Fast Task-Based Adaptive Sampling for 3D Single-Photon Multispectral Lidar Data

Mohamed Amir Alaa Belmekki, *Student Member, IEEE*, Rachael Tobin, Gerald S. Buller, Stephen McLaughlin, *Fellow, IEEE*, and Abderrahim Halimi, *Senior Member, IEEE*

***Abstract*—3D single-photon LiDAR imaging plays an important role in numerous applications. However, long acquisition times and significant data volumes present a challenge for LiDAR imaging. This paper proposes a task-optimized adaptive sampling framework that enables fast acquisition and processing of high-dimensional single-photon LiDAR data. Given a task of interest, the iterative sampling strategy targets the most informative regions of a scene which are defined as those minimizing parameter uncertainties. The task is performed by considering a Bayesian model that is carefully built to allow fast per-pixel computations while delivering parameter estimates with quantified uncertainties. The framework is demonstrated on multispectral 3D single-photon LiDAR imaging when considering object classification and/or target detection as tasks. It is also analysed for both sequential and parallel scanning modes for different detector array sizes. Results on simulated and real data show the benefit of the proposed optimized sampling strategy when compared to state-of-the-art sampling strategies.**

***Index Terms*—Adaptive sampling, 3D multispectral imaging, single-photon LiDAR, Bayesian estimation, Poisson statistics, robust estimation, classification, target detection.**

I. Introduction

LIGHT detection and ranging (LiDAR) used with time-correlated single-photon detection is receiving significant interest as an emerging approach in numerous applications such as Defence, automotive [1], [2], environmental sciences [3], long-range depth imaging [4]–[9], underwater [10], [11] or through fog [12] depth imaging, and multispectral imaging [13]–[15]. Such a single-photon LiDAR system operates by illuminating the scene using a pulsed laser source, and recording the arrival times of the reflected photons with respect to the time of laser pulse emission. By performing this measurement repeatedly over many laser pulses, it is possible to form a timing histogram from which a high resolution measurement of the photon time-of-flight can be made. By measuring the time-of-flight at each pixel location, it is then possible to make a depth estimate at each part of the scene. This can be done by scanning pixels sequentially, or more efficiently by considering state-of-the-art Single-Photon Avalanche Diode (SPAD) detector arrays which allow the parallel photon acquisition of all pixels [16]. The resulting histograms of counts contain useful information regarding the presence/absence of an object, and allow a 3D model of the observed target to be built using its estimated depth and reflectivity profiles. This process can be repeated using different laser wavelengths to obtain additional multispectral information on the observed scene. Current limitations that preclude extensive use of LiDAR in real world applications include the high acquisition time necessary to collect enough target photons, in addition to the high background level when imaging in ambient light conditions, which affects the quality of the reconstructed 3D scene. Multispectral LiDAR can lead to large data volumes which highlights the need to reduce measurement points and only target informative regions. Indeed, reducing the target's laser power is also important due to eye safety constraints for autonomous navigation, and to obtain energy efficient systems. Minimising illumination levels is also important when considering medical applications such as single-photon microscopy in which the observed cells are light sensitive and subject to photostructural damage.

Many strategies have been proposed to circumvent the above challenges either by improving the acquisition [17]–[21], or by presenting advanced algorithms to restore damaged or sparse photon data [14], [22]–[24] of multidimensional single-photon LiDAR images. We distinguish two main approaches to improve data acquisition using subsampling, those considering spatially structured scanned points, and those based on random points. Structured point scanning includes foveated based scanning, which mimic the vision system found in the animal kingdom [18], [25], [26]. Such systems consider a structured scanning array (e.g., a circle) which allocates denser points in regions of interest and sparser points in the remaining regions. The position of the region of interest is updated dynamically based on the observed scene. The second family includes random scanning points which can be associated with a static or dynamic sampling procedure. The static methods introduced in [19], [21] showed satisfactory results under challenging conditions. Nonetheless, they are still incompatible with real-time requirements and are computationally expensive as they use reversible-jump Markov chain Monte Carlo (RJ-MCMC) methods as an inference tool. Random sampling was also considered in sparse-to-depth reconstruction using deep learning approaches. Methods such

