion of the useful signal within a high and possibly non-uniform background level [17]–[19]. To obtain more detailed information about the observed target, one approach is to use multiple laser wavelengths which inevitably lead to larger data volumes which may necessitate the requirement for advanced algorithms to only select useful pixels [20], [21] or to account for shared data structures and correlations [22]–[25].

Several solutions have been proposed in the literature to tackle these challenges. We distinguish three broad families: statistical, learning-based and hybrid methods. The former builds on a statistical model and solves the resulting inference using stochastic simulation methods [6], [15], [26], [27], or optimization algorithms [19], [25], [28]. These principled methods benefit from a good interpretability but are subject to the definition of good features to represent the data. The second family learns important features from training data with an available ground-truth, and then uses the learned features to process new measured data [10], [29]–[31]. These approaches are dependent on the training data, and might require expensive network retraining if the imaging conditions change (e.g., different noise level). The third family uses a plug-and-play (PnP) approach [32] by combining methods of different families to improve performance [33], [34]. Beside providing good results, these methods can lack interpretability (e.g., in terms of convergence) and increasing interest is now devoted to providing principled PnP formulations as in [35], [36].

This paper combines the advantages of these families by proposing a principled statistical-based algorithm, that can use state-of-the-art algorithms as a guide for robust processing of multispectral 3D Lidar data acquired through obscurants. An approximate likelihood distribution is considered and a hierarchical Bayesian model is proposed to exploit the data Poisson statistics, the multi-scale information (known to improve noise

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCI.2021.3111572, IEEE Transactions on Computational Imaging

and photon-sparsity robustness [19], [25], [33], [37]), and prior knowledge on the depth and reflectivity maps. This hierarchical model ensures the robustness of the proposed strategy to the mismatch between the simplified observation model and the actual one. In contrast to the hierarchical Bayesian models in [6], [22] which required computationally expensive Markov chain simulation methods for parameter inference, the proposed formulation allows for independent parameter updates, leading to efficient parallel implementations and fast inference. This is achieved by introducing latent variables that are connected to the parameters of interest using Markov random fields, hence accounting for spatial correlations between pixels while ensuring parameter independence. Inspired by the PnP approaches that incorporate state-of-the-art denoisers [32], [34], we propose a weight-based model which uses the results of state-of-the-art algorithms as a guide to improve performance. The parameter's posterior distribution is obtained by combining the likelihood and proposed prior distributions. This distribution provides parameter estimates together with their uncertainties which are essential for result analysis and decision making. More precisely, we used a coordinate descent algorithm [38]–[40] to approximate the maximum a-posteriori estimator of all parameters, leading to simple iterative updates based on analytical or well known operators (e.g., weighted median filter [41], [42]). The new algorithm is tested on simulated and real underwater data showing promising results in terms of robustness to noise, interpretability and computational cost when compared to state-of-the-art algorithms.

The paper is structured as follows. Section II introduces the observation model and formulates the considered approximated likelihood. The proposed hierarchical Bayesian model is presented in Section III, and the choice of the guidance weights is described in details in Section IV. Section V then introduces the estimation algorithm used to approximate the maximum a-posteriori estimate of the parameters. Section VI analyses the proposed algorithm's performance when considering synthetic data with known ground-truth. Results on real data are presented in Section VII. Conclusions and future work are finally reported in Section VIII.

II. Problem formulation

This section introduces the observation model for multi-spectral Lidar, followed by the likelihood approximation used in this paper. The last part presents the multi-scale information which is a key ingredient to restore the parameters of interest.

A. Observation model

In addition to object reflectivity, the TCSPC Lidar system measures the depth profiles by illuminating the scene and measuring the time-of-flight of the returned photons. These photons are then collected in a histogram of counts, denoted $y_{n,t}$, and representing the received photon counts at pixel location $n \in \{1, \cdots, N\}$, and time-of-flight (ToF) bin $t \in \{1, \cdots, T\}$. In the case of multi-spectral imaging, the system illuminates the scene using K wavelengths leading to K histograms where each sample is denoted by $y_{n,t,k}$, with $k \in \{1, \cdots, K\}$. It is often assumed that the resulting histograms of counts follow a Poisson distribution $\mathcal{P}(.)$ as follows [6], [19], [25]:

$$y_{n,t,k} \sim \mathcal{P}\left(s_{n,t,k}\right) \tag{1}$$

where $s_{n,t,k}$ represents the average photon counts in the nth pixel, tth time bin and kth wavelength. In presence of at-most one target per-pixel, the signal can be approximated as follows

$$s_{n,t,k} = r_{n,k} f_k\left(t - d_n\right) + b_{n,t,k} \tag{2}$$

where f_k represents the system impulse response (SIR) of the kth wavelength, which can be measured during system calibration, $r_{n,k} \geq 0$ represents the reflectivity of the observed object assumed different for different wavelengths, $d_n \geq 0$ represents the object distance which is related to the object depth profile and assumed the same for all wavelengths (it is expressed in time bins or converted to meters using the speed of light c as follows $cd_n/2$), and $b_{n,t,k} \geq 0$ represents the background which gathers all photon events that do not originate from reflections at the target surface, i.e., the dark counts of the detector and the environment background due to the ambient illumination or photon scattering when imaging through obscurants. When imaging through turbid media, the background will have a non-uniform shape with respect to the depth observation timing window [17], [43], hence the dependence of b on t. Our goal is to estimate the depth and reflectivity parameters when imaging in extreme conditions due to imaging though obscurants (high and non-uniform background) or sparse photon imaging (e.g., rapid or long-range imaging).

B. Approximated Poisson likelihood

Assuming independence between the observed pixels $y_{n,t,k}$ leads to the joint likelihood

$$P(\boldsymbol{Y}|\boldsymbol{d}, \boldsymbol{R}, \boldsymbol{B}) = \prod_{n=1}^{N}\prod_{k=1}^{K}\prod_{t=1}^{T} \frac{s_{n,t,k}^{y_{n,t,k}}}{y_{n,t,k}!} \exp\left(-s_{n,t,k}\right) \tag{3}$$

where $\boldsymbol{d}$ is an $N \times 1$ vector gathering depth values, $\boldsymbol{R}$ is an $N \times K$ matrix gathering reflectivity values, and $\boldsymbol{B}, \boldsymbol{Y}$ are $N \times T \times K$ tensors of background values and photon counts, respectively. Let's assume the absence of background counts or the availability of a background rejection method to isolate signal counts (as introduced later in Section V-G). Assuming that $\sum_{t=1}^{T} f_k\left(t - d_n\right) = 1, \forall k$ for all realistic d_n, the likelihood reduces to (see Appendix for more details)

$$\begin{aligned} P({}^{s}\boldsymbol{y}_n|\boldsymbol{r}_n, d_n) \quad &\propto \quad \prod_{k=1}^{K}\left[\mathcal{G}\left(r_{n,k}; 1 + \bar{s}_{n,k}, 1\right)\bar{Q}\left({}^{s}\boldsymbol{y}_{n,k}\right)\right] \\ &\times \quad \prod_{t,k}\left[f_k\left(t - d_n\right)\right]^{{}^{s}y_{n,t,k}} \end{aligned} \tag{4}$$

where $\propto$ stands for proportional to, $\mathcal{G}(x; \gamma, \theta) \propto x^{\gamma-1}\exp(-x/\theta)$ denotes the gamma distribution with shape and scale parameters denoted γ, θ, ${}^{s}\boldsymbol{y}_{n,k}$ represents the histogram of target reflected (or signal) counts, $\bar{Q}\left({}^{s}\boldsymbol{y}_{n,k}\right)$ is a normalization constant that depends on the signal counts ${}^{s}\boldsymbol{y}_{n,k}$ (but not on the parameters of interest $\boldsymbol{r}_n, d_n$) and $\bar{s}_{n,k} = \sum_{t=1}^{T} {}^{s}y_{n,t,k}$ represents the sum of signal counts. It should be noted from (4) that the maximum likelihood (ML)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCI.2021.3111572, IEEE Transactions on Computational Imaging

estimate of the reflectivity at the nth pixel location for the kth wavelength is given by

$$r_{n,k}^{\mathrm{ML}} = \bar{s}_{n,k}. \tag{5}$$

Similarly, the depth maximum likelihood estimate is obtained using a simple log-matched filtering of the histogram with the SIR, as follows

$$d_n^{\mathrm{ML}} = \mathrm{argmax}_d \sum_{t,k} {}^{s}y_{n,t,k} \log[f_k(t-d)]. \tag{6}$$

It is common to approximate the SIR at each wavelength with the Gaussian function $f_k\left(\mu - d_n\right) \approx \mathcal{N}(d_n; \mu, \sigma_k^2)$ [28], [44]. In this case the likelihood in (4) becomes

$$\begin{aligned} P({}^{s}\boldsymbol{y}_n | \boldsymbol{r}_n, d_n) \quad &\propto\!\!\!\propto \quad \prod_{k=1}^{K} \left[\mathcal{G}\left(r_{n,k}; 1+\bar{s}_{n,k}, 1\right) \bar{Q}\left({}^{s}\boldsymbol{y}_{n,k}\right)\right] \\ &\times \quad \mathcal{N}(d_n; d_n^{\mathrm{ML}}, \bar{\sigma}^2) \end{aligned} \tag{7}$$

where $\propto\!\!\!\propto$ stands for approximately proportional to, $\mathcal{N}(x; \mu, \sigma_k^2)$ represents the Gaussian distribution with average μ and variance σ_k^2, $\bar{\sigma}^2 = \left(\sum_k \frac{\bar{s}_n}{\sigma_k^2}\right)^{-1}$ and $d_n^{\mathrm{ML}} = \bar{\sigma}^2 \sum_{k=1}^{K} \frac{\sum_{t=1}^{T} t\, {}^{s}y_{n,t,k}}{\sigma_k^2}$ is given analytically when considering Gaussian approximation for the SIR. Considering these approximations, Eq. (7) indicates that the depth and reflectivity parameters are independent and that they appear within conventional Gaussian and gamma distributions, which is crucial for the design of the proposed Bayesian strategy. Indeed, the quality of the ML depth and reflectivity estimators is known to be poor in challenging scenarios, hence the need to account for known parameter properties to improve reconstruction. This can be done within the Bayesian framework adopted in this paper.

