Fast online 3D reconstruction of dynamic scenes from individual single-photon detection events

Fast online 3D reconstruction of dynamic scenes from individual single-photon detection events

Yoann Altmann, *Member, IEEE,* Stephen McLaughlin, *Fellow, IEEE* and Michael E. Davies, *Fellow, IEEE*

***Abstract*—In this paper, we present an algorithm for online 3D reconstruction of dynamic scenes using individual times of arrival (ToA) of photons recorded by single-photon detector arrays. One of the main challenges in 3D imaging using single-photon Lidar is the integration time required to build ToA histograms and reconstruct reliably 3D profiles in the presence of non-negligible ambient illumination. This long integration time also prevents the analysis of rapid dynamic scenes using existing techniques. We propose a new method which does not rely on the construction of ToA histograms but allows, for the first time, individual detection events to be processed online, in a parallel manner in different pixels, while accounting for the intrinsic spatiotemporal structure of dynamic scenes. Adopting a Bayesian approach, a Bayesian model is constructed to capture the dynamics of the 3D profile and an approximate inference scheme based on assumed density filtering is proposed, yielding a fast and robust reconstruction algorithm able to process efficiently thousands to millions of frames, as usually recorded using single-photon detectors. The performance of the proposed method, able to process hundreds of frames per second, is assessed using a series of experiments conducted with static and dynamic 3D scenes and the results obtained pave the way to a new family of real-time 3D reconstruction solutions.**

***Index Terms*—3D reconstruction, Single-photon Lidar, Bayesian filtering, online estimation**

I. Introduction

Fast reconstruction of 3D scenes using single-photon light detection and ranging (Lidar) technology is an important challenge which is important in applications such as autonomous driving [1], environmental monitoring [2]–[4] and defence [5]. A growing number of 3D imaging modalities is becoming increasingly popular [6], and single-photon Lidar offers appealing advantages, including low-power, a capability for long-range imaging [7], [8] or imaging in complex media such as fog/smoke [9] and underwater [10], [11] with excellent range resolution (of the order of millimetres [12]). Recently, several algorithms have also been proposed to analyse distributed objects [13]–[18], i.e., when multiple surfaces are visible within each pixel.

Despite pushing the boundaries of 3D reconstruction in extreme environments, single-photon Lidar still suffers from 1) relatively long integration times required to obtain sufficiently reliable data and 2) significant computational requirements to process the resulting large volume of data recorded by single-photon imaging systems. Recent advances in single-photon avalanche diode (SPAD) detector arrays [19], [20] have allowed significant reductions in acquisition times over raster scanning systems [12], [21]–[23], enabling acquisitions with video frame rates. Yet, robust, automated and scalable methods allowing for fast analysis of single-photon data are still required. One of the main bottlenecks of most state-of-the-art 3D reconstruction methods [16], [24]–[29] is that they rely on the construction of histograms of photon times of arrival (ToA) (or batches of detection events), which, when synchronised with a pulsed laser (time correlated single-photon counting, TCSPC) correspond to photon times of flight (ToF), used to infer object ranges. One important exception is the so-called "first-photon" imaging approach [30] whereby the reflectivity and 3D profiles of the scene can be recovered using a single photon per pixel. However, the approach in [30] targets primarily raster scanning Lidar systems, allowing the variable per-pixel acquisition times, i.e., until the first photon is detected.

In this work, we consider Lidar data acquired using SPAD arrays and investigate a new 3D reconstruction algorithm that does not rely on ToF histograms, but on individual photon detection events. More precisely, we address the problem of 3D reconstruction after each time period (defined in Section II) during which each SPAD detector can record at most one detection event. This approach is particularly relevant for applications where the objects in the scene can move significantly faster than the integration period or the number of laser repetitions required to build sufficiently populated ToF histograms. In such cases, the relative movement of the scene with respect to the sensor can produce a 3D blur that produces broader peaks or even multiple returns in some Lidar waveforms, which can jeopardise the 3D reconstruction task. By reconstructing a 3D profile after each (short) time period, our approach, which processes sequentially individual detection events rather than ToF histograms, is significantly less prone to 3D blur. It is however important to note that our method assumes a single visible surface per pixel at each time instant. Generalisation of this work to multiple surfaces, which is a significantly more complex problem using individual detections, is out of scope of this work.

Adopting a Bayesian approach, we consider a likelihood model based on the standard single-photon Lidar observation model in the low-flux regime. We then introduce a dynamic model for the spatiotemporal (ST) evolution of the 3D profile. Due to the complex nature of the likelihood (mixture of two distributions) and the structure of the prior model, the standard online estimation methods based on (extended) Kalman filtering [31] cannot be used directly. As the complexity of the resulting model grows prohibitively with the number of detection events, i.e., over time, we adopt an approximate estimation strategy based on assumed density filtering (ADF)

[32]–[34], whereby the posterior distribution of the 3D profile estimated for a given frame is projected onto a family of more tractable distributions (Gaussian distributions here), which reduces significantly the complexity of the sequential estimation procedure. Although particle filters [35] could also be considered for approximate inference, this would lead to an increased computational cost induced by the approximation of densities using a large number of particles. In Sections II and III, we discuss how our ADF-based approach benefits from update rules which can be computed analytically and thus do not require more computationally intensive iterative optimization procedures as in [11], [15], [18] nor Monte Carlo sampling steps as in [16], [21]. It is important to note that the resulting method, which uses each frame (during which at most one photon can be detected per pixel) only once, enables online 3D reconstruction of dynamic scenes with limited memory requirements. Indeed, the individual frames are processed sequentially, resulting in a fixed computational cost per frame which is important for any real time implementation. Moreover, thanks to its intrinsically parallel algorithmic architecture, the proposed method is extremely scalable to large arrays and long sequences of frames. Another important advantage is that it does not require the knowledge of the (potentially time varying) ambient illumination level.

To summarise, the main contributions of this work are:

- A new Bayesian model for sequential 3D reconstruction using individual photon-detection events
- An online estimation strategy, proposed to the best of our knowledge for the first time, for reconstruction of dynamic 3D scenes from individual photon detection events. This method based on assumed density filtering is highly scalable and computationally attractive.

The remainder of the paper is organised as follows. Section II first recalls the classical observation model for 3D reconstruction using single-photon measurements in the photon-starved regime and describes the Bayesian model and inference strategy proposed for 3D reconstruction using a single frame. The generalisation of this method to online 3D reconstruction of dynamic scenes is detailed in Section III. Results of simulations conducted with simulated single-pixel data and sequences of frames are presented and discussed in Section IV and conclusions are finally reported in Section V.

II. Single frame analysis

A. Observation model

In this work, we consider a sequence of N frames, where each frame of duration T consists of P pixels. This paper addresses the reconstruction of dynamic 3D scenes where each single-photon detector, associated with one pixel, is able to record at most one detection event per pixel and per frame.

Let's first consider an active illumination scenario where the laser emits pulses of light with a repetition/illumination period $T_r = T$. As described in [28], assuming that a single surface is visible in each pixel, within each frame n, the average photon flux at the detector/pixel p can be modelled as

$$\lambda_{p,n}(t) = r_{p,n} s(t - 2d_{p,n}/c) + b_{p,n}, \forall t \in [0; T_r)\,, \tag{1}$$

where $d_{p,n}$ is the instantaneous distance of the object, c is the speed of light in the homogeneous medium between the imaging system and the detector and $r_{p,n}$ is an amplitude parameter related to the reflectivity of the object. Moreover, $b_{p,n}$ represents the instantaneous ambient illumination and dark count level in the pth pixel, which can potentially vary among pixels. Note that $r_{p,n}$ and $b_{p,n}$ also account for the quantum efficiency of the detectors that is not further detailed here for brevity (see [28] for details). Moreover, $s(\cdot)$ is the overall impulse response of the imaging system, which includes the shape of the pulse emitted by the laser and the temporal response of the single-photon detector. As in [28], we assume that $s(\cdot)$ is known as it can be measured during the calibration of the Lidar system, and that it can be well approximated by a Gaussian profile with variance s^2. As will be discussed in Section II, the proposed method can also be applied when the shape of this impulse response is not Gaussian and changes from one pixel to another due, for instance, to the inhomogeneity of the P detectors.