Manuscript received July 16, 2021; revised December 3, 2021; accepted January 25, 2022. Date of publication February 14, 2022; date of current version February 28, 2022. This work was supported in part by the U.K. Royal Academy of Engineering under Research Fellowship Scheme under Grant RF/201718/17128 and in part by EPSRC under Grants EP/T00097X/1, EP/S000631/1, and EP/S026428/1. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Henry Arguello. *(Corresponding author: Mohamed Amir Alaa Belmekki.)*

The authors are with the School of Engineering and Physical Sciences, Heriot-Watt University, EH14 4AS Edinburgh, U.K. (e-mail: mb219@hw.ac.uk; r.tobin@hw.ac.uk; g.s.Buller@hw.ac.uk; s.mclaughlin@hw.ac.uk; a.halimi@hw.ac.uk).

Digital Object Identifier 10.1109/TCI.2022.3150974

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

as [27]–[29] used random scanning to obtain dense depth maps using sparse random measurements acquired from a LiDAR sensor and an RGB image from a camera. However, these approaches require an RGB guide to upsample the sparse map into a dense one. In [20], a dynamic sampling scheme was proposed showing significant improvement in depth restoration over static sampling schemes. However, this algorithm assumes a negligible contribution from ambient light photons which is often not satisfied for real-life scenarios. Deep learning was also investigated for adaptive depth acquisition. The authors of [17] proposed a method that uses an RGB image to approximate the depth map and hence locate regions of interest, which are re-scanned to refine the depth map. An improvement was proposed in [30] by combining the robust reconstruction algorithm [27] and an estimate of uncertainty using ensemble learning to iteratively improve data sampling. However, learning based algorithm requires the availability of the training data and a guiding RGB image to improve performance, and could be sensitive to the differences between training and testing data. Furthermore, all the aforementioned deep learning approaches are only optimized for a depth reconstruction task, and require depth values as input instead of photons detected by the Lidar system.

This paper proposes a new framework for task-optimized adaptive sampling (AS) of the scene to jointly improve both the acquisition and processing of single-photon sensing systems without making use of any other imaging modality. Based on a selected task, the data acquisition step can be optimized by dynamically sampling the most informative locations of a scene as in [20], [31]. As indicated in Fig. 1, the proposed scene-based sampling strategy is based on three main steps, (i) select the points to scan and their dwell time, (ii) use scanned points to perform a task using a statistical framework with uncertainty quantification, (iii) construct a map of regions of interest (ROI) to define the next set of scanning points. In this paper, we demonstrate this framework on a LiDAR 3D imaging application and focus on the task of target detection based on object spectral signatures. A Bayesian framework is adopted to perform this task since (i) it allows regularization of the ill-posed problem resulting from the scene sub-sampling, (ii) it benefits from marginalisation tools that lead to fast analytical estimates of the parameters of interest, and (iii) it quantifies parameter uncertainties which will be used to define the ROI map. The proposed model accounts for data Poisson statistics and parameter prior information, to build a posterior distribution of the parameters of interest. These parameters include spatial labels to locate pixels with or without a reflective surface, the class of each pixel based on a known spectral library, and depth estimates for pixels having an object. The proposed AS framework and Bayesian algorithm are analysed when considering both structured and random sampling scenarios (pixel wise or array scanning), and validated on sparse data with high background levels. The study shows promising results when compared to static or dynamic sampling strategies.

To summarize the main contributions of the paper are:

- The use of a scene-based sampling approach that is optimized for multiple tasks in the context of single-photon LiDAR imaging. This sampling approach allows fast data acquisition by sampling the most informative regions to perform a task on a particular scene.
- A new computationally efficient Bayesian algorithm to perform multispectral classification, depth estimation and target detection.