C. Multiscale information

A common approach to improve the performance of maximum likelihood estimation for Lidar data is to consider multiscale information, as already exploited in several state-of-the-art 3D Lidar denoising algorithms [19], [25], [33], [37]. The key observation is that spatially downsampled histograms, which are still Poisson distributed, lead to depth and reflectivity estimates with lower noise at a price of a reduced spatial resolution, and the potential to mix histograms of objects at different depths. In this paper, we adopt a similar strategy by considering L downsampled versions of the histogram of counts. For each wavelength k, spatially downsampled version of the histograms $\boldsymbol{Y}$ are first computed based on predefined L graphs of neighbours $\phi^{1,\cdots,L}$ leading to $\boldsymbol{Y}_k^{\ell}$ (for example, $q^{(2)} = 3\times 3$ neighbours for $\phi^{(2)}$, and $q^{(3)} = 5\times 5$ neighbours for $\phi^{(3)}$, $\cdots$). The latter can be efficiently computed using convolutions in the case of a regular grid but our algorithm can be equally applied to a non-uniform sampling grid of the pixels. Assuming independence between these histograms leads to L likelihood distributions as follows

$$\begin{aligned} P({}^{s}\boldsymbol{y}_n^{(\ell)} | \boldsymbol{r}_n^{(\ell)}, d_n^{(\ell)}) \quad &\propto\!\!\!\propto \quad \prod_{k=1}^{K} \left[\mathcal{G}\left(r_{n,k}^{(\ell)}; 1+\bar{s}_{n,k}^{(\ell)}, 1\right) \bar{Q}\left({}^{s}\boldsymbol{y}_{n,k}^{(\ell)}\right)\right] \\ &\times \quad \mathcal{N}\left(d_n^{(\ell)}; d_n^{\mathrm{ML}(\ell)}, \left(\bar{\sigma}^{(\ell)}\right)^2\right) \end{aligned} \tag{8}$$

$\forall \ell \in 1, \cdots, L$, where $\left(\bar{\sigma}^{(\ell)}\right)^2 = \left(\sum_k \frac{\bar{s}_n^{(\ell)}}{\sigma_k^2}\right)^{-1}$, $\ell = 1$ is the original cube, and for example, $\ell = 2$ corresponds to a 3×3 downsampling, $\ell = 3$ to a 5×5 downsampling, etc.

III. Hierarchical Bayesian model

Estimating depth and reflectivity parameters in extreme conditions is an ill-posed problem which requires the use of prior information to alleviate its indeterminacy. A Bayesian strategy is considered to combine the approximate likelihood described above, with parameter prior distributions accounting for known parameter properties. The resulting posterior distribution will be exploited by deriving Bayesian point estimators and additional measures of uncertainty about the estimates. The following sub-sections introduce the proposed Bayesian model.

A. Prior distribution for depth

Our model assumes the observation of L depth maps $\boldsymbol{d}^{(\ell)}$ obtained from multi-scale downsampled histograms, and having different noise levels as highlighted by the Gaussian variances in (8). Object depth profiles exhibit homogeneous surfaces (i.e., spatial correlation) separated by a discontinuous jump between different surfaces. This requires enforcing spatial correlation between the pixels of a surface, while preserving edges of isolated objects or between separated surfaces. To incorporate this information, we introduce an $N\times 1$ latent variable $\boldsymbol{x}$ that is connected to all multi-scale depth maps, to provide a robust reconstruction of the true depth map by considering correlations between pixels. To preserve edges separating different surfaces, we propose the following mixture of Laplace conditional prior distributions for $\boldsymbol{x}$ as follows

$$\begin{aligned} &x_n | d_n^{(1,\cdots,L)}, w_{\nu_n,n}^{(1,\cdots,L)}, \epsilon_n \sim \\ &\prod_{n'\in\nu_n} \left[\prod_{\ell=1}^{L} \mathcal{L}\left(x_n; d_{n'}^{(\ell)}, \frac{\epsilon_n}{w_{n',n}^{(\ell)}}\right)\right] \end{aligned} \tag{9}$$

where $\mathcal{L}(x; \mu, \epsilon) = 1/(2\epsilon)\exp\left(-|x-\mu|/\epsilon\right)$ represents the Laplace distribution with average μ and diversity parameter ϵ, ν_n represents the spatial neighbourhood of the nth pixel, $d_{n'}^{(\ell)}$ denotes the mean, $\epsilon_n > 0$ is the variance of $\boldsymbol{x}_n$ and $w_{n',n}^{(\ell)} \geq 0$ are constant weights to be defined. Note that (9) preserves edges as it considers the sparsity promoting ℓ_1-norm of the differences between $\boldsymbol{x}$ and $\boldsymbol{D} = \left[\boldsymbol{d}^{(1)}, \cdots, \boldsymbol{d}^{(L)}\right]$. The weights $w_{n',n}^{(\ell)} \geq 0$ are essential as they allow guiding the connections between $\boldsymbol{x}$ and $\boldsymbol{D}$ using any available side-information (e.g., obtained from other sensors in the case of multi-modal imaging, or by using state-of-the-art denoising algorithms in the case of plug-and-play approaches). It is also worth noting that prior (9) is connected to the Bayesian lasso model [45], [46]. Indeed, (9) could be obtained by marginalizing the exponentially-distributed variance hyperparameter of a Gaussian mixture prior. Finally, (9) does not enforce positivity on the depth parameter $\boldsymbol{x}$, however, this will be ensured as indicated in Section V-B.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCI.2021.3111572, IEEE Transactions on Computational Imaging

B. Prior distribution for reflectivity

In a similar fashion to depth, spatial smoothness can be enforced on the reflectivity by considering latent variables as in the gamma Markov random field prior [47]. However, this prior will lead to underestimated reflectivity values as already highlighted in [27]. In this work, we introduce an $N \times K$ latent variable $\boldsymbol{M}$ assigned a Gaussian prior distribution as follows

$$m_{n,k}|r_{\nu_n,k}^{(1,\cdots,L)}, v_{\nu_n,n}^{(1,\cdots,L)}, \psi_{n,k}^2 \sim \prod_{n' \in \nu_n} \left[\prod_{\ell=1}^{L} \mathcal{N}\left(r_{n',k}^{(\ell)}, \frac{\psi_{n,k}^2}{v_{n',n;k}^{(\ell)}} \right) \right] \tag{10}$$

where $v_{n',n;k}^{(\ell)} \geq 0$ are constant weights to be defined, and $\psi_{n,k}^2$ represents the variance of the latent variable and contains reflectivity uncertainty information for the kth wavelength. The $N \times 1$ latent variable associated with the kth wavelength, denoted $\boldsymbol{m}_k$, contains reflectivity information through its relation to $r_{n,k}^{(\ell)}$ and will serve as the reflectivity estimate for the kth wavelength.

Although this is not a conjugate prior, it will lead to non-negative analytical estimates for $\boldsymbol{M}, \boldsymbol{R}$ as indicated in Section V.

C. Priors of the variance hyperparameters

The variance parameters $\epsilon_n, \forall n$ (resp. $\psi_{n,k}^2, \forall n, k$) should be positive. Assuming prior independence between the parameters $\epsilon_n, \forall n$ (resp. $\psi_{n,k}^2, \forall n, k$) and accounting for their positivity, we assign a conjugate inverse gamma distribution for these parameters as follows

$$\begin{aligned} f(\boldsymbol{\epsilon}) &= \prod_{n=1}^{N} \mathcal{IG}(\epsilon_n; \alpha_d, \beta_d) \\ f(\boldsymbol{\Psi}) &= \prod_{k=1}^{K} \prod_{n=1}^{N} \mathcal{IG}\left(\psi_{n,k}^2; \alpha_r, \beta_r\right) \end{aligned} \tag{11}$$

where $\alpha_r, \beta_r, \alpha_d, \beta_d$ are positive user fixed hyperparameters. In absence of additional knowledge, these hyperparameters are fixed to obtain a non-informative prior.

D. Posterior distribution

The joint posterior distribution of this Bayesian model can be computed from the following hierarchical structure (after dropping indices for clarity)

$$\begin{aligned} f(\boldsymbol{x}, \boldsymbol{D}, \boldsymbol{M}, \boldsymbol{R}, \boldsymbol{\epsilon}, \boldsymbol{\Psi}|\boldsymbol{Y}) \propto f(\boldsymbol{Y}|\boldsymbol{R}, \boldsymbol{D}) f(\boldsymbol{D}, \boldsymbol{x}|\boldsymbol{\epsilon}, \boldsymbol{W}) \\ f(\boldsymbol{R}, \boldsymbol{M}|\boldsymbol{\Psi}, \boldsymbol{V}) f(\boldsymbol{\epsilon}) f(\boldsymbol{\Psi}) \end{aligned} \tag{12}$$

where $\boldsymbol{W}$ and $\boldsymbol{V}$ gather the depth and reflectivity non-negative weights, respectively. $f(\boldsymbol{Y}|\boldsymbol{R}, \boldsymbol{D})$ is given in (8), $f(\boldsymbol{D}, \boldsymbol{x}|\boldsymbol{\epsilon}, \boldsymbol{W})$ in (9), $f(\boldsymbol{R}, \boldsymbol{M}|\boldsymbol{\Psi}, \boldsymbol{V})$ in (10), and $f(\boldsymbol{\epsilon}), f(\boldsymbol{\Psi})$ in (11). This distribution contains complete information regarding the parameters of interest $\boldsymbol{x}, \boldsymbol{D}, \boldsymbol{R}, \boldsymbol{M}$ and their uncertainties $\boldsymbol{\epsilon}, \boldsymbol{\Psi}$. A common approach is to extract Bayesian point estimators such as the maximum a-posteriori (MAP) estimator or the minimum mean square estimator (MMSE). In this paper, we consider the MAP estimator of all parameters. It should be noted that the depth related parameters $\boldsymbol{D}, \boldsymbol{x}, \boldsymbol{\epsilon}$ and the reflectivity ones $\boldsymbol{R}, \boldsymbol{M}, \boldsymbol{\Psi}$ are independent allowing parallel optimization with respect to both set of parameters. Finally, Fig. 1 presents a directed acyclic graph (DAG) which summarizes the main parameters of the proposed hierarchical Bayesian model.

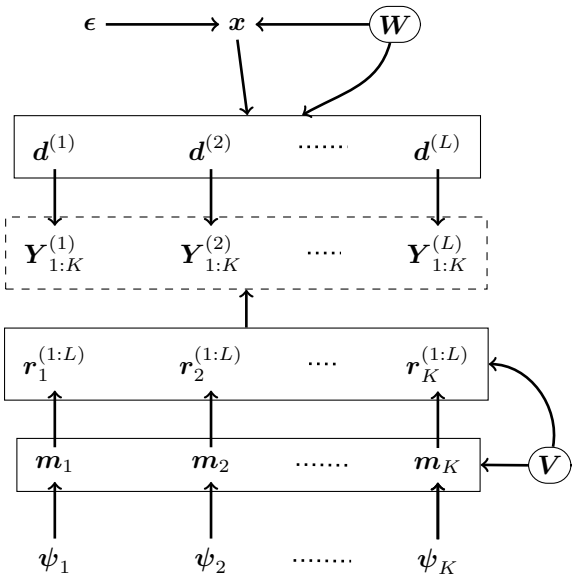

Fig. 1. DAG for the observations, parameter and hyperparameters of the proposed model. The guiding weights appear in circles, the observations in a dashed box and the rectangular box gathers the joint multi-scale or multi-wavelengths parameters.

IV. Incorporating guidance using weights selection

The choice of the weights is very important and will have a direct impact on the algorithm performance. Several strategies have been considered in the literature where the choice can be based on the spatial distance between points, similarity of their values, etc [48]–[50]. In this paper, we assume the presence of guiding information (e.g., by using other algorithms, or sensors) and define these weights while considering multi-scale and multi-wavelength information.