Over the nth illumination period, the detection rate is thus given by

$$\Lambda_{p,n} = \int_0^{T_r} \lambda_{p,n}(t)\mathrm{d}t = r_{p,n} S + B_{p,n} \tag{2}$$

where $B_{p,n} = T_r b_{p,n}$ and where we assume that the object distance is not too close from the minimum (0) and maximum $(T_r c/2)$ admissible ranges such that the integral $S = \int_0^{T_r} s(t - 2d_{p,n}/c)dt$ remains constant, whatever the value of $d_{p,n}$. In the low-flux regime, we have $r_{p,n}(t)S + B_{p,n} \ll 1$, such that the probability of two photons reaching the same detector in a given interval T_r is small and such that the dead-time of the detector can be neglected. In that case, the probability of detection is given by $\pi_{p,n} = 1 - \exp\left[-\Lambda_{p,n}\right] \approx \Lambda_{p,n}$ and the probability of a detected photon being associated with the original emitted pulse, denoted by $w_{p,n}$ is given by $w_{p,n} = (r_{p,n}S)/\Lambda_{p,n}$. Let $z_{p,n} \in (0; 1)$ be a binary label indicating a detection event (i.e., when $z_{p,n} = 1$) in pixel p for the nth frame, such that

$$f(z_{p,n} = 1|\pi_{p,n}) = \pi_{p,n}. \tag{3}$$

When $z_{p,n} = 1$, the observation model for the measured time of arrival $y_{p,n} \in [0; T_r)$ in pixel p and frame n can be expressed as

$$f(y_{p,n}|z_{p,n} = 1, w_{p,n}, d_{p,n}) = w_{p,n} f_s\left(y_{p,n} - \frac{2d_{p,n}}{c}\right) + (1 - w_{p,n})\mathcal{U}_{[0;T_r)}(y_{p,n}), \tag{4}$$

where $\mathcal{U}_{[0;T_r)}(\cdot)$ is the uniform distribution defined on $[0; T_r)$, and with $f_s(t - 2d_{p,n}/c) = S^{-1}s(t - 2d_{p,n}/c), \forall (p, n)$. Moreover, we use the notation $y_{p,n} = \emptyset$ when no detections were recorded in the pth pixel within the nth frame.

Assume now that a frame lasts N_r laser repetition periods, i.e., $T = N_r T_r$ and that the detector is only able to record at most one detection event during that frame. If the observation conditions have not changed during the N_r repetitions, the probability of detection is given by $\tilde{\pi}_{p,n} = 1 - \exp\left[-N_r \Lambda_{p,n}\right]$. However, in the low-flux regime, Eq. (4) still applies. Consequently, although each frame can result from more than

one illumination period, the observation models (3) and (4) is still valid by replacing $\pi_{p,n}$ by $\tilde{\pi}_{p,n}$ in (3), provided that the observation conditions have not changed over the period T. This observation can be useful for practical applications since the in low-flux regime, imposing $r_{p,n}S + B_{p,n} \ll 1$ (for each T_r interval) leads to extremely sparse detection events and large volumes with $T = T_r$, while using $T = N_r T_r$ (for a given T_r) allows both reduced data volume and higher per-frame detection rates.

In this paper, we address the problem of estimating $\boldsymbol{D} = \{d_{p,n}\}_{p,n}$ from the set of observations $\boldsymbol{Y} = \{y_{p,n}\}_{p,n}$. As discussed in the introduction of this paper, although it is possible to develop batch-based methods for recovering $\boldsymbol{D}$ given all the NP observations [21], [28], such approaches can become computationally prohibitive for large numbers of pixels, but more importantly for long temporal sequences. Moreover, these existing approaches do not specifically deal with time varying scenes, and do not use a spatiotemporal models. Thus, here we adopt a sequential approach where the N frames are processed one by one and only once, allowing for fast estimation and reduced memory requirements. In the remainder of the paper, we thus use (3)-(4) as our observation model.

The next paragraph introduces the Bayesian model and estimation strategy used to process a single frame, assuming that $\boldsymbol{W} = \{w_{p,n}\}_{p,n}$ is known. The generalisation of the proposed updated to online 3D reconstruction, including the sequential estimation of $\boldsymbol{W}$ will be discussed in Section III.

B. Estimation strategy

As mentioned above, we first investigate the estimation of $\boldsymbol{d}_n = \{d_{p,n}\}_p$ from a set of measurements $\boldsymbol{y}_n = \{y_{p,n}\}_p$ associated with the nth frame. Assuming the detection events in different pixels are mutually independent (given the other parameters in (4)), the joint likelihood can be expressed as

$$f(\boldsymbol{y}_n|\boldsymbol{z}_n, \boldsymbol{w}_n, \boldsymbol{d}_n) = \prod_{p=1}^{P} f(y_{p,n}|z_{p,n}, w_{p,n}, d_{p,n}), \tag{5}$$

with $\boldsymbol{z}_n = \{z_{p,n}\}_p$, $\boldsymbol{w}_n = \{w_{p,n}\}_p$ and $f(y_{p,n} = \emptyset | z_{p,n} = 0, w_{p,n}, d_{p,n}) = 1$.

To obtain a tractable and computationally efficient ADF-based estimation strategy, we propose to define independent prior distributions for the target ranges in a given frame, i.e., $f(\boldsymbol{d}_n|\boldsymbol{\Theta}_n) = \prod_{p=1}^{P} f(d_{p,n}|\boldsymbol{\theta}_{p,n})$. Despite the apparent lack of prior correlation between the elements of $\boldsymbol{d}_n$ (given the set of parameters in $\boldsymbol{\Theta}_n$), it is possible to enforce ST correlations by defining $\boldsymbol{\Theta}_n$ using $\boldsymbol{d}_{n-1}$, as will be discussed in Section III. For now, let's assume that each distance $d_{p,n}$ is assigned a fully specified mixture of M Gaussian distributions as follows

$$f(d_{p,n}|\boldsymbol{\theta}_{p,n}) \sim \sum_{m=1}^{M} u_{p,n}^{(m)} \mathcal{N}(d_{p,n}; \mu_{p,n}^{(m)}, \sigma_{p,n}^{2(m)}), \tag{6}$$

where $\mathcal{N}(\cdot; \mu_{p,n}^{(m)}, \sigma_{p,n}^{2(m)})$ is a Gaussian distribution with mean $\mu_{p,n}^{(m)}$ and variance $\sigma_{p,n}^{2(m)}$, $\boldsymbol{\theta}_{p,n} = \left\{\mu_{p,n}^{(m)}, \sigma_{p,n}^{2(m)}\right\}_m$ and $\boldsymbol{\Theta}_n = \{\boldsymbol{\theta}_{p,n}\}_p$. The weights $\{u_{p,n}^{(m)}\}_m$ of the Gaussian mixture model (GMM) in (6) satisfy $\sum_{m=1}^{M} u_{p,n}^{(m)} = 1, \forall (p,n)$ and their value, as well as that of M will be discussed in Section III. Since the joint likelihood (5) and the joint prior distribution (6) can be factorised over the P pixels, the resulting posterior distribution given by

$$\begin{aligned} f(\boldsymbol{d}_n|\boldsymbol{y}_n, \boldsymbol{z}_n, \boldsymbol{w}_n, \boldsymbol{\theta}_{p,n}) &\propto f(\boldsymbol{y}_n|\boldsymbol{z}_n, \boldsymbol{w}_n, \boldsymbol{d}_n) f(\boldsymbol{d}_n|\boldsymbol{\Theta}_n)) \\ &\propto \prod_{p=1}^{P} f(y_{p,n}|z_{p,n}, w_{p,n}, d_{p,n}) f(d_{p,n}|\boldsymbol{\theta}_{p,n}), \end{aligned} \tag{7}$$

can also be factorised over the P pixels and the P range parameters in $\boldsymbol{d}_n$ can thus be estimated independently, in a parallel manner. Consequently, we simply summarise the update for one parameter $d_{p,n}$, i.e., for pixel p. If $z_{p,n} = 0$, $d_{p,n}$ does not appear in the data likelihood. In that case, the posterior distribution of $d_{p,n}$ reduces to its prior (6). If $z_{p,n} = 1$, the posterior distribution of $d_{p,n}$ is the following mixture

$$\begin{aligned} &f(d_{p,n}|y_{p,n}, z_{p,n} = 1, w_{p,n}, \boldsymbol{\theta}_{p,n}) \\ &\quad \propto f(d_{p,n}|\boldsymbol{\theta}_{p,n}) f(y_{p,n}|z_{p,n} = 1, w_{p,n}, d_{p,n}), \end{aligned} \tag{8}$$