A. Depth weights $\boldsymbol{W}$

Assuming the presence of an outlier free multi-scale guiding depth $\underline{\boldsymbol{d}}^{(\ell)}, \ell = 1, \cdots, L$, our selection of the multi-scale weights $\boldsymbol{W}$ encourages the depth map at a given scale $\boldsymbol{d}^{(\ell)}$ to be close to $\underline{\boldsymbol{d}}^{(\ell)}$ (with a graph of neighbours, for example 3×3 neighbours in a uniform grid). More precisely, we assign low weights for pixels that differ significantly from their corresponding pixels in $\underline{\boldsymbol{d}}^{(\ell)}$ as follows

$$\begin{aligned} w_{n,n'}^{(\ell)} &= w_{\text{norm}} \left[\prod_{\ell'=1}^{\ell-1} \left(1 - w_{n,n'}^{(\ell')} \right) \right] \\ &\times \exp\left(- \frac{|d_n^{\text{ML}(\ell)} - \underline{d}_{n'}^{(\ell)}|}{2\zeta q^{(\ell)}} \right) \end{aligned} \tag{13}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCI.2021.3111572, IEEE Transactions on Computational Imaging

for $\ell \in \{1, \cdots, L\}$, where w_{norm} is a normalization constant ensuring $\sum_{\ell,n'} w_{n,n'}^{(\ell)} = 1$, the coefficient ζ is easily fixed based on physical considerations related to the impulse response width and it is weighted by the downsampling coefficient $q^{(\ell)} = (2\ell - 1) \times (2\ell - 1)$ to account for the multi-resolution effect. In (13), the product over ℓ' promotes higher weights for lower ℓ, i.e., a high $w_{n,n'}^{(\ell')}$ will enforce low values for $w_{n,n'}^{(\ell)}$ with $\ell > \ell'$.

We are now left with the task of finding a reliable multi-scale depth guide which is robust to outliers. This information can be obtained by considering other sensing modalities such as Radar, Sonar, when available. It can also be obtained by applying an off-the-shelf depth reconstruction algorithm to the Lidar data (e.g., [19]). The latter strategy is adopted in this paper. We consider two methods, the first, denoted GD1 for guide depth 1, is inspired by [5] which adopted the rank order mean approach to unmix signal from background counts. Here, we first detect background corrupted pixels (those without $\sqrt{q^{(2)}}$ neighbours having close depth values) and then replace them with the median of surrounding valid points. The second strategy, denoted GD2, represents $\boldsymbol{d}^{\text{ML}(\ell=1:L)}$ as a point cloud and applies an outlier rejection algorithm to remove corrupted values (i.e., using *pcdenoise* in Matlab [51]). We note finally that the weights could be updated with iterations leading to a pseudo-Bayesian approach [52], but this is out of the scope of this paper and will be left for future work.

B. Reflectivity weights $\boldsymbol{V}$

The reflectivity weights are obtained from the multi-scale images $\boldsymbol{r}_k^{(\ell)}, \forall k, \ell$, but we note that they can also be learned using additional reflectivity maps acquired by other sensors when available. Assuming the presence of $\underline{\boldsymbol{r}}_k^{(\ell)}, \forall k, \ell$ reflectivity guides and depth weights $\boldsymbol{W}$, we consider a multi-scale bilateral filtering approach [34], [49], [53] and define the reflectivity weights as follows

$$v_{n,n';k}^{(\ell)} = v_{\text{norm}}\, w_{n,n'}^{(\ell)} \exp\left(-\frac{|r_{n,k}^{\text{ML}(\ell)} - \underline{r}_{n',k}^{(\ell)}|}{2\eta_{n,k} q^{(\ell)}}\right) \tag{14}$$

where v_{norm} is a normalization constant ensuring $\sum_{\ell,n'} v_{n,n';k}^{(\ell)} = 1$, and $\eta_{n,k}$ is a constant weighted by the downsampling coefficient $q^{(\ell)}$. As indicated in (14), correlation between depth and reflectivity images is introduced through the use of $\boldsymbol{W}$ to define $\boldsymbol{V}$. This will promote close points in space having similar depths to share similar reflectivities, in addition to exploit the multiscale depth guidance information to reject or mitigate the effect of measured outliers in both $\boldsymbol{D}$ and $\boldsymbol{R}$. Note that reflectivity texture will be preserved by considering the $\boldsymbol{R}$ dependent exponential term in (14). The reflectivity variables $\boldsymbol{r}_k^{(\ell)}, \forall k, \ell$, follow a gamma distribution and hence show data dependent noise levels. To account for this effect, we assume a signal dependent variance $\eta_{n,k}$, which is fixed as follows

$$\eta_{n,k} = \max\left(0.1, r_{n,k}^{\text{ML}(L)}\right). \tag{15}$$

Several reflectivity restoration algorithms can be used to obtain the guides $\underline{\boldsymbol{r}}_k^{(\ell)}, \forall k, \ell$, based on the considered imaging scenarios. Algorithms based on Poisson statistics can be used in the sparse photon regime [28], [54], [55], while other state-of-the-art denoising algorithms [56], [57] can be considered in dense photon regimes. In this paper, we consider three guidance methods. The first guidance intensity (denoted GI1) considers $\underline{\boldsymbol{r}}_k^{(\ell)} = \boldsymbol{r}_k^{\text{ML}(\ell)}, \forall k, \ell$ which leads to a multi-scale generalization of the bilateral filter. Indeed, these multi-scale maps already contain filtering properties which will provide good performance in practice. The second guidance (GI2) considers the Poisson based reconstruction method [54] (used with authors defaults parameters) which is applied to each scale and wavelength of $\boldsymbol{r}_k^{\text{ML}(\ell)}, \forall k, \ell$ to obtain $\underline{\boldsymbol{r}}_k^{(\ell)}, \forall k, \ell$. As a third guidance (GI3), we considered the state-of-the-art learning based DnCNN denoiser [57], also applied to each scale and wavelength $\boldsymbol{r}_k^{\text{ML}(\ell)}, \forall k, \ell$. Finally, note that reflectivity multi-spectral correlations are introduced through the depth weights, which are shared between all wavelengths. Additional correlations can be easily included through the weights $\boldsymbol{V}$ when building the reflectivity guides.

V. Estimation algorithm

We propose to use the MAP estimators for all parameters and hyperparameters $\boldsymbol{x}, \boldsymbol{D}, \boldsymbol{R}, \boldsymbol{M}, \boldsymbol{\epsilon}, \boldsymbol{\Psi}$. More precisely, the maximum of the posterior distribution in (12) is approximated using a coordinate descent algorithm [38], [39]. This algorithm sequentially maximizes the conditional distributions associated with each parameter until convergence to a local minimum of the negative log-posterior. The algorithm's main steps are presented in Algo. 1 and described with more details in the following sections. Note that the resulting depth updates alternates between robust to outliers non-linear parameter estimation (line 11) and a filtering step (line 12), which are commonly observed steps in several state-of-the-art algorithms [32], [34], [35] and optimization algorithms [58]. Note also that reflectivity and depth iterates are independent and can be run in parallel. Note finally that reflectivity updates are analytically obtained ensuring fast estimation.

A. Updating $\boldsymbol{x}$

The parameters of $\boldsymbol{x}$ are independent allowing parallel updating of $x_n, \forall n$. It is clear from (12) that the conditional distribution of $\boldsymbol{x}$ results from (9). Minimizing the negative-log of the conditional distribution reduces to

$$\hat{x}_n = \underset{x}{\text{argmin}}\, \mathcal{C}(x) = \underset{x}{\text{argmin}} \sum_{\ell,n' \in \nu_n} w_{n',n}^{(\ell)} |x - d_{n'}^{(\ell)}|. \tag{16}$$

This is a weighted median filter (WMF) which has several efficient implementations (e,g, [41]). Note that the solution of (16) will be non-negative provided that $d_{n'}^{(\ell)} \geq 0$, which is ensured during initialization.

B. Updating $\boldsymbol{D}$

The variables $d_1^{(\ell)}, \cdots, d_N^{(\ell)}$ are independent and spatial correlation is introduced through the latent variable $\boldsymbol{x}$. This is interesting as it allows the parallel implementation of $d_n^{(\ell)}, \forall n, \ell$ with respect to n and ℓ. The conditional distribution

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCI.2021.3111572, IEEE Transactions on Computational Imaging

Algorithm 1 Estimation algorithm

Input:
$\boldsymbol{Y}_k, \forall k$; L; $\phi^{1,\cdots,L}$
Generate low resolution data:
Generate low-resolution histograms $\boldsymbol{Y}_k^{(\ell)}, \ell \in \{1,\cdots,L\}$ using $\phi^{1,\cdots,L}$
Estimate background level $\hat{\boldsymbol{B}}_k, \forall k$ as in (27)
Estimate $\boldsymbol{d}^{\mathrm{ML}(\ell)}, \boldsymbol{r}_k^{\mathrm{ML}(\ell)}, \forall k$ as in (6), (5)
Compute guiding weights $\boldsymbol{W}, \boldsymbol{V}$ as in (13), (14)
Coordinate descent algorithm
while conv$= 0$ **do**
  Update $x_n, \forall n$ using WMF in (16)
  Update $\boldsymbol{d}^{(\ell)}, \forall \ell$ using threshold operator in (17)
  Update $\boldsymbol{\epsilon}$ using analytical mode in (19)
  Update $\boldsymbol{m}_k, \forall k$ using analytical mode in (20)
  Update $\boldsymbol{r}_k^{(\ell)}, \forall \ell$ using analytical mode in (22)
  Update $\boldsymbol{\Psi}$ using analytical mode in (24)
  Set conv$= 1$ if the convergence criteria are satisfied
end while
Output:
$\boldsymbol{x}, \boldsymbol{M}, \boldsymbol{\epsilon}, \boldsymbol{\Psi}$

of $\boldsymbol{D}$ is obtained by combining the likelihood in (8), and the prior in (9). Straightforward computations show that the update $d_n^{(\ell)}$ is given by

$$d_n^{(\ell)} = \underset{d}{\operatorname{argmin}} \frac{\left[d - d_n^{\mathrm{ML}(\ell)}\right]^2}{2\left(\bar{\sigma}^{(\ell)}\right)^2} + \sum_{n' \in \nu_n} \frac{w_{n,n'}^{(\ell)} |d - x_{n'}|}{\epsilon_{n'}^2}. \quad (17)$$

This is a generalization of the well known soft-threshold operator which can be analytically solved as in [59]. Note that the solution of (17) will be non-negative provided that $x_{n'} \geq 0$ and $d_n^{\mathrm{ML}(\ell)} \geq 0$ which is ensured during initialization.

C. Updating depth variance: $\boldsymbol{\epsilon}$

The conditional distribution of ϵ_n is an inverse-gamma distribution given by

$$\epsilon_n | \boldsymbol{x}, \boldsymbol{D}, \boldsymbol{W} \sim \mathcal{IG}\left[L + \bar{N} + \alpha_d, \mathcal{C}\left(x_n\right) + \beta_d\right] \quad (18)$$

whose mode is given by

$$\hat{\epsilon}_n = \frac{\mathcal{C}\left(x_n\right) + \beta_d}{L + \bar{N} + \alpha_d + 1} \quad (19)$$

where $\bar{N}$ is the number of spatial neighbours.

D. Updating $\boldsymbol{M}$

It is clear from (12) that the conditional distribution of $\boldsymbol{M}$ results from (10). This is a normal distribution whose mean is analytically given by

$$\hat{m}_{n,k} = \frac{\sum_{\ell, n' \in \nu_n} v_{n',n;k}^{(\ell)} \, r_{n',k}^{(\ell)}}{\sum_{\ell, n' \in \nu_n} v_{n',n;k}^{(\ell)}}. \quad (20)$$

This equation highlights a weighted sum of the multi-scale reflectivity maps $\boldsymbol{r}$, as for the bilateral filter.

E. Updating $\boldsymbol{R}$

The parameters of $\boldsymbol{R}$ are independent allowing parallel updating of $r_{n,k}^{(\ell)}, \forall n, k, \ell$. The conditional distribution of $\boldsymbol{R}$ is obtained by combining the likelihood in (8), and the prior in (10). Minimizing the negative-log of the conditional distribution reduces to

$$\hat{r}_{n,k}^{(\ell)} = \underset{r}{\operatorname{argmin}} \left\{ r - \bar{s}_{n,k}^{(\ell)} \log r + \mathcal{H}(r) \right\} \quad (21)$$

where $\mathcal{H}(r) = \frac{1}{2\psi_r}\left(r - \mu_r\right)^2$ with $\psi_r^{-1} = \sum_{n'} \frac{v_{n',n;k}^{(\ell)}}{\psi_{n',k}}$ and $\mu_r = \sum_{n'} \left(\frac{v_{n,n';k}^{(\ell)} m_{n',k}}{\psi_{n',k}} \right)$. The minimum is analytically provided by [60]

$$\hat{r}_{n,k}^{(\ell)} = \frac{\mu_r - \psi_r + \sqrt{\left(\mu_r - \psi_r\right)^2 + 4\psi_r \bar{s}_{n,k}^{(\ell)}}}{2}. \quad (22)$$

F. Updating $\boldsymbol{\Psi}$

The conditional distribution of the reflectivity variance $\psi_{n,k}$ is an inverse-gamma distribution given by

$$\psi_{n,k} | \boldsymbol{M}, \boldsymbol{R}, \boldsymbol{V} \sim \mathcal{IG}\left[\frac{L + \bar{N}}{2} + \alpha_r, \mathcal{K} + \beta_r\right] \quad (23)$$

with $\mathcal{K} = \sum_{\ell, n' \in \nu_n} \frac{v_{n',n,k}^{(\ell)} \left(m_{n,k} - r_{n',k}^{(\ell)}\right)^2}{2}$. The mode is analytically given by

$$\hat{\psi}_{n,k} = \frac{\mathcal{K} + \beta_r}{\frac{L+\bar{N}}{2} + \alpha_r + 1}. \quad (24)$$

G. Background estimation

Our algorithm assumes known signal counts, which can be obtained after removing background counts from observed histograms. In the presence of obscurants, the background can be non-uniform $b_{n,t,k}$, i.e., in addition to pixels and wavelengths it also depends on time bins related to the depth dimension. Assuming a spatially homogeneous distribution of the obscurant, the background level can be assumed smooth. This means that after downsampling, $y_{n,t,k}^{(L)}$ can be represented by the sum of a smooth function $\hat{b}_{n,t,k}$ and a sparse signal due to target reflections. Unmixing these two signals is a common signal processing problem and can be solved using several tools, e.g., Robust PCA [61]. In this paper, we only require an approximative estimate of $\hat{b}_{n,t,k}$ and are more interested in efficient solutions. More precisely, we assume the background has the same temporal shape for all pixels and estimate this shape as follows

$$\bar{b}_{t,k} = \operatorname{median}\left(y_{\sqcap_n,t,k}^{(L)}\right) \quad (25)$$

where $\sqcap_n$ represent the indices of the lowest 10% values of $y_{n,t,k}^{(L)}$ to only consider background and reject signal returns. For a given time bin, this strategy assumes that at least 10% of pixels only contain background without a target, which is often satisfied except when observing a perfectly lateral plane having the same depth value for all pixels. Akin to [62], the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCI.2021.3111572, IEEE Transactions on Computational Imaging

noise level of each pixel is estimated using the median as follows

$$\underline{b}_{n,k} = \text{median}\left(y^{(L)}_{n,:,k}\right). \tag{26}$$

The smooth background is then obtained by

$$\hat{b}_{n,t,k} = \max\left(0, \underline{b}_{n,k} + \bar{b}_{t,k} - \bar{\bar{b}}_k\right) \tag{27}$$

with $\bar{\bar{b}}_k = \sum_t \bar{b}_{t,k}/T$. Knowing the background level, the approximate signal counts can be extracted as follows

$${}^{s}y^{(\ell)}_{n,t,k} = \max(y^{(\ell)}_{n,t,k} - \hat{b}_{n,t,k},\, 0), \forall n \tag{28}$$

for $t \in [t_l, t_h]$; where $t_l = \max(1, d_n^{\text{ML}(\ell)} - I_k^l)$, $t_h = \min(T, d_n^{\text{ML}(\ell)} + I_k^r)$, where I_k^l and I_k^r represent the attack and trailing width of the kth SIR.

H. Stopping criteria

Two criteria are considered to stop the iterative coordinate decent algorithm for depth and reflectivity. The first is maximum number of iterations. The second evaluates the estimated parameter values and stops the algorithm if the relative difference between successive iterates is smaller than a threshold as in [63]

$$\left\|\boldsymbol{x}^{(i+1)} - \boldsymbol{x}^{(i)}\right\|_1 \leq \xi\left(\left\|\boldsymbol{x}^{(i)}\right\|_1 + \xi\right). \tag{29}$$

where i denotes the algorithm iterations and $\xi = 0.001$ is a threshold.

VI. Results on simulated data

This section evaluates the proposed algorithm on simulated data. The section first introduces comparisons algorithms and evaluation criteria. Then we analyse the robustness of the proposed algorithm with respect to sparse and high-background regimes and compare it on a single-wavelength 3D Lidar data. Finally, we generalize the analysis to multiple wavelengths scenarios. All simulations have been performed on a Matlab R2020a on a computer with Intel(R) Core(TM) i7-4790 CPU@3.60GHz and 32GB RAM.

A. Comparison algorithms and evaluation criteria

To highlight the robustness and benefit of the proposed algorithm, it is compared to several state-of-the-art algorithms including:

- The unmixing algorithm (UA) [19]: considers multi-scale information for robust reconstruction of depth and reflectivity images. It assumes the presence of one surface on all pixels, and is used when analysing robustness to noise and photon-sparse regime imaging for one surface scenes on single spectral data.
- The RT3D algorithm [34]: assumes the presence of multiple surfaces per-pixel and is used when analysing robustness to noise and photon-sparse regime imaging on single spectral data.
- The MUSAPOP algorithm [27]: assumes the presence of multiple surfaces per-pixel and is used when analysing multi-spectral Lidar data.
- The MNR3D algorithm [25]: is used when analysing multi-spectral Lidar data. Note that we post-processed the algorithm outputs (a depth map for each wavelength) to obtain one depth map for all wavelengths. This is done by capturing the position of the maximum of the sum of cleaned cubes.
- The Classical algorithm (denoted Class.): estimates $\boldsymbol{d}^{\text{ML}}, \boldsymbol{r}_k^{\text{ML}}, \forall k$ as in (6), (5) from the observed histograms (without removing background)
- The B-Class. algorithm: removes estimated background level as in (27), then estimates $\boldsymbol{d}^{\text{ML}}, \boldsymbol{r}_k^{\text{ML}}, \forall k$ as in (6), (5), respectively (see lines 5-6 in Algo. 1).

Comparison results will be analysed qualitatively (by showing reconstruction scenes) and quantitatively using several criteria. The depth performance is measured based on the depth absolute error (DAE) measure DAE= $\frac{1}{N'}\left\|\boldsymbol{d}^{\text{ref}} - \boldsymbol{d}^{\text{est}}\right\|_1$, where N' represents the number of pixels having a target, and $\boldsymbol{d}^{\text{ref}}$ and $\boldsymbol{d}^{\text{est}}$ are the reference and estimated depth maps with a target, respectively. Similarly, intensity is evaluated using the intensity normalized absolute error IAE= $\frac{\left\|\boldsymbol{r}^{\text{ref}} - \boldsymbol{r}^{\text{est}}\right\|_1}{\|\boldsymbol{r}^{\text{ref}}\|_1}$. In addition, we consider the metrics used in [27] to evaluate point clouds. More precisely, we consider the percentage of true detections as a function of the distance τ, where a true detection occurs if an estimated point of a given nth pixel has a reference point in its surrounding such that $|\hat{d}_n - d_n^{\text{ref}}| \leq \tau$. The sum of the estimated points that can not be assigned to any true point at a distance of τ are considered as false detections. Average normalized IAE is considered for intensity, where pixels with no or false detections are assumed to introduce an error of $\frac{r_n^{\text{ref}}}{\|\boldsymbol{r}^{\text{ref}}\|_1}$.

B. Robustness to sparsity or background counts

This section evaluates the algorithm performance under different cases, including the photon sparse regime (low average photon-per-pixels) and low signal-to-background ratio (SBR), where average SBR= $\frac{\sum_{k=1}^{K}\sum_{n=1}^{N} r_{n,k}}{\sum_{k=1}^{K}\sum_{n=1}^{N}\sum_{t=1}^{T} b_{n,t,k}}$. The background estimation strategy is also evaluated when comparing the results of Class. and B-Class. algorithms. Simulations are performed on the Art scene extracted from the Middlebury dataset[1], as it is a cluttered scene used to evaluate many algorithms [19], [64] (see Fig. 2 (a)). An intensity image is first constructed using the luminance of the RGB image. The 283×183 depth and intensity images are then used to generate a 20ps time bin histograms of counts as in (1), while considering a real system impulse response (leading-edge of 3 bins and trailing-edge of 26 bins). The resulting cube of histograms is of size 283×183 pixels and $T = 300$ time bins. To investigate several scenarios, we generate multiple histogram cubes by varying the SBR ratio logarithmically in $[0.01, 100]$ and the average photons-per-pixel (PPP) in $[0.1, 1000]$ (this PPP combines signal and background counts, useful signal counts can be deduced from the PPP and SBR values). In addition, we consider two background shapes, a conventional uniform shape (i.e., $b_{n,t,k} = b_{n,k}$) where the background level

[1]Available in: http://vision.middlebury.edu/stereo/data/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCI.2021.3111572, IEEE Transactions on Computational Imaging