which is a mixture of $2M$ Gaussian distributions when $f_s(\cdot)$ is also Gaussian. Note that although $f(d_{p,n}|y_{p,n}, z_{p,n} = 1, w_{p,n}, \boldsymbol{\theta}_{p,n})$ seems to be only known up to a multiplicative constant, its normalising constant, as well as the mixture weights and the mean/variances of each component of the mixture can be computed analytically by integrating (8) with respect to (w.r.t.) $d_{p,n}$. The moments of $f(d_{p,n}|y_{p,n}, z_{p,n} = 1, w_{p,n}, \boldsymbol{\theta}_{p,n})$, and in particular its mean and variance can then be computed as for any mixture of distributions [36, Chap. 1]. These summary statistics are then used to obtain a point estimate (i.e., the mean) of $d_{p,n}$, as well as corresponding measures of uncertainty (through the variance). When $f_s(\cdot)$ is not Gaussian, it is in general not possible to compute analytically the mean and variance of $f(d_{p,n}|y_{p,n}, z_{p,n} = 1, w_{p,n}, \boldsymbol{\theta}_{p,n})$, but it is possible to resort to numerical integration tools [37], [38] such as Gaussian quadrature or Laplace approximation to approximate the integrals

$$\int f_s(y_{p,n} - 2d_{p,n}/c) \mathcal{N}(d_{p,n}; \mu_{p,n}^{(m)}, \sigma_{p,n}^{2(m)}) \mathrm{d}d_{p,n}, \tag{9}$$

and in turn the moments of $f(d_{p,n}|y_{p,n}, z_{p,n} = 1, w_{p,n}, \boldsymbol{\theta}_{p,n})$.

III. Online estimation

A. Approximation using Assumed Density Filtering

Estimating the posterior mean and variance of $d_{p,n}$ presents a great advantage for online estimation, beyond simply providing summary statistics about the current range profile. It allows, by propagating simply the first and second-order moments of the current posterior distributions, the use of a tractable adaptive estimation procedure. Indeed, if the prior distribution of $d_{p,n}$ consists of M components (as in (6)), its posterior will contain $2M$ components and if a classical Gaussian random walk is then used to model $f(d_{p,n+1}|d_{p,n})$, the posterior distribution of $d_{p,n+1}$ will present $4M$ terms after marginalisation of $d_{p,n}$. That number will thus increase prohibitively as n increases. The basic principle of assumed density filtering in this case is to approximate

$f(d_{p,n}|y_{p,n}, w_{p,n}, \boldsymbol{\theta}_{p,n})$ by a more tractable distribution that can then be used to build a new prior distribution for $d_{p,n+1}$. While it is possible to construct complex approximations of $f(d_{p,n}|y_{p,n}, w_{p,n}, \boldsymbol{\theta}_{p,n}) \propto f(d_{p,n}|\boldsymbol{\theta}_{p,n})$ using a fixed (reduced) number of Gaussian components, here we simply use an approximation based on a single Gaussian $q(d_{p,n})$. In a similar fashion to classical assumed density filtering [32], [33] and expectation-propagation [34], this approximation is found by minimising the following Kullback-Leibler divergence

$$KL\left[f(d_{p,n}|y_{p,n}, w_{p,n}, \boldsymbol{\theta}_{p,n})||q_{p,n}(d_{p,n})\right] \tag{10}$$

w.r.t. $q_{p,n}(d_{p,n})$ which belongs to the family of Gaussian distributions. This minimisation reduces to matching the mean and variance of $f(d_{p,n}|y_{p,n}, w_{p,n}, \boldsymbol{\theta}_{p,n})$ and $q_{p,n}(d_{p,n})$, hence the discussion about the estimation of the moments of $f(d_{p,n}|y_{p,n}, w_{p,n}, \boldsymbol{\theta}_{p,n})$ in Section II-B.

B. Spatiotemporal dynamic model for the range profile

A classical choice for modelling relatively slowly evolving parameters relies on (Gaussian) random walks. Whilst this approach is easy to implement, it does not allow, using simply $f(d_{p,n+1}|d_{p,n}), \forall p$, for rapid changes as might occur when the imaging system or the scene moves orthogonally to the direction of observation, whereby a foreground object can disappear from one pixel and appear in neighbouring pixels. To alleviate issues associated with such changes while keeping the estimation strategy tractable, we define, for each pixel, a local neighbourhood $\mathcal{V}_p$ of M neighbours (including the current pixel) and define the following prior model $f(d_{p,n+1}|\boldsymbol{\theta}_{p,n+1})$

$$\propto \sum_{p' \in \mathcal{V}_p} \nu_{p'} f_{\gamma^2}(d_{p,n+1}|d_{p',n}) q_{p',n}(d_{p',n}), \tag{11}$$

where $\{q(d_{p,n})\}_{p,n}$ are the Gaussian approximating posterior distributions of $\{d_{p,n}\}_{p,n}$ computed by minimising (10) and $f_{\gamma^2}(\cdot|d_{p',n}) = \mathcal{N}(\cdot; d_{p',n}, \gamma^2)$ is a Gaussian random walk which models (through its variance γ^2) the expected amount of movement of the objects of the scenes along the direction of observation, between two frames. More precisely, this mostly allows displacements smaller than 3γ along that direction (using the three-sigma rule of thumb). To incorporate larger displacements which cannot be captured the proposed GMM, γ can be increased but this makes the model in Eq. (11) less informative and the results more prone to noise. Note that in practice γ^2 should be smaller than the variance of the likelihood $f_s(\cdot)$ of a signal detection event (e.g., s^2 in the Gaussian case) for the inference process to benefit from the ST model in Eq. (11). This is however the case for current Lidar systems using fast laser repetition rates.

In a similar fashion to (8), Eq. (11) is a finite mixture a M Gaussian distributions whose weights, and individual means and variances, gathered in $\boldsymbol{\theta}_{p,n+1}$ can be computed analytically by integration of the right-hand side of (11) w.r.t. $\{d_{p',n}\}_{p' \in \mathcal{V}_p}$. Using this strategy, the number of components of $f(d_{p,n+1}|\boldsymbol{\theta}_{p,n+1})$ remains the same as for $f(d_{p,n}|\boldsymbol{\theta}_{p,n})$, that is, M. The parameter M, which controls the size of the neighbourhood structure, will depend on the actual distance between pixels and the expected transverse velocity of the dynamic objects of the scene. In practice, the period T is expected to be short enough such that an object present in a given pixel is not expected to move by more than a few pixels (in the transverse direction) and M can be kept small. In this work, we used $M = 5$ using 4 neighbouring pixels, assuming the scene is moving slowly compared to the time scale given by T. Larger neighbourhoods can be used if objects are expected to move by several pixels in the image plane between successive frames. Note that the overall computational cost of the method per frame will grow linearly with M (the number of modes in each posterior distribution (8) is $2M$).

C. Estimation of the other model parameters

Interestingly, the proposed 3D reconstruction method does not rely on the knowledge of the detection probabilities $\{\boldsymbol{\pi}_n\}_n$ since they do not intervene in the estimation of $\boldsymbol{d}_n$ which only relies on $\{\boldsymbol{y}_n\}_n$. In particular, this method does require knowledge of the number N_r of illumination periods during each frame, which is used in the probabilities of detection $\{\tilde{\pi}_{p,n}\}_{p,n}$ (see discussion below Eq. (4)). Thus, the only important and generally unknown parameters are the probabilities of signal detection events in $\boldsymbol{W}$.

In a similar fashion to the approach we proposed for $\boldsymbol{D}$, the elements of $\boldsymbol{W}$ can be included in a Bayesian model and assigned sequentially prior distributions for online estimation, i.e., by computing the posterior distribution of $(\boldsymbol{d}_n, \boldsymbol{w}_n)$ at each frame, and by approximating this distribution to build a tractable prior distribution $f(\boldsymbol{d}_{n+1}, \boldsymbol{w}_{n+1}|\boldsymbol{d}_n, \boldsymbol{w}_n)$. However, this is not the approach we adopt here as it makes the estimation procedure more computationally demanding, in particular when computing the marginal moments, or more generally expectations w.r.t the posterior distribution of $(\boldsymbol{d}_n, \boldsymbol{w}_n)$ during the KL divergence minimisation.