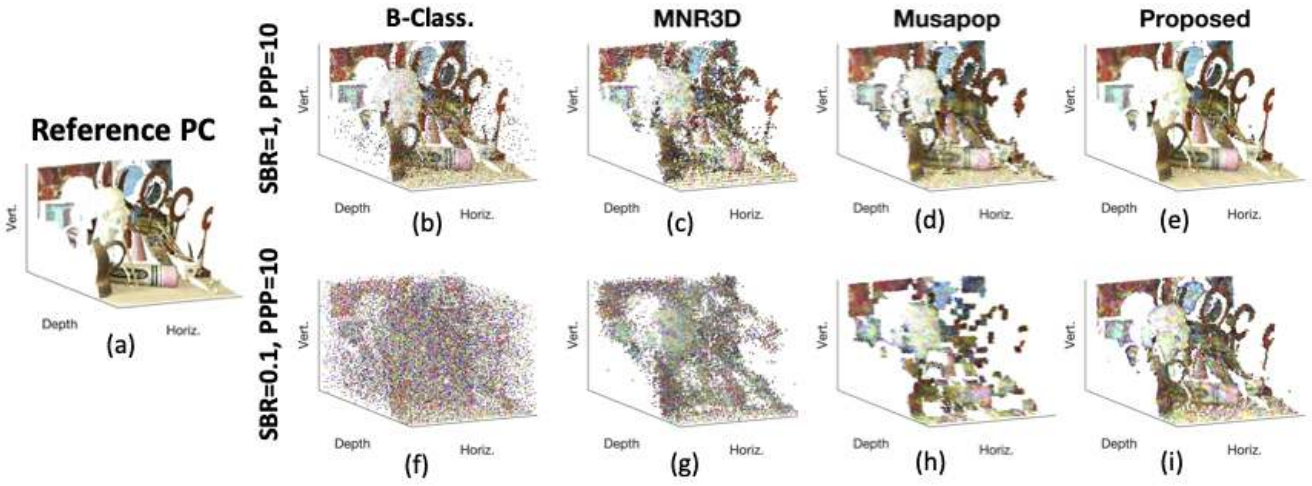

Fig. 2. 3D representations of the art scene for two levels of SBR and PPP obtained with uniform background. (a) Reference point cloud, (b-f) B-Class., (c-g) MNR3D [25], (d-h) MUSAPOP [27], (e-i) proposed algorithm. (b-c-d) SBR=1, PPP=10 per wavelength, (e-f-g) SBR=0.1, PPP=10 per wavelength.

is the same for all time bins, and a gamma shaped background (i.e., $b_{n,t,k} \sim \mathcal{G}(\alpha, \beta)$ where $\mathcal{G}$ denotes a gamma distribution with parameters $\alpha = 2$ and $\beta = 30$) often encountered when imaging through obscurants (such as underwater or through fog [17]). The proposed algorithm is considered with the following parameters $L = 3$ with $q^{(2)} = 3\times3$ and $q^{(3)} = 9\times9$, $\zeta = 2.7$cm (i.e., 9 time bins), while considering the first depth and intensity guides (GD1 and GI1). It is compared with the Classical and B-Class. algorithms (matched filter before and after removing non-uniform background), and the robust UA algorithm whose depth and intensity regularization parameters were tuned to provide best DAE performance. RT3D is not considered here as the scene is only composed of one surface per pixel. Fig. 3 shows the log scale DAE performance of the considered algorithms when considering uniform (left column) and gamma shaped backgrounds (right column). All algorithms show good results for high SBR and PPP and the performance degrades when decreasing SBR or PPP or when considering a non-uniform background. The proposed algorithm is more robust as it shows the lowest DAE even for extreme cases (DAE$\approx$ 0.01 for SBR=1 and PPP=1 photons). The UA algorithm presents second best results, and shows robust results. However, performance is slightly reduced for high SBR and PPP levels due to considering a Gaussian IRF instead of the asymmetric one used to simulate the data. The B-Class. algorithm is more robust than Class. which highlights the importance of the background removal step (consequently, we will only compare with B-Class. without showing Class. in following sections of the paper). Fig. 4 shows similar behaviours when considering the recovered intensity images, i.e., best robustness by the proposed algorithm followed by the UA algorithm. While all algorithms perform well for high SBR and PPP levels, it is worth noting that UA presented best IAEs in this case although its results tend to be over-smoothed (see Fig. 5).

In addition to depth and intensity maps, the proposed algorithm also provides their corresponding uncertainty maps (variance of the estimates), which help with decision making.

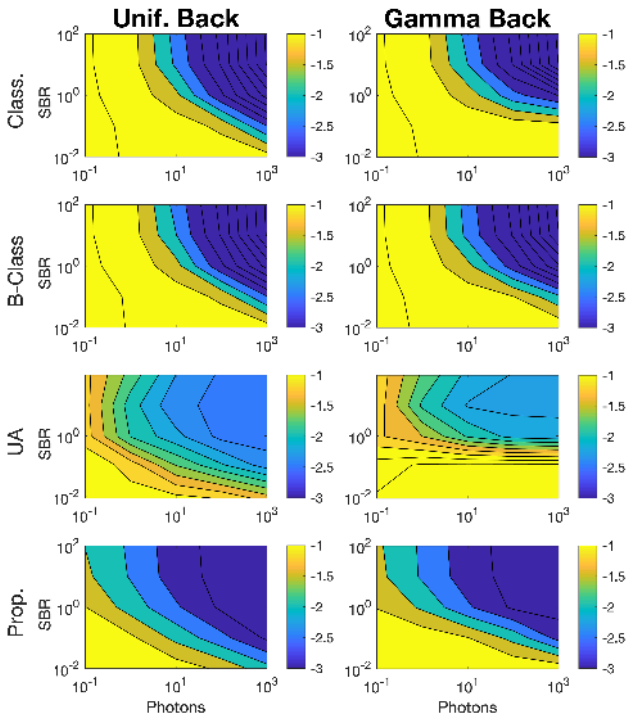

Fig. 3. Depth absolute errors (in log scale) obtained for the art scene with different algorithms w.r.t. SBR and PPP levels. (top-row) Class. algorithm, (second row) B-Class. algorithm, (third row) UA algorithm [19], (fourth row) proposed algorithm. (Left-column) data with uniform background, (right-column) data with gamma background. The lower DAE the better.

Fig. 5 shows the depth and reflectivity maps together with their uncertainty maps for SBR=1 and PPP=10 photons for uniform background (i.e., 5 signal photons on average). It is observed that the proposed algorithm provides sharp depth maps due to to the use of ℓ_1 based sparsity inducing prior. Note that higher uncertainty is observed on low reflectivity areas, near object edges and on corrupted regions due to high-background levels.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCI.2021.3111572, IEEE Transactions on Computational Imaging

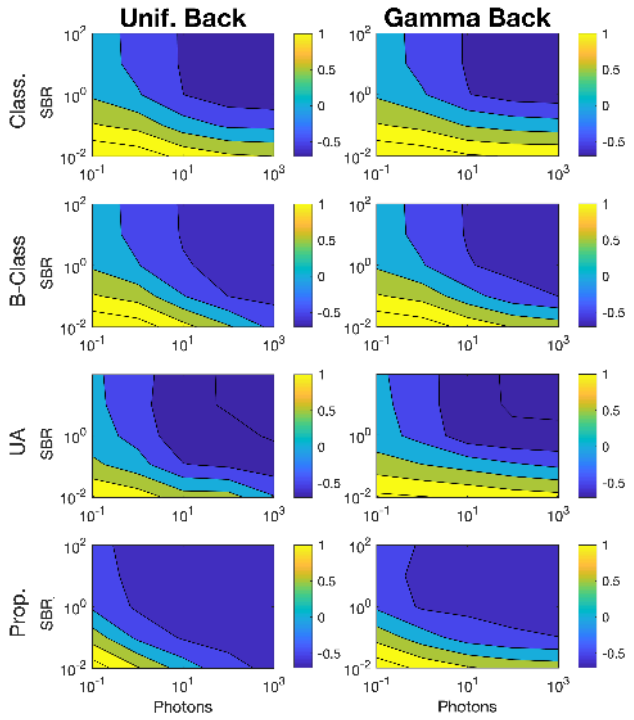

Fig. 4. Normalized intensity absolute errors (in log scale) obtained for the art scene with different algorithms w.r.t. SBR and PPP levels. (top-row) Class. algorithm, (second row) B-Class. algorithm, (third row) UA algorithm [19], (fourth row) proposed algorithm. (Left-column) data with uniform background, (right-column) data with gamma background. The lower IAE the better.

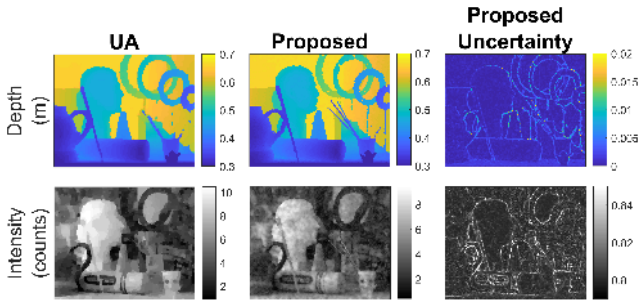

Fig. 5. Estimated depth and reflectivity maps with the UA [19] and proposed algorithms for SBR=1 and PPP=10 photons and uniform background. The proposed algorithm provides additional uncertainty maps.

An advantage of the proposed algorithm is that it can benefit from state-of-the-art algorithms and use their results as a guide. We investigate here the performance of the proposed algorithm when considering two depth guides (GD1 and GD2) and three intensity guides (GI1, GI2, GI3). We repeat the same experiment as above while fixing SBR=1 and varying PPP. Fig. 6 shows the DAE and IAE performance of the four variants, indicating an overall similar performance with a slight advantage for GD1 when compared to GD2. GI3 provides similar results as GI1, GI2 and is not represented for clarity. In what follows, we consider the GD1 and GI1 guides for all experiments.

Table I finally reports the computational time of the algorithms considered indicating fast performance when compared

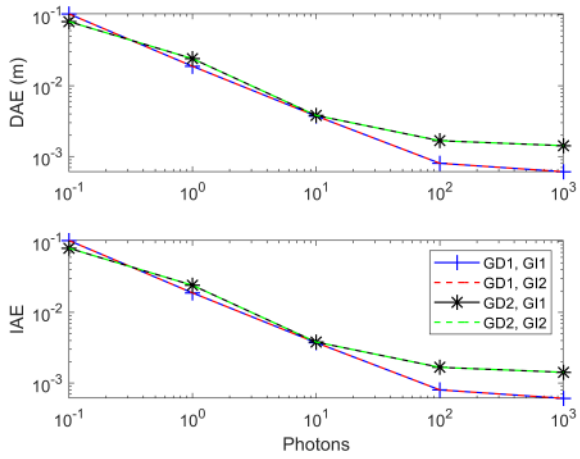

Fig. 6. Proposed algorithm using different depth and intensity guides on the art scene at SBR=1 and different ppp levels. (top) depth errors, (bottom) normalized intensity errors. GI3 provides similar results as GI1, GI2 and is not represented for clarity.

TABLE I
AVERAGE COMPUTATIONAL TIME (IN SECONDS) OF THE COMPARED ALGORITHMS ON $283 \times 183 \times 300 \times K$ DATA CUBE GENERATED WITH A UNIFORM AND GAMMA BACKGROUND, WHERE K REPRESENTS THE NUMBER OF WAVELENGTHS.

		Average photons per pixel (PPP)				
		0.1	1	10	100	1000
Art, K=1 wavelength	Class.	0.4	0.4	0.4	0.4	0.4
	B-Class.	2.3	2.4	2.4	2.4	2.4
	UA	37	42	65	62	136
	Prop.	4.2	4.2	3.9	3.8	3.8
Art, K=3 wavelengths	Class.	10	10	10	10	
	B-Class.	17	18	17	17	
	MNR3D	110	185	146	97	
	MUSAPOP	548	628	840	1169	
	Prop.	21	21	19	19	

to the UA algorithm. Note that the computational bottleneck of the proposed algorithm occurs when building downsampled cubes necessary for background estimation, whose time is indicated by B-Class. times (representing $> 60\%$ of proposed algorithm computational times). It should be also noted that most operations of the proposed algorithm are pixel-wise or bin-wise independent, and a much reduced computational time is expected by using parallel computing tools.