Instead, we use the following simple heuristic method which provides satisfactory results in practice. Let $\bar{\boldsymbol{w}}_n$ be an estimate of $\boldsymbol{w}_n$ obtained from the previously observed data $\{\boldsymbol{y}_n\}_{n=1,\ldots,n-1}$. Our aim here is to propose an estimate $\bar{\boldsymbol{w}}_{n+1}$ of $\boldsymbol{w}_{n+1}$, which depends on $\bar{\boldsymbol{w}}_n$ and the data $\boldsymbol{y}_n$. We first define an instantaneous estimator $\hat{\boldsymbol{w}}_n = \{\hat{w}_{p,n}\}_p$ with $\hat{w}_{p,n} = \bar{w}_{p,n}$ if $y_{p,n} = \emptyset$. If $y_{p,n} \neq \emptyset$, $\hat{w}_{p,n}$ is obtained from (8) where $\boldsymbol{w}_n$ has been replaced by $\bar{\boldsymbol{w}}_n$. More precisely, the posterior distribution $f(d_{p,n}|y_{p,n}, z_{p,n} = 1, \bar{w}_{p,n}, \boldsymbol{\theta}_{p,n})$ consists of a mixture of $2M$ Gaussian distributions with different weights. One half of the Gaussian components correspond to possible positions of the surface assuming the detected photon is a background photon. They are obtained by multiplying the GMM prior by the uniform distribution in Eq. (4). The other M Gaussian components correspond to possible positions of the surface assuming the detected photon is a "signal" photon and they are obtained by multiplying the GMM prior with the term involving $f_s(\cdot)$ in Eq. (4). Thus $\hat{\boldsymbol{w}}_n$ is obtained by summing the weights of the latter M components, which corresponds to the posterior probability of the current detection event to be a signal detection. The updated vector of probabilities is obtained using $\bar{\boldsymbol{w}}_{n+1} = (1-\alpha)\bar{\boldsymbol{w}}_n + \alpha\hat{\boldsymbol{w}}_n$, where $\alpha \in (0;1)$ is an attenuation parameter to be tuned depending on the expected variations of $\boldsymbol{w}_n$ over time.

Note that it is also possible to apply a smoothing post-processing step, e.g., standard gaussian filtering to $\bar{\boldsymbol{w}}_{n+1}$ to further refine the estimate of $\boldsymbol{w}_{n+1}$ since these parameters are often expected to be spatially correlated in each frame. As mentioned above, this strategy is simple and does not significantly degrade the performance of the 3D reconstruction method in most scenarios. The pseudo-code of the proposed method, referred to as O3DSP (for Online 3D reconstruction using Single-Photon data) is presented in Algo. 1.

ALGORITHM 1

O3DSP algorithm

Fixed input parameters: Variance of RW for dynamic model: γ^2, Neighbourhood size M, temporal smoothing parameter α, parameter of GMM ν.
Initialization $(n = 0)$
Set $(\bar{w}_{p,1}, q_{p,0}(\cdot)), \forall p$.
for $n = 1, \dots N$ **do**
  for $p = 1, \dots P$ **do**
    Compute prior model $f(d_{p,n}|\boldsymbol{\theta}_{p,n})$ from (11).
    Compute exact posterior distribution $f(d_{p,n}|y_{p,n}, \bar{w}_{p,n}, \boldsymbol{\theta}_{p,n})$ in (8).
    Compute $q_{p,n}(d_{p,n})$ using (10).
    Set the estimated depth $\hat{d}_{p,n}$ as the mean of $q_{p,n}(d_{p,n})$.
    if $y_{p,n} = \emptyset$ **then**
      Set $\hat{w}_{p,n} = \bar{w}_{p,n}$.
    else
      Compute $\hat{w}_{p,n}$ using (8) by replacing $\boldsymbol{w}_n$ by $\bar{\boldsymbol{w}}_n$.
    end if
  end for
  Compute $\bar{\boldsymbol{w}}_{n+1} = (1-\alpha)\bar{\boldsymbol{w}}_n + \alpha\hat{\boldsymbol{w}}_n$.
  Optional: Apply smoothing operator to $\bar{\boldsymbol{w}}_{n+1}$.
end for

Another important issue that might arise is the occurrence of a new object in the field of view. A particularly challenging scenario is the appearance of an object initially occluded by another object. In such cases, it is possible to add an extra component in (11), e.g., whose mean and variance can be related to the mean/median and dispersion of the $\{d_{p',n}\}_{p'}$, respectively. This approach would be efficient to capture new objects appearing between a foreground object and the background. However, as will be shown in Section IV, such extra term does not seem necessary as the proposed ST model naturally enforces large variances around edges, which in turn allows initially occluded to be detected. Note that more complex and principled strategies should be develop to handle more challenging occlusion scenarios and situations where pixels do not contain any objects, which are out of scope of this work. This point will be discussed in the conclusion of this study. New objects can also enter the field of view from any side. To address this problem, we include, for the pixels around the edges of the image, and additional Gaussian component (with a large variance) in the mixture (11) such that the resulting prior allows at the same time, ranges similar to those in nearby pixels but also significantly different ranges induced by the presence of new objects.

Finally, the proposed algorithm can also be applied in the presence of faulty pixels for which $\pi_{p,n} = 0$. For these pixels, the range information will be inferred using the inpainting capability of the model in (11).

IV. RESULTS

In this section, we discuss the performance of O3DSP through a series of experiments conducted with simulated data whereby ground truth is available for comparison. We first investigate the main parameters influencing the reconstruction performance using individual pixels, i.e., without accounting for information provided by neighbouring pixels. Then, we investigate the reconstruction of static and dynamic scenes using photon-starved measurements.

A. Single-pixel experiments

In Sections II and III, we have assumed that the measured times of arrivals follow continuous distributions, i.e., they are either uniformly distributed over $[0; T_r)$ or Gaussian distributed. However, SPAD detectors have a finite timing resolution, whereby the measured times of arrival follow discrete distributions defined on a finite support. Fortunately, state-of-the-art SPADs [19]–[21] present a timing resolution which is much smaller that the support of $f_s(\cdot)$ and thus than T_r. Consequently, assuming continuous measurements does not significantly bias the estimation performance. Should the temporal resolution of the SPADs be coarser, O3DSP can still be applied using dither on the discrete measured ToAs [39].

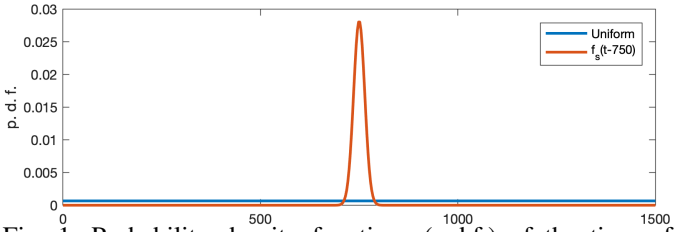

Fig. 1: Probability density functions (p.d.f.) of the time of arrival of a signal photon (red) for $d = 750$ and a background photon (blue), for $T_r = 1500$ and $s^2 = 200$.

In all the simulation results presented in this paper, we use the arbitrary illumination period $T_r = 1500$ (unless stated otherwise) and $f_s(\cdot)$ is modelled by a Gaussian distribution with variance $s^2 = 200$ and without loss of generality, with use $c/2 = 1$. The distributions of the times of arrival of signal and background photons for $d = 750$ are depicted in Fig. 1. To initialise the algorithm, we used $\bar{w}_{p,0} = 0.5, \forall p$ and the Gaussian initial approximations $q_{p,0}(\cdot), \forall p$ are set identically such that their mean is $T_r/2$ and their variance allows the entire interval $(1, T_r)$ to be in the high probability region. This leads to a weakly informative initialisation that we use to assess the convergence of the algorithm. As will be discussed below, more efficient initialisations can also be used.

First, we investigate, the impact of $w_{p,n}$ on the estimation of $d_{p,n}$ for a given probability of detection $\pi_{p,n}$. Here the number of frames is set to $N = 500$, $\pi_{p,n} = 0.5$, $d_{p,n} = 300$ and $\alpha = 0.01$. Figs. 2 and 3, compare the convergence of $\{d_{p,n}\}_n$ and $\{\bar{w}_{p,n}\}_n$ for $w_{p,n} = 0.8$ (Fig. 2) and $w_{p,n} = 0.3$ (Fig. 3). The top subplots depict the frames during which background (in black) and signal (in red) detections are recorded. The middle subplots depict the mean (red lines) and ± 3 standard deviation intervals (black dashed lines) obtained

by minimising (10). The bottom subplots represent the online estimates $\{\bar{w}_{p,n}\}_n$ (red lines) of $w_{p,n}$. As can be seen from Figs. 2 and 3, the estimate of $d_{p,n}$ converges faster with $w_{p,n} = 0.8$ than with $w_{p,n} = 0.3$ (faster convergence to the ground truth and smaller uncertainty). This phenomenon is to be expected as the number of signal detections increases with $w_{p,n}$, which in turn increases the amount of information about $d_{p,n}$. On the other hand, the convergence of $\bar{w}_{p,n}$ seems similar in both cases (around 200-300 frames).

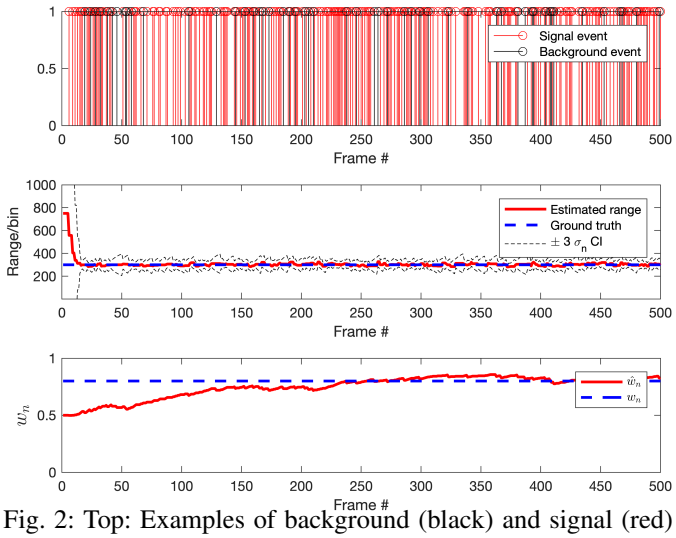

Fig. 2: Top: Examples of background (black) and signal (red) detection events for $N = 500$, $\pi_{p,n} = 0.5$, $w_{p,n} = 0.8$. Middle: Estimation of $\{d_{p,n}\}_n$ for $\alpha = 0.01$ and $\gamma^2 = 100$. Bottom: online estimates $\{\bar{w}_{p,n}\}_n$ (red lines) of $w_{p,n}$ for $\alpha = 0.01$ and $\gamma^2 = 100$.

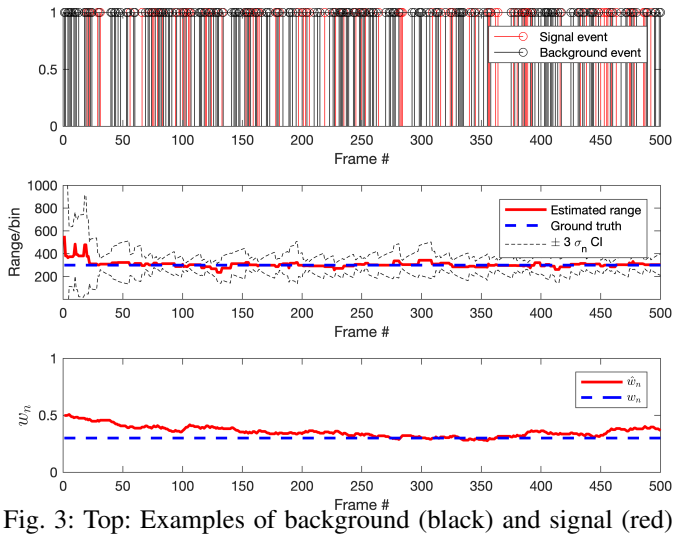

Fig. 3: Top: Examples of background (black) and signal (red) detection events for $N = 500$, $\pi_{p,n} = 0.5$, $w_{p,n} = 0.3$. Middle: Estimation of $\{d_{p,n}\}_n$ for $\alpha = 0.01$ and $\gamma^2 = 100$. Bottom: online estimates $\{\bar{w}_{p,n}\}_n$ (red lines) of $w_{p,n}$ for $\alpha = 0.01$ and $\gamma^2 = 100$.

Fig. 4 shows the estimation of $d_{p,n}$ and $w_{p,n}$ with $\pi_{p,n} = 0.8$, $d_{p,n} = 300$ and $w_{p,n} = 0.3$. As expected, the convergence of $\{d_{p,n}\}_n$ is faster than in Fig. 3 since its estimation is directly related to the number of signal detections which increases with $\pi_{p,n}$ (for a fixed $w_{p,n}$).

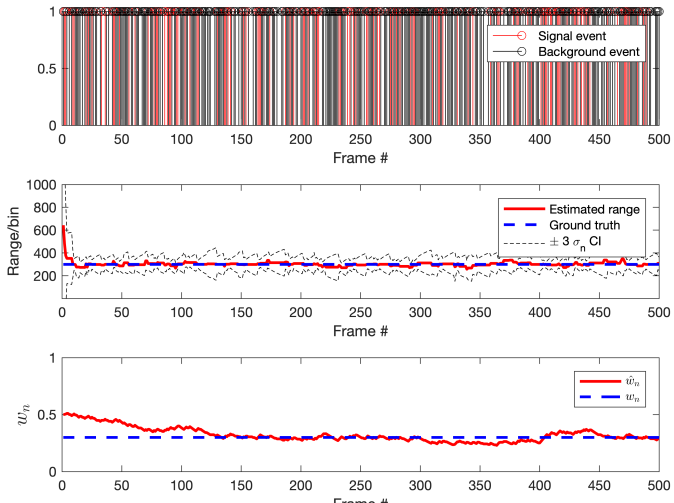

Fig. 4: Top: Examples of background (black) and signal (red) detection events for $N = 500$, $\pi_{p,n} = 0.8$, $w_{p,n} = 0.3$. Middle: Estimation of $\{d_{p,n}\}_n$ for $\alpha = 0.01$ and $\gamma^2 = 100$. Bottom: online estimates $\{\bar{w}_{p,n}\}_n$ (red lines) of $w_{p,n}$ for $\alpha = 0.01$ and $\gamma^2 = 100$.

Reducing the probability of detection has an impact on the estimation of $d_{p,n}$ and $w_{p,n}$, as can be seen in Fig. 5, where $\pi_{p,n} = 0.1$ and $w_{p,n} = 0.3$. In this case, with an average of 50 detection events for $N = 500$ frames (30% of which being signal detections), the convergence speed of $\{\bar{w}_{p,n}\}_n$ is reduced and the uncertainty about $d_{p,n}$ increases due to the lack of information provided by the data. In such difficult scenarios, the proposed method might not converge toward the correct solution without using additional information, which is a well known potential limitation of ADF [34]. However, as will be shown in Section IV-B, the proposed ST model using information contained in neighbouring pixels (see (11)) yields satisfactory results in the photon-starved regimes considered here.

We also evaluate the performance of O3DSP by analysing a single-pixel measurement where the object range describes a sine wave and where $w_{p,n}$ experiences two sudden changes (see Fig. 6). This figure has been obtained with $\pi_{p,n} = 0.5$. As can be seen in the top and bottom subplots of Fig. 6, the probability of signal detection $w_{p,n}$ is changed successively from $w_{p,n} = 0.3$ to $w_{p,n} = 0.8$ and back to $w_{p,n} = 0.3$. This figure shows that O3DSP is able to satisfactorily track the changes of $d_{p,n}$ without noticeable delay and that about $200 - 300$ frames are required for $\bar{w}_{p,n}$ to converge around the correct value.

To highlight the benefits of our online approach over batch-based methods we also consider the single-pixel measurements used in Fig. 6 and compare our approach to the classical cross-correlation method (see details in [21]). This approach is chosen as it is the fastest batch-based method which processes all the pixels independently. Although the comparison could have been performed using image sequences and more advanced methods, the competing methods would have led to

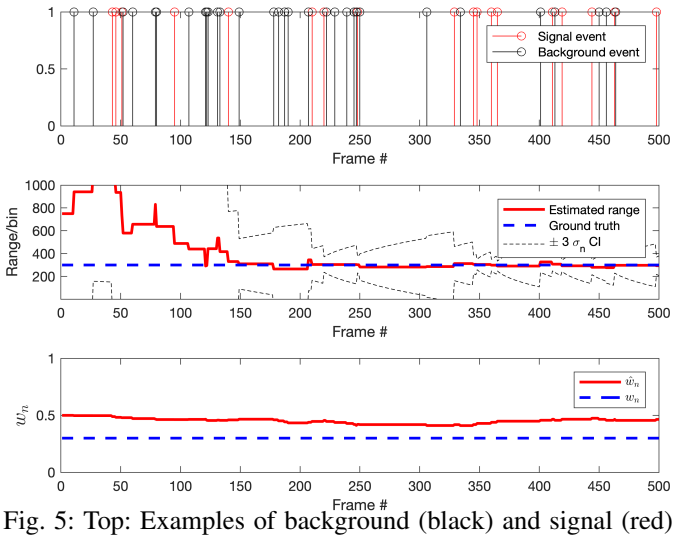

Fig. 5: Top: Examples of background (black) and signal (red) detection events for $N = 500$, $\pi_{p,n} = 0.1$, $w_{p,n} = 0.3$. Middle: Estimation of $\{d_{p,n}\}_n$ for $\alpha = 0.01$ and $\gamma^2 = 100$. Bottom: online estimates $\{\bar{w}_{p,n}\}_n$ (red lines) of $w_{p,n}$ for $\alpha = 0.01$ and $\gamma^2 = 100$.