C. Evaluation on multispectral 3D Lidar data

This section analyses the performance of the proposed algorithm for multi-spectral Lidar imaging. We consider 3 wavelengths of the Art scene to generate three histograms of counts with the same realistic IRF as in Section VI-B while varying SBR and PPP levels. The proposed algorithm is considered with the same hyperparameters (as in Section VI-B), and is compared with the MUSAPOP algorithm (used with the authors' parameters) and the MNR3D algorithm. The latter algorithms are designed to process multi-spectral data and delivers point clouds hence the use of probability of detection, and number of false detections to evaluate performance. Note that MNR3D operates on the data cubes which requires large memory, thus it has been applied on

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCI.2021.3111572, IEEE Transactions on Computational Imaging

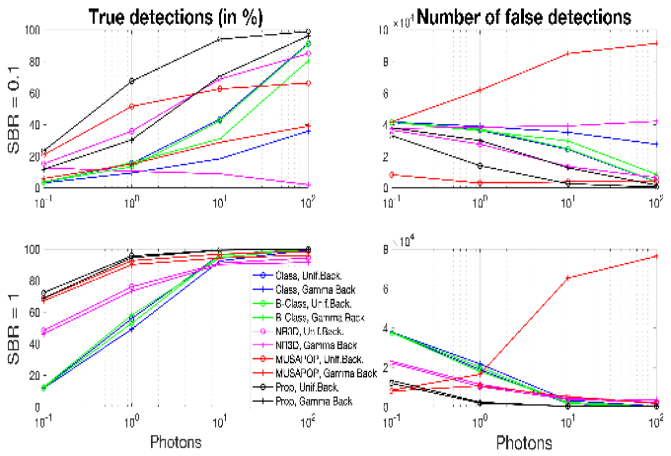

Fig. 7. PD and false detections of the Class, B-Class., MNR3D, MUSAPOP and proposed algorithms for different SBR, PPP levels and background shapes, with an error distance of $\tau = 10$bins.

spatially downsampled data of size 142×92 pixels and 300 time bins and final results have been upsampled using nearest neighbour interpolation. Fig. 7 represents these two criteria for five algorithms when considering a uniform (cross marker) and gamma shaped backgrounds (circle marker) for two SBR levels and several PPPs at $\tau = 10$ bins (where PPP is the average number of photons per pixel and per wavelength). The proposed algorithm presents best performance (highest true detections and lowest false ones) highlighting its robustness due to the efficient use of the multi-scale information. Fig. 2 shows an example of the obtained point clouds with different algorithms for uniform background. MUSAPOP is better than B-Class. and MNR3D, but fails to process the noisy case where only one signal photon is present against 9 background counts on average (SBR=0.1, PPP=10 photons). The proposed algorithm presents best performance and is robust to noise. In addition, it does not join disconnected surfaces (due to the use of sparsity inducing Laplace prior for the depth) and shows sharp intensity values (due to the weighted and depth guided reflectivity reconstruction). Table I finally highlights the fast computational time of the proposed algorithm when compared to the MUSAPOP and MNR3D algorithms.

VII. Results on real data

A. Results on real 3D underwater data

The proposed algorithm is validated on real underwater Lidar data of a moving target (painted metal flange 13 mm thick, diameter of 70 mm, with 7 mm diameter holes) put at a stand-off distance of 1.7m from the end of the water tank nearest the sensor (see the target in Fig. 8 (a)). The data were acquired in lab settings using a CMOS Si-SPAD detector array based system acquiring binary frames at a rate of 500fps, with 1ms acquisition time per frame, 700 time bins and 34ps per bin [11]. The 128×192 pixels binary frames were pre-processed by building histograms of counts every 10ms (max of 10 counts per histogram). Different concentrations of a commercially available antacid medicine, called Maalox, were mixed with water to obtain varying scattering levels of the imaging environment. With high Maalox concentrations, the turbid water is highly scattering leading to a non-uniform background as shown in Fig. 9.

The proposed algorithm is considered using the hyper-parameters of Section VI-B and the guides GD1 and GI1. Results are compared with the B-Class. algorithm and the RT3D algorithm. The UA algorithm is not considered as it assumes the presence of a target in all pixels which is not satisfied in this case. Fig. 8 shows the 3D point clouds obtained with the different algorithms for clear water (b-c-d) and turbid water (e-f-g)[2]. All algorithms performed well in clear water. However, both RT3D and B-Class. performed poorly in turbid water due to non-uniform background affecting the data, and leading to the detection of a fake object in front of the true target. The proposed algorithm successfully eliminates the background counts and retrieve a good reconstruction of the target even under these extreme imaging conditions. In addition, the proposed algorithm also provides uncertainty maps for the estimated parameters, as indicated in Fig. 10. These maps show higher uncertainties when imaging through scattering water, and near object edges. More results when considering other frames at AL=1.2 are provided in video 1, video 2, and at AL=4.8 in video 3, video 4.

B. Results on real photon starved multispectral data

The proposed algorithm is validated on real multispectral Lidar data of a static lego target (see Fig. 11 (left) for a reference acquired at 40 ms acquisition time per pixel) [23]. This data has 200×200 pixels, $T = 1500$ bins (a bin represents 2ps) and $L = 4$ wavelengths acquired at $473, 532, 589$ and 640 nm. We are interested in the sparse photon regime and analyse performance with six acquisition times per pixel as follows $0.05, 0.1, 0.5, 1, 10, 40$ ms, leading to average photon counts of $1.4, 2.9, 14.5, 29, 289, 1159$ photons per pixel and per wavelength with SBR $\in [50, 60]$. Fig. 11 shows the reference point cloud (obtained with the B-Class. algorithm on 40ms data after correcting outliers) and the reconstruction results with the B-Class., MUSAPOP, MNR3D and proposed algorithms at $500\mu s$ acquisition time per pixel. The B-Class. provides multiple false detections, and a noisy reflectivity image. MNR3D operates on the data cubes which requires large memory, thus it has been applied on spatially downsampled cubes of size 100×100 pixels and 300 time bins, and the final results have been upsampled using nearest neighbour interpolation. Both MNR3D and MUSAPOP improve results compared to B-Class. but show blocky-points, and blurred reflectivity. In contrast, the proposed algorithm provided the best performance with distinct surfaces and sharp reflectivity image even at this low level of photons. Fig. 12 indicates the true detection probability and number of false detection for the studied algorithms, indicating the superiority of the proposed algorithm.

[2] Attenuation length (AL) is an indication of the effect of optical attenuation, and is the distance over which the light intensity is reduced to 1/exp(1) of its original value. AL = αd, when the light propagates a distance d in water with attenuation coefficient α [11].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCI.2021.3111572, IEEE Transactions on Computational Imaging

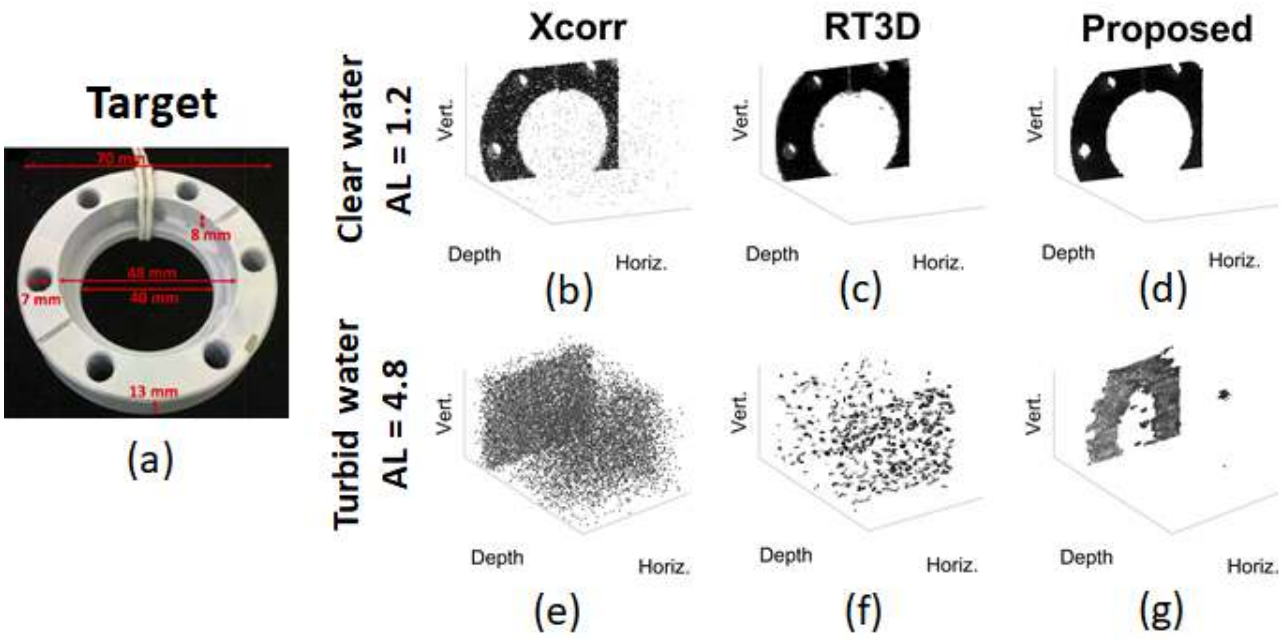

Fig. 8. (a) Target used for underwater imaging experiments. 3D representations of underwater scenes with (b-e) B-Class., (c-f) RT3D [34] and (d-g) the proposed algorithm. (b-c-d-e) clear water with AL=1.2, (f-g-h-i) turbid water with AL=4.8 (these AL values are for transceiver to target, and not round-trip).

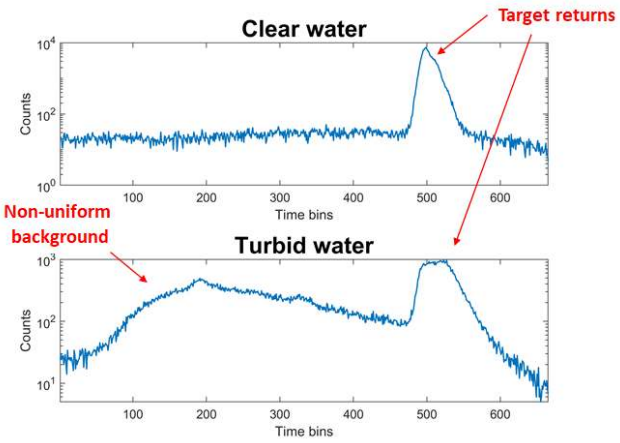

Fig. 9. Examples of the obtained histogram summed over all pixels for (top) clear water with laser power 1.2 mW, (bottom) turbid water with laser power 8 mW, when imaging the flange target located around the 500 time bin. The bottom curve highlights a non-uniform background (see the region 100 - 400 bins). The shape of the peak is different in the two histograms because the target return is visible in different pixels during the two measurements, as the target moves across the field of view of the camera.

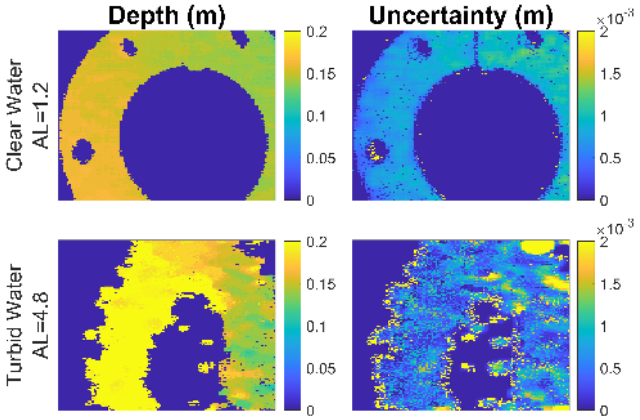

Fig. 10. Estimated Depth values and uncertainty maps using the proposed method for (top) clear and (bottom) turbid water.

VIII. Conclusions

This paper addressed the combination of several challenging problems using a new robust Bayesian algorithm for the reconstruction of multispectral single-photon Lidar data. The algorithm exploited multi-scale information to improve depth and reflectivity estimates under extreme conditions due to low light level illumination or imaging through turbid media. The framework has the ability to incorporate other state-of-the-art denoisers/estimators or fuse information from additional sensing sources, in addition to providing uncertainty measures regarding the estimates which is crucial for decision making. The algorithm has been validated on different scenarios using both simulated and real data, leading to best results in presence of a high and possibly non-uniform background noise, and a significant speed improvements over other state-of-the-art algorithms. Future work will generalize the proposed strategy to process multiple detections per-pixel as observed in object's edges or when imaging through semi-transparent surfaces. Current implementation was done in Matlab, and a computational improvement is expected by using parallel computing tools (such as GPUs) which is being investigated.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCI.2021.3111572, IEEE Transactions on Computational Imaging

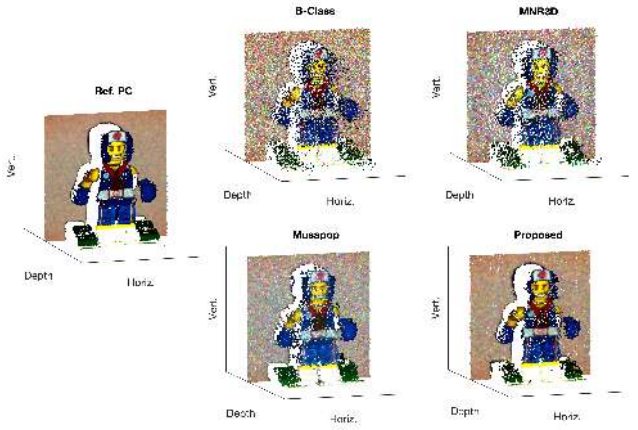

Fig. 11. 3D representations of Lego scene with different algorithms obtained at $500\mu s$ acquisition time per pixel and 14 average counts per pixel and wavelength.

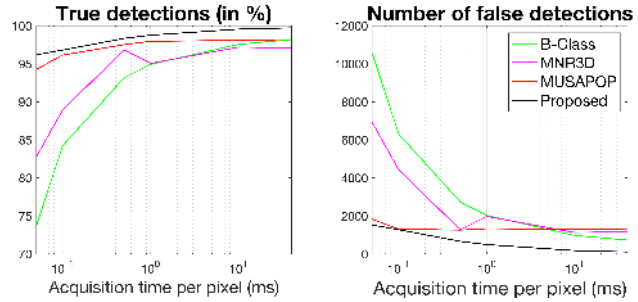

Fig. 12. PD and false detections of the B-Class., MNR3D [25], MUSAPOP [27], and proposed algorithms for the Lego data at different acquisition times per pixel with an error distance of $\tau = 30$bins.

APPENDIX
APPROXIMATE LIKELIHOOD

In absence of background counts, (3) can be written for the nth pixel as follows

$$P({}^{s}\boldsymbol{y}_n|\boldsymbol{r}_n, d_n) \propto \prod_{t,k} \frac{\left[r_{n,k} f_k\left(t-d_n\right)\right]^{{}^{s}y_{n,t,k}} \exp\left[-r_{n,k} f_k\left(t-d_n\right)\right]}{{}^{s}y_{n,t,k}!} \tag{30}$$

or equivalently

$$\begin{aligned} P({}^{s}\boldsymbol{y}_n|\boldsymbol{r}_n, d_n) \quad \propto \quad & \prod_{k=1}^{K} \left[r_{n,k}\right]^{\sum_{t=1}^{T} {}^{s}y_{n,t,k}} \\ \times \quad & \prod_{t,k} \left[f_k\left(t-d_n\right)\right]^{{}^{s}y_{n,t,k}} \frac{1}{{}^{s}y_{n,t,k}!} \\ \times \quad & \prod_{k=1}^{K} \exp\left[-\sum_{t=1}^{T} r_{n,k} f_k\left(t-d_n\right)\right] \end{aligned} \tag{31}$$

Assuming $\sum_{t=1}^{T} f_k\left(t-d_n\right) = 1, \forall k$ leads to $\sum_{t=1}^{T} r_{n,k} f_k\left(t-d_n\right) = r_{n,k} \sum_{t=1}^{T} f_k\left(t-d_n\right) = r_{n,k}$. In addition, denoting $\bar{s}_{n,k} = \sum_{t=1}^{T} {}^{s}y_{n,t,k}$ gives

$$\begin{aligned} P({}^{s}\boldsymbol{y}_n|\boldsymbol{r}_n, d_n) \quad \propto \quad & \prod_{k=1}^{K} r_{n,k}^{\bar{s}_{n,k}} \exp\left(-r_{n,k}\right) \\ \times \quad & \prod_{t,k} \left[f_k\left(t-d_n\right)\right]^{{}^{s}y_{n,t,k}} \frac{1}{{}^{s}y_{n,t,k}!}. \end{aligned} \tag{32}$$

Adopting a Bayesian strategy and assuming that the parameters $\boldsymbol{r}_n, d_n$ are random, (32) shows that $r_{n,k}$ follows a gamma distribution as indicated in (4). This was expected as gamma is a conjugate distribution of the Poisson distribution.

REFERENCES

[1] A. M. Wallace, A. Halimi, and G. S. Buller, "Full waveform lidar for adverse weather conditions," *IEEE Trans. Vehicular Tech.*, vol. 69, no. 7, pp. 7064–7077, 2020.

[2] J. Rapp, J. Tachella, Y. Altmann, S. McLaughlin, and V. K. Goyal, "Advances in single-photon lidar for autonomous vehicles: Working principles, challenges, and recent advances," *IEEE Signal Processing Magazine*, vol. 37, no. 4, pp. 62–71, 2020.

[3] G. S. Buller and A. Wallace, "Ranging and three-dimensional imaging using time-correlated single-photon counting and point-by-point acquisition," vol. 13, no. 4, pp. 1006–1015, July 2007.

[4] A. Kirmani, D. Venkatraman, D. Shin, A. Colaço, F. N. C. Wong, J. H. Shapiro, and V. K. Goyal, "First-photon imaging," *Science*, vol. 343, no. 6166, pp. 58–61, 2014.

[5] D. Shin, A. Kirmani, V. K. Goyal, and J. H. Shapiro, "Photon-efficient computational 3-d and reflectivity imaging with single-photon detectors," *IEEE Transactions on Computational Imaging*, vol. 1, no. 2, pp. 112–125, 2015.

[6] 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 Process.*, vol. 25, no. 5, pp. 1935–1946, Mar. 2015.

[7] A. M. Pawlikowska, A. Halimi, R. A. Lamb, and G. S. Buller, "Single-photon three-dimensional imaging at up to 10 kilometers range," *Opt. Express*, vol. 25, no. 10, pp. 11 919–11 931, May 2017.

[8] 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," *Photon. Res.*, vol. 8, no. 9, pp. 1532–1540, Sep 2020.

[9] Z.-P. Li, J.-T. Ye, X. Huang, P.-Y. Jiang, Y. Cao, Y. Hong, C. Yu, J. Zhang, Q. Zhang, C.-Z. Peng, F. Xu, and J.-W. Pan, "Single-photon imaging over 200km," *Optica*, vol. 8, no. 3, pp. 344–349, Mar 2021.

[10] D. B. Lindell, M. O'Toole, and G. Wetzstein, "Single-photon 3d imaging with deep sensor fusion," *ACM Trans. Graph.*, vol. 37, no. 4, pp. 113:1–113:12, July 2018.

[11] A. Maccarone, F. M. D. 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. Express*, vol. 27, no. 20, pp. 28 437–28 456, Sep 2019.

[12] S. C. Medin, J. Murray-Bruce, D. Castañón, and V. K. Goyal, "Beyond binomial and negative binomial: Adaptation in bernoulli parameter estimation," *IEEE Trans. Comput. Imaging*, vol. 5, no. 4, pp. 570–584, 2019.

[13] A. Halimi, P. Ciuciu, A. McCarthy, S. McLaughlin, and G. S. Buller, "Fast adaptive scene sampling for single-photon 3d lidar images," in *IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Guadeloupe, West Indies, Dec. 2019.

[14] Q. Legros, J. Tachella, R. Tobin, A. Mccarthy, S. Meignen, G. S. Buller, Y. Altmann, S. Mclaughlin, and M. E. Davies, "Robust 3d reconstruction of dynamic scenes from single-photon lidar using beta-divergences," *IEEE Trans. Image Process.*, vol. 30, pp. 1716–1727, 2021.

[15] 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.

[16] R. Tobin, A. Halimi, A. McCarthy, P. Soan, and G. S. Buller, "Robust real-time 3D imaging of moving scenes through atmospheric obscurant using single-photon LiDAR," *Scientific Reports*, vol. 11, p. 11236, 2021.

[17] G. Satat, M. Tancik, and R. Raskar, "Towards photography through realistic fog," in *Computational Photography (ICCP), 2018 IEEE International Conference on*. IEEE, 2018, pp. 1–10.

[18] R. Tobin, A. Halimi, A. McCarthy, M. Laurenzis, F. Christnacher, and G. S. Buller, "Three-dimensional single-photon imaging through obscurants," *Opt. Express*, vol. 27, no. 4, pp. 4590–4611, Feb 2019.

[19] 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, no. 3, pp. 445–459, Sept. 2017.

[20] A. Halimi, A. Wallace, G. Buller, and S. McLaughlin, "Fast surface detection using single-photon detection events," in *IEEE SSPD*, Sept. 2020.

[21] J. Tachella, Y. Altmann, S. McLaughlin, and J. . Tourneret, "Fast surface detection in single-photon lidar waveforms," in *EUSIPCO-19*, A Coruna, Spain, 2019, pp. 1–5.

[22] 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, no. 4, pp. 658–670, Dec 2017.

[23] Q. Legros, S. Meignen, S. McLaughlin, and Y. Altmann, "Expectation-maximization based approach to 3d reconstruction from single-waveform multispectral lidar data," *IEEE Trans. Comput. Imaging*, vol. 6, pp. 1033–1043, 2020.