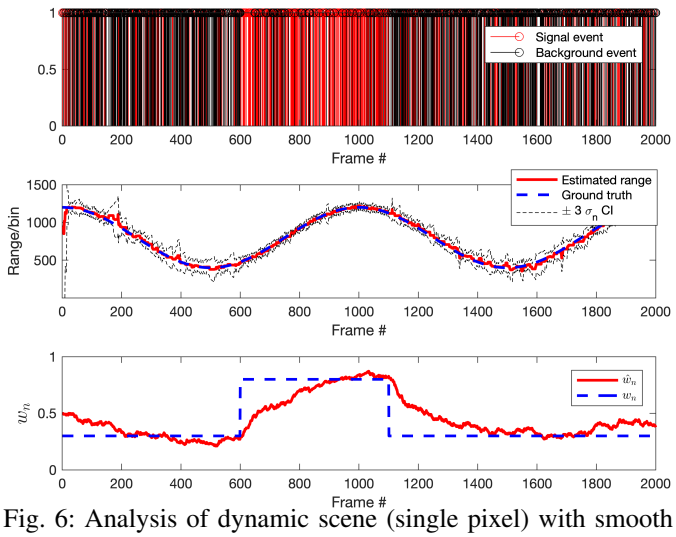

Fig. 6: Analysis of dynamic scene (single pixel) with smooth changes of $d_{p,n}$ and sudden changes of $w_{p,n}$. Top: Examples of background (black) and signal (red) detection events for $N = 2000$, $\pi_{p,n} = 0.5$. Middle: Estimation of $\{d_{p,n}\}_n$ for $\alpha = 0.01$ and $\gamma^2 = 100$. Bottom: online estimates $\{\bar{w}_{p,n}\}_n$ (red lines) of $w_{p,n}$.

significantly higher computational costs. To apply the cross-correlation, we first discretise the detection events uniformly over $[0; T_r)$ with a stepsize of 1, which is much smaller and $s^2 = 200$ such that the discretisation bias can be neglected. For each batch of N_0 frames, the depth is then estimated by finding the delay that maximises the cross-correlation between the histogram of times of arrival within this batch and the discretised version of $s(\cdot)$. Fig. 7 compares the depth estimates obtained via cross-correlation for batches of $N_0 = 10$, $N_0 = 50$ and $N_0 = 100$ frames to those obtained using O3DSP. While small values of N_0 can lead to more accurate instantaneous estimates of the ranges, this figure shows that the results are also more sensitive to background detections due to the small number of detections within each batch of N_0 frames. Note that in extreme cases where $\pi_{p,n}$ is small, there might even be no detection in some batches. Note also in the top plot of Fig. 7 that the performance of the cross-correlation method is affected by the relative amount of background detections (larger errors for $w_{p,n} = 0.3$ than for $w_{p,n} = 0.8$, i.e., for $n \in [600; 1100]$). Here, we initialised the proposed method using weakly informative parameters but it could be initialised using a batch-based method, such as cross-correlation, with the first few frames to improve the convergence speed.

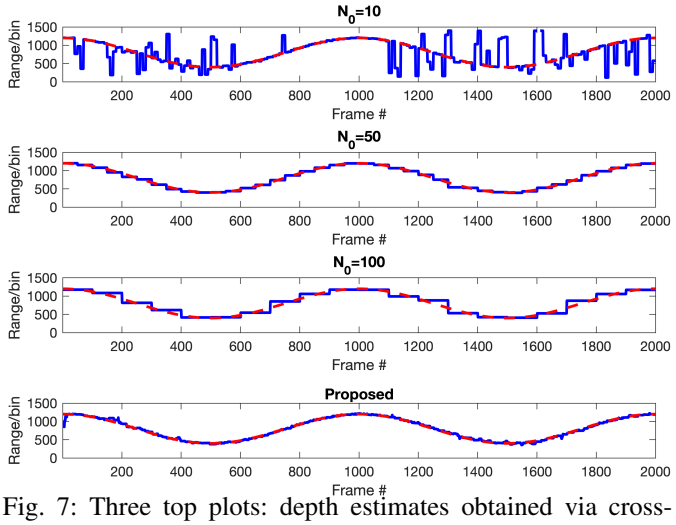

Fig. 7: Three top plots: depth estimates obtained via cross-correlation for batches of $N_0 = 10$, $N_0 = 50$ and $N_0 = 100$. Bottom: Estimation of $\{d_{p,n}\}_n$ using the proposed method with $\alpha = 0.01$ and $\gamma^2 = 100$. The solid blue (resp. dashed red) curves depict the estimated (resp. actual) ranges. The data used to generate this figure are the same as for Fig. 6.

B. Analysis of static and dynamic 3D scenes

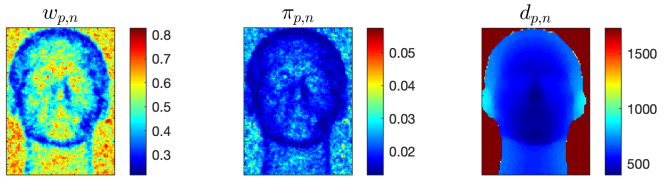

Fig. 8: Ground truth parameters used for assessing the performance of the proposed method for reconstruction of a static scene.

In this section, we first analyse the performance and convergence speed of O3DSP using simulated data based on real Lidar measurements conducted in [21], [22]. More precisely, we consider a series of $N = 5000$ frames composed of 129×95 pixels and associated with a static scenes whose range profile, probabilities of signal detection $\{w_{p,n}\}_p$ and probabilities of detection $\{\pi_{p,n}\}_p$ are depicted in Fig. 8. Here, we used $T_r = 2500$. Note that for most pixels $\pi_{p,n} < 5\%$,

which corresponds to realistic observation conditions in the photon-starved regime.

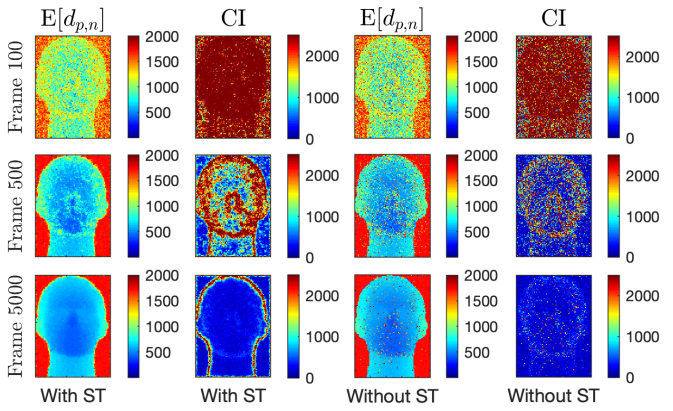

Fig. 9: Online range estimation performance for a static scene: Estimated instantaneous means (first and third columns) and ± 3 standard deviation confidence intervals (CI) (second and fourth column) after $N = 100$ frames (top), $N = 500$ frames (middle) and $N = 5000$ frames (bottom). The two columns on the left-hand side (resp. right-hand side) have been obtained with (resp. without) the proposed ST model.

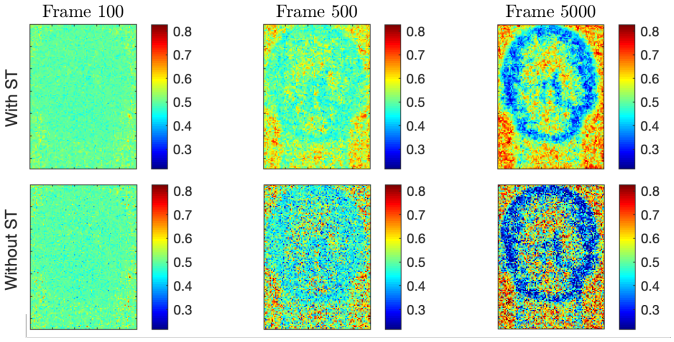

Fig. 10: Online range estimation performance for a static scene: Estimated images of $\bar{\boldsymbol{w}}_n$ after $N = 100$ frames (left), $N = 500$ frames (middle) and $N = 5000$ frames (right). The top (resp. bottom) row has been obtained with (resp. without) the proposed ST model.