[24] Y. Altmann, A. Maccarone, A. Halimi, A. McCarthy, G. Buller, and S. McLaughlin, "Efficient range estimation and material quantification from multispectral lidar waveforms," in *2016 Sensor Signal Processing for Defence (SSPD)*, Sept 2016, pp. 1–5.

[25] A. Halimi, R. Tobin, A. McCarthy, J. Bioucas-Dias, S. McLaughlin, and G. S. Buller, "Robust restoration of sparse multidimensional single-photon lidar images," *IEEE Trans. Comput. Imaging*, vol. 6, pp. 138–152, 2020.

[26] S. Hernandez-Marin, A. M. Wallace, and G. J. Gibson, "Multilayered 3D lidar image construction using spatial models in a bayesian framework," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 6, pp. 1028–1040, June 2008.

[27] 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*, vol. 6, pp. 208–220, 2020.

[28] A. Halimi, Y. Altmann, A. McCarthy, X. Ren, R. Tobin, G. S. Buller, and S. McLaughlin, "Restoration of intensity and depth images constructed using sparse single-photon data," in *Proc. EUSIPCO*, 2016, pp. 86–90.

[29] Z. Sun, D. B. Lindell, O. Solgaard, and G. Wetzstein, "Spadnet: deep rgb-spad sensor fusion assisted by monocular depth estimation," *Opt. Express*, vol. 28, no. 10, pp. 14 948–14 962, May 2020.

[30] J. Peng, Z. Xiong, X. Huang, Z.-P. Li, D. Liu, and F. Xu, "Photon-efficient 3d imaging with a non-local neural network," in *Computer Vision – ECCV 2020*. Cham: Springer International Publishing, 2020, pp. 225–241.

[31] A. Ruget, S. McLaughlin, R. K. Henderson, I. Gyongy, A. Halimi, and J. Leach, "Robust super-resolution depth imaging via a multi-feature fusion deep network," *Opt. Express*, in press.

[32] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *2013 IEEE Global Conference on Signal and Information Processing*, 2013, pp. 945–948.

[33] W. Marais and R. Willett, "Proximal-gradient methods for poisson image reconstruction with bm3d-based regularization," in *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2017, pp. 183–187.

[34] J. Tachella, Y. Altmann, N. Mellado, A. McCarthy, R. Tobin, G. S. Buller, J.-Y. Tourneret, and S. J. McLaughlin, "Real-time 3d reconstruction from single-photon lidar data using plug-and-play point cloud denoisers," *Nature Communications*, vol. 10, no. 1, p. 4984, 2019.

[35] Y. Romano, M. Elad, and P. Milanfar, "The little engine that could: Regularization by denoising (red)," *SIAM Journal on Imaging Sciences*, vol. 10, no. 4, pp. 1804–1844, 2017.

[36] Y. L. V. Monga and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *IEEE Signal Processing Magazine.*, 2020, to appear.

[37] J. Tachella, Y. Altmann, X. Ren, A. McCarthy, G. S. Buller, S. McLaughlin, and J.-Y. Tourneret, "Bayesian 3d reconstruction of complex scenes from single-photon lidar data," *SIAM Journal on Imaging Sciences*, vol. 12, no. 1, pp. 521–550, 2019.

[38] D. P. Bertsekas, *Nonlinear programming*. Belmont, Massachusetts: Athena Scientific, 1995.

[39] J. Sigurdsson, M. Ulfarsson, and J. Sveinsson, "Hyperspectral unmixing with l_q regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 6793–6806, Nov. 2014.

[40] A. Halimi, C. Mailhes, J.-Y. Tourneret, and H. Snoussi, "Bayesian estimation of smooth altimetric parameters: Application to conventional and delay/Doppler altimetry," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 2207–2219, Mar. 2016.

[41] Q. Zhang, L. Xu, and J. Jia, "100+ times faster weighted median filter (wmf)," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, June 2014, pp. 2830–2837.

[42] G. Arce, *Weighted Median Filters*. John Wiley & Sons, Ltd, 2004, pp. 139–250.

[43] M. Pfennigbauer, C. Wolf, J. Weinkopf, and A. Ullrich, "Online waveform processing for demanding target situations," in *Laser Radar Technology and Applications XIX; and Atmospheric Propagation XI*, M. D. Turner, G. W. Kamerman, L. M. W. Thomas, and E. J. Spillar, Eds., vol. 9080, International Society for Optics and Photonics. SPIE, 2014, pp. 142 – 151.

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

[45] T. Park and G. Casella, "The bayesian lasso," *Journal of the American Statistical Association*, vol. 103, no. 482, pp. 681–686, 2008.

[46] M. A. T. Figueiredo, "Adaptive sparseness for supervised learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1150–1159, 2003.

[47] O. Dikmen and A. Cemgil, "Gamma markov random fields for audio source modeling," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 3, pp. 589–601, March 2010.

[48] W. Liu, X. Chen, J. Yang, and Q. Wu, "Robust color guided depth map restoration," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 315–327, 2017.

[49] S. Chan, A. Halimi, F. Zhu, I. Gyongy, R. K. Henderson, R. Bowman, S. McLaughlin, G. S. Buller, and J. Leach, "Long-range depth imaging using a single-photon detector array and non-local data fusion," *Scientific Reports*, vol. 9, no. 1, p. 8075, 2019.

[50] X. Liu, D. Zhai, R. Chen, X. Ji, D. Zhao, and W. Gao, "Depth super-resolution via joint color-guided internal and external regularizations," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 1636–1645, April 2019.

[51] R. B. Rusu, Z. C. Marton, N. Blodow, M. Dolha, and M. Beetz, "Towards 3d point cloud based object maps for household environments," *Robotics and Autonomous Systems*, vol. 56, no. 11, pp. 927–941, 2008.

[52] P. Milanfar, "A tour of modern image filtering: New insights and methods, both practical and theoretical," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 106–128, 2013.

[53] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, 1998, pp. 839–846.

[54] 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.

[55] J. Salmon, Z. Harmany, C.-A. Deledalle, and R. Willett, "Poisson noise reduction with non-local PCA," *Journal of Mathematical Imaging and Vision*, vol. 48, no. 2, pp. 279–294, 2014.

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

[57] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[58] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan 2011.

[59] R. Nassif, C. Richard, A. Ferrari, and A. H. Sayed, "Proximal multitask learning over networks with sparsity-inducing coregularization," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6329–6344, 2016.

[60] M. Figueiredo and J. Bioucas-Dias, "Restoration of poissonian images using alternating direction optimization," *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3133–3145, Dec 2010.

[61] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, June 2011.

[62] I. Gyongy, S. W. Hutchings, A. Halimi, M. Tyler, S. Chan, F. Zhu, S. McLaughlin, R. K. Henderson, and J. Leach, "High-speed 3d sensing via hybrid-mode imaging and guided upsampling," *Optica*, vol. 7, no. 10, pp. 1253–1260, Oct 2020.

[63] K. Madsen, H. B. Nielsen, and O. Tingleff, "Methods for non-linear least squares problems (2nd ed.)," Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, p. 60, 2004.

[64] D. Scharstein and C. Pal, "Learning conditional random fields for stereo," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, June 2007, pp. 1–8.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCI.2021.3111572, IEEE Transactions on Computational Imaging

Abderrahim Halimi (S'11-M'14-SM'20) received the Eng. degree in electronics from the National Polytechnic School of Algiers, Algeria, in 2009, and both the M.Sc. and Ph.D. degrees in signal processing from the Institut National Polytechnique de Toulouse, Toulouse, France, in 2010 and 2013, respectively. He was a Post-Doctoral Research Associate from 2013 to 2018, and is currently an Assistant Professor and Royal Academy of Engineering (RAEng) Research Fellow within the School of Engineering and Physical Sciences in Heriot-Watt University. His research activities focus on statistical signal and image processing, with a particular interest in Bayesian inverse problems with applications to remote sensing (hyperspectral imaging, satellite altimetry), single-photon depth imaging and medical imaging.

Aurora Maccarone is from Palermo, Italy. She graduated in Physics from Universitá di Palermo (Palermo, Italy), and she received a PhD in Physics and Applied Photonics from Heriot-Watt University (Edinburgh, UK) in 2016. Following her doctoral studies, Aurora worked at Heriot-Watt University where she took her first position as a PDRA, and since 2020 she is a Research Fellow funded by the Royal Academy of Engineering. Her research interests include single-photon technologies, imaging in challenging conditions, and quantum communications.

PLACE PHOTO HERE

Robert Lamb Robert Lamb obtained his BSc in physics at the University of York in 1983, an MSc in Applied and Modern Optics from the University of Reading in 1989 and his PhD in phase conjugation in high power lasers from Imperial College, London in 1994. He began his career with the UK MoD (later QinetiQ) investigating phase conjugation with stimulated Brillouin scattering and multi-wavelength lasers with stimulated Raman scattering in high power lasers. He also developed low-light level remote sensing using time correlated single photon counting for target identification and ultra-low light level sensing. In 2006 he was appointed Head of Future Systems Group at SELEX Galileo (now Leonardo) and is now the Chief Technologist for Electro-optics, with responsibility for future technology investment and innovation in lasers and photonics for airborne targeting systems. He is an Honorary Professor at Heriot-Watt University, a QinetiQ Fellow and a Fellow of the Institute of Physics.

Gerald S. Buller received the B.Sc. (Hons.) degree in natural philosophy from the University of Glasgow, Glasgow, U.K., in 1986 and the Ph.D. degree in physics from Heriot-Watt University, Edinburgh, U.K., in 1989. He is currently a Professor of physics at Heriot-Watt University. In 2002, he co-founded Helia Photonics Ltd., based in Livingston, U.K. In 2015, he received an EPSRC Established Career Fellowship in Quantum Technology to research sparse photon imaging. His current research interests include single-photon detection methods, particularly at infrared wavelengths, and applications involving the use of single photons. These applications include single-photon depth imaging and novel quantum communications protocols. He is a Fellow of the Royal Society of Edinburgh, the UK Institute of Physics, and the Optical Society of America.

Stephen McLaughlin (F'11) was born in Clydebank, Scotland in 1960. He received the B.Sc. degree in Electronics and Electrical Engineering from the University of Glasgow in 1981 and the Ph.D. degree from the University of Edinburgh in 1990. From 1981 to 1984 he was a Development Engineer in industry. In 1986 he joined the University of Edinburgh where he studied the performance of linear adaptive algorithms in high noise and nonstationary environments. In 1988 he joined the academic staff at Edinburgh, and from 1991 until 2001 he held a Royal Society University Research Fellowship to study nonlinear signal processing techniques. In 2002 he was awarded a personal Chair in Electronic Communication Systems at the University of Edinburgh. In October 2011 he joined Heriot-Watt University as a Professor of Signal Processing and Head of the School of Engineering and Physical Sciences. His research interests lie in the fields of adaptive signal processing and nonlinear dynamical systems theory and their applications to biomedical, energy and communication systems. Prof McLaughlin is a Fellow of the Royal Academy of Engineering, of the Royal Society of Edinburgh, of the Institute of Engineering and Technology and of the IEEE.