First, we compare the performance of O3DSP processing all the pixels independently, i.e., without smoothing of $\bar{\boldsymbol{w}}$ and with $\nu = 1$ to the version using the proposed ST model. In this case, we used $M = 5$ neighbours, $\nu = 0.99$ and $\bar{\boldsymbol{w}}$ was smoothed using a Gaussian filter with standard deviation 0.5 pixels. In the two scenarios, we used $(\alpha, \gamma^2) = (0.1, 10)$. Fig. 9 depicts the estimated means and variances of the range estimates after 100, 500 and 5000 frames (top to bottom), with (left columns) and without (right columns) the ST model. These results illustrate the benefits of the ST model which improves the convergence speed of the algorithm and which reduces the number of isolated pixels with poorly estimated range (see bottom row of Fig. 9). O3DSP with the ST model is able to clearly identify regions of high uncertainty, i.e., the boundaries of the head where the range is likely to change suddenly, should the head move. Moreover, the uncertainty increases with the range difference between close pixels. For instance, the uncertainty is larger at the boundary of the head than in the neck/chin boundary. Similarly, Fig. 10 compares the estimated values of $\{w_{p,n}\}$ obtained with (top row) and without (bottom row) the ST model and spatial smoothing of $\bar{\boldsymbol{w}}$. Here $\bar{w}_{p,1}$ has been set to $\bar{w}_{p,1} = 0.5, \forall p$. This figure illustrates that the ST model not only improves the depth estimation but also the estimation of $\bar{\boldsymbol{w}}$ when used in conjunction with the spatial smoothing of $\bar{\boldsymbol{w}}$, which further improves the convergence of $\bar{\boldsymbol{w}}$.

To assess quantitatively the convergence of the method, we use the range root mean square error (RMSE) defined as

$$\text{RMSE}_n = \sqrt{\frac{1}{P}||\hat{\boldsymbol{d}}_n - \boldsymbol{d}_n||_2^2}, \tag{12}$$

where $\boldsymbol{d}_n$ and $\hat{\boldsymbol{d}}_n$ are the actual and estimated range profiles in the frame n, respectively. Fig. 11 confirms that the proposed ST model improves the convergence speed and estimation performance in terms of RMSE. To ease the visualisation of these results, the generated data associated with this static scene, as well as the estimated range profiles are provided in a supplementary video (Video 1) associated with this paper.

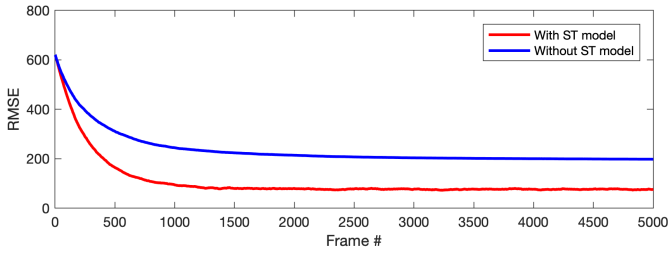

Fig. 11: Range RMSEs obtained with (red lines) and without (blue lines) the proposed spatiotemporal (ST) model for the static scene considered in Fig. 8.

For completeness, we also generated data with the same parameters as above but with probabilities of detection $\{\pi_{p,n}\}_{p,n}$ multiplied by 10, when compared to those depicted in Fig. 8 (middle subplot), leading to an average probability of detection of 20% per pixel and per frame. Fig. 12 compares the convergence of the RMSEs for the original data (referred to as "low detection probability") and the new data set (referred to as "high detection probability"). As expected, increasing $\pi_{p,n}$ yields faster convergence and lower RMSEs at convergence due to the additional amount of (more frequent) detections available.

Finally, we applied our algorithm to the 3D reconstruction of a synthetically generated dynamic scene which consists of flat homogeneous rectangles, in front of a static backplane. For this experiment, we used $N = 2400$ frames of 100×100 pixels with $\pi_{p,n} = 0.5, \forall (p, n)$ and $T_r = 2500$. During the first 800 frames, two objects are present. The first object is static while the second object describes a counterclockwise circular trajectory, centred at the centre of the image (rotation of $0.45°$ per frame). During this rotation, the second object completely occludes the first one which then reappears. During

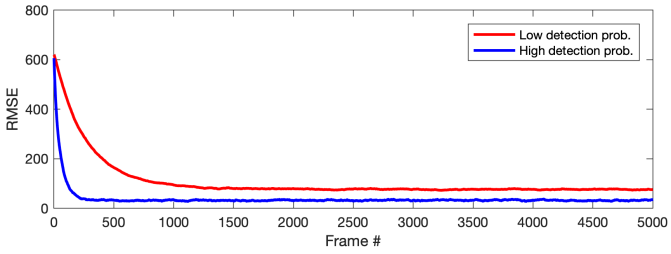

Fig. 12: Range RMSEs obtained for the static scene using the parameters defined in Fig. 8 (red lines) and by multiplying the probabilities of detection in Fig. 8 (middle subplot) by 10 (blue lines).

the next 800 frames, the firs objects disappears suddenly and the second one describes the same trajectory as before (while its range remains unchanged) but its size varies. At frame 1600, a third object enters the field of view from the left and describes an horizontal movement (constant range), while the first object moves away from the backplane. Moreover, we set $w_{p,n} = 0.5$ for the pixels associated with the backplane and $w_{p,n} = 0.7$ for those associated with the two objects. This scenario is chosen to assess the robustness of the algorithm to occlusions and appearance of new objects. The parameters of the algorithm have been set to $M = 5$, $\alpha = 0.1$, $\nu = 0.5$ and $\gamma^2 = 100$. The observed data as well as the estimated range profiles are provided in the second supplementary video associated with this paper (see Video 2). As an example, Fig. 13 depicts estimated range profiles and associated uncertainties for three frames, namely before, during, and after the occlusion of one of the objects. Here, the range uncertainty is measured using the width on the confidence intervals (CI) defined as 6 times (± 3) the standard deviations of the approximating Gaussians. For the three frames, we observe, as expected, higher uncertainties at the boundaries of the small rectangles. Moreover, this figure illustrates that the proposed method is able to recover occluded objects when they become visible again.

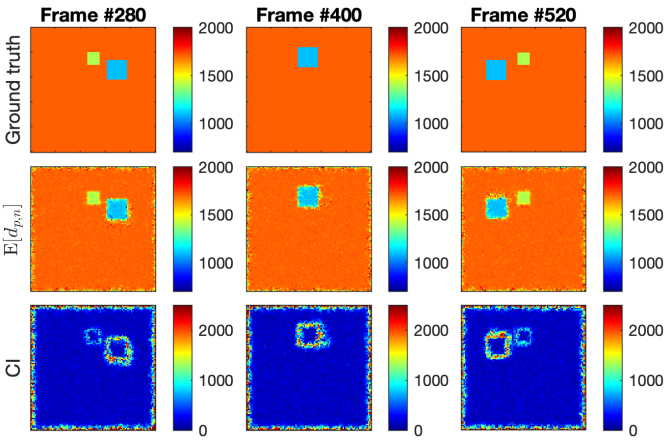

Fig. 13: Example of range estimation for a dynamic scene with occlusion of one object. The full estimated range sequence can be seen in the supplementary Video 2.

As mentioned in Sections II and III, one important property of the method is that, for a given frame, all updates (expect the smoothing step in line 11 of Algo. 1) can be performed in parallel, using only estimates from one previous frame. In this work, the method has been implemented using Matlab 2017b running on a MacBook Pro with 16GB of RAM and a 2.9 GHz Intel Core i7 processor, leading to a average processing time of 4ms per frame (with $P = 10^4$ pixels).

V. Conclusion

In this work, we presented a first 3D reconstruction algorithm using individual photon detection events for online analysis of dynamic scenes. Based on assumed density filtering, the proposed method is computationally efficient as the data are processed partly in a parallel fashion (pixels in a given frame) and sequentially (successive frames). The results presented in this paper have illustrated the flexibility and ability of the method to be used for static and slowly moving scenes (compared to the frame rate). Whilst the code has not been fully optimised, preliminary results conducted with a tailored implementation using a Titan Xp GPU indicate significant computational improvement (well below 1ms per frame), paving the way to new and efficient streaming and processing of data directly from actual SPAD detector arrays. While the proposed method is able to track relatively slow changes of the 3D profile, ongoing work include the development of more sophisticated models, able the better predict the dynamics of the 3D profile and in particular, sudden changes associated with the appearance of objects or the occurrence of new objects. This problem is also related to the potential presence of an unknown number of objects per pixel, as in [16] for instance, which should be addressed in future work, in particular for fast object detection. Although the range estimation does not seem to be significantly affected by the quality of the estimation of the probability of signal detection in the scenarios investigated, it would be also interesting to investigate in future studies whether the proposed methodology can be made more robust to extreme ambient illuminations where $w_{p,n} \ll 1$.

Acknowledgment

This work was supported by the Royal Academy of Engineering under the Research Fellowship scheme RF201617/16/31, the ERC advanced grant C-SENSE, project 694888, and by the Engineering and Physical Sciences Research Council (EPSRC) (grant EP/S000631/1) and the MOD University Defence Research Collaboration (UDRC) in Signal Processing. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research. The authors would also like to thank Dr Aurora Maccarone and Julian Tachella (Heriot-Watt University) for interesting discussions during the preparation of this work.

References

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

[2] C. Mallet and F. Bretar, "Full-waveform topographic Lidar: State-of-the-art," *ISPRS Journal of photogrammetry and remote sensing*, vol. 64, no. 1, pp. 1–16, 2009.

[3] M. A. Canuto, F. Estrada-Belli, T. G. Garrison, S. D. Houston, M. J. Acuña, M. Kováč, D. Marken, P. Nondédéo, L. Auld-Thomas, C. Castanet, D. Chatelain, C. R. Chiriboga, T. Drápela, T. Lieskovský, A. Tokovinine, A. Velasquez, J. C. Fernández-Díaz, and R. Shrestha, "Ancient lowland maya complexity as revealed by airborne laser scanning of northern guatemala," *Science*, vol. 361, no. 6409, 2018. [Online]. Available: http://science.sciencemag.org/content/361/6409/eaau0137

[4] A. M. Wallace, A. McCarthy, C. J. Nichol, X. Ren, S. Morak, D. Martinez-Ramirez, I. H. Woodhouse, and G. S. Buller, "Design and evaluation of multispectral Lidar for the recovery of arboreal parameters," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 8, pp. 4942–4954, 2014.

[5] J. Gao, J. Sun, J. Wei, and Q. Wang, "Research of underwater target detection using a slit streak tube imaging Lidar," in *Proc. Academic International Symposium on Optoelectronics and Microelectronics Technology (AISOMT)*, Harbin, China, Mar. 2012, pp. 240–243.

[6] R. Horaud, M. Hansard, G. Evangelidis, and C. Ménier, "An overview of depth cameras and range scanners based on time-of-flight technologies," *Machine Vision and Applications*, vol. 27, no. 7, pp. 1005–1020, Oct 2016. [Online]. Available: https://doi.org/10.1007/s00138-016-0784-4

[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. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-25-10-11919

[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," 2019.

[9] 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. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-27-4-4590

[10] A. Maccarone, A. McCarthy, X. Ren, R. E. Warburton, A. M. Wallace, J. Moffat, Y. Petillot, and G. S. Buller, "Underwater depth imaging using time-correlated single-photon counting," *Optics express*, vol. 23, no. 26, pp. 33 911–33 926, 2015.

[11] 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, Sept 2017.

[12] A. McCarthy, R. J. Collins, N. J. Krichel, V. Fernández, A. M. Wallace, and G. S. Buller, "Long-range time-of-flight scanning sensor based on high-speed time-correlated single-photon counting," *Appl. Opt.*, vol. 48, no. 32, pp. 6241–6251, Nov 2009. [Online]. Available: http://ao.osa.org/abstract.cfm?URI=ao-48-32-6241

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

[14] R. Tobin, A. Halimi, A. McCarthy, X. Ren, K. J. McEwan, S. McLaughlin, and G. S. Buller, "Long-range depth profiling of camouflaged targets using single-photon detection," *Optical Engineering*, vol. 57, no. 3, p. 031303, 2017.

[15] A. Halimi, R. Tobin, A. McCarthy, S. McLaughlin, and G. S. Buller, "Restoration of multilayered single-photon 3d Lidar images," in *Proc European Signal Processing Conference (EUSIPCO)*, Aug 2017, pp. 708–712.

[16] J. Tachella, Y. Altmann, X. Ren, A. McCarthy, G. Buller, S. McLaughlin, and J. 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. [Online]. Available: https://doi.org/10.1137/18M1183972

[17] J. Tachella, Y. Altmann, J.-Y. Tourneret, and S. McLaughlin, "3D reconstruction using single-photon Lidar data exploiting the widths of the returns," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2019.

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

[19] X. Ren, P. W. R. Connolly, A. Halimi, Y. Altmann, S. McLaughlin, I. Gyongy, R. K. Henderson, and G. S. Buller, "High-resolution depth profiling using a range-gated cmos spad quanta image sensor," *Opt. Express*, vol. 26, no. 5, pp. 5541–5557, Mar 2018. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-26-5-5541

[20] R. K. Henderson, N. Johnston, F. Mattioli Della Rocca, H. Chen, D. Day-Uei Li, G. Hungerford, R. Hirsch, D. McLoskey, P. Yip, and D. J. S. Birch, "A 192 x 128 time correlated spad image sensor in 40-nm cmos technology," *IEEE Journal of Solid-State Circuits*, pp. 1–10, 2019.

[21] 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 Transactions on Image Processing*, vol. 25, no. 5, pp. 1935–1946, 2016.

[22] ——, "Target detection for depth imaging using sparse single-photon data," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 3256–3260.

[23] 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. Signal Processing Conference (EUSIPCO)*, Budapest-Hungary, Sep. 2016, pp. 86–90.

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

[25] D. Shin, F. Xu, D. Venkatraman, R. Lussana, F. Villa, F. Zappa, V. K. Goyal, F. N. Wong, and J. H. Shapiro, "Photon-efficient imaging with a single-photon camera," *Nature Communications*, vol. 7, 2016.

[26] Y. Altmann and S. McLaughlin, "Range estimation from single-photon lidar data using a stochastic EM approach," in *Proc. Signal Processing Conference (EUSIPCO)*, Rome, Italy, Sep. 2018, pp. 1–5.

[27] D. B. Lindell, M. OToole, and G. Wetzstein, "Single-Photon 3D Imaging with Deep Sensor Fusion," *ACM Trans. Graph. (SIGGRAPH)*, no. 4, 2018.

[28] 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.

[29] J. Rapp, Y. Ma, R. M. A. Dawson, and V. K. Goyal, "Dead time compensation for high-flux ranging," 2018.

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

[31] C. K. Chui and G. Chen, *Kalman Filtering with Real-time Applications*. Berlin, Heidelberg: Springer-Verlag, 1987.

[32] S. L. Lauritzen, "Propagation of probabilities, means, and variances in mixed graphical association models," *Journal of the American Statistical Association*, vol. 87, no. 420, pp. 1098–1108, 1992. [Online]. Available: http://www.jstor.org/stable/2290647

[33] X. Boyen and D. Koller, "Tractable inference for complex stochastic processes," in *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, ser. UAI'98. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1998, pp. 33–42. [Online]. Available: http://dl.acm.org/citation.cfm?id=2074094.2074099

[34] T. P. Minka, "Expectation propagation for approximate bayesian inference," in *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 2001, pp. 362–369.

[35] A. J. Haug, *Bayesian estimation and tracking: a practical guide*. Hoboken, NJ: John Wiley & Sons, 2012. [Online]. Available: http://cds.cern.ch/record/1487095

[36] S. Frühwirth-Schnatter, *Finite Mixture and Markov Switching Models*, ser. Springer Series in Statistics. Springer New York, 2006. [Online]. Available: https://books.google.co.uk/books?id=f8KiI7eRjYoC

[37] M. P. Wand, J. T. Ormerod, S. A. Padoan, and R. Frhwirth, "Mean field variational Bayes for elaborate distributions," *Bayesian Anal.*, vol. 6, no. 4, pp. 847–900, 2011.

[38] A. Perelli, M. A. Lexa, A. Can, and M. E. Davies, "Denoising message passing for X-ray computed tomography reconstruction," *CoRR*, vol. abs/1609.04661, 2016. [Online]. Available: http://arxiv.org/abs/1609.04661

[39] J. Rapp, R. M. A. Dawson, and V. K. Goyal, "Improving Lidar depth resolution with dither," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, Oct 2018, pp. 1553–1557.