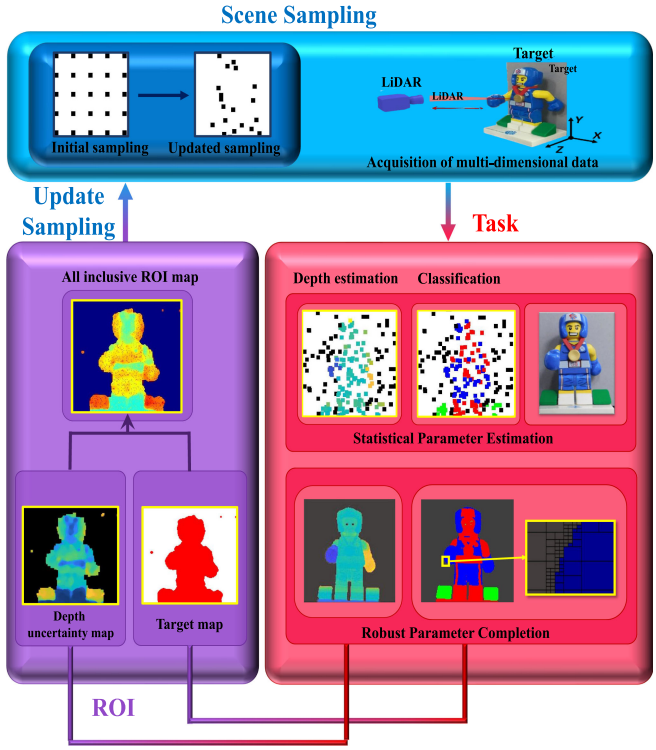


Fig. 1. Representation of the proposed adaptive sampling process. ROI stands for region of interest.

The paper is structured as follows. Section II presents the task-based adaptive sampling approach. The Bayesian model for the classification task and the associated estimation strategy are presented in Sections III and IV, respectively. Results on simulated and real data are analysed in Sections V and VI. The Conclusions and future work are finally reported in Section VII.

II. Task-Optimized Adaptive Sampling

Single-photon LiDAR acquire histograms of counts for each pixel location leading to large data volumes as we increase the spatial resolution (range resolution represented by the number of time bins, or cross-range resolution represented by the number of pixels) and/or the spectral resolution. For instance, a data set with 100×100 pixels, 1000 time bins, and 32 wavelengths can yield an excessively large number of data samples ($>10^8$) which will lead to a significant computational load and prohibitive memory requirements. Adaptive sampling appears as a promising strategy to reduce data volume [17], [20], [25], [26], [30] by focusing the scanning on pixels containing target's returns, and scanning less those only containing background reflections.

The proposed adaptive sampling framework is summarized in Fig. 1. Assuming the presence of a high-resolution (HR) sampling grid of N pixels, the approach aims to iteratively

sample N_s pixels to progressively improve the performance of a pre-defined task (e.g., depth estimation, classification, etc). The location of these N_s pixels is subject to different constraints based on the scanning scenario considered. Indeed, the N_s pixels could be non-uniformly located in the scene when considering pixel-wise scanning of the scene, or have a spatial structure when considering an array scanning system. Both scenarios will be investigated in the following section. Assuming no prior information about the scene of interest, the first iteration will scan N_s pixels uniformly spaced across the scene. The collected samples are then used to perform a specific task such as target detection, classification, depth estimation, etc. The HR sampling grid of size N is constructed by an inpainting process using the cumulative scanned samples and a non-linear operator (in this paper a median filter). The uncertainty of the estimated parameters are then used to establish a map of regions of interest that will serve as a basis for defining the positions and the acquisition times of the new sampled pixels. This iterative procedure will continue until convergence, as indicated in Algo. 1 where i represents the current iteration. The next subsections describe in detail each of these steps.

Algorithm 1: Task-Optimized Adaptive Sampling.

1: Initialization
2: Initialize: N_s, pixel locations $\nu_{N_s}^{(1)}$, pixels acquisition times are $t_{\nu_{N_s}}^{(1)}$, N, conv=0
3: **while** conv$= 0$ **do**
4: Scan N_s pixels with locations $\nu_{N_s}^{(i)}$ and acquisition times $t_{\nu_{N_s}}^{(i)}$
5: Processing step: Fast and robust task performance (object detection, classification, depth estimation, etc)
6: Regions of interest: Computation of a probability map of ROI
7: Update sampling step: Generation of new positions $\nu_{N_s}^{(i)}$ (using a MH algorithm) and acquisition times $\boldsymbol{t}_{\nu_{N_s}}^{(i)}$
8: Convergence: conv$= 1$, if the stopping criteria are satisfied.
9: **end while**

A. Fast and Robust Task Performance

Sensing aims to collect information regarding some phenomena and to perform a pre-defined task, such as depth estimation in LiDAR, cell detection in microscopy, target tracking in defence, etc. This step of the algorithm performs this pre-defined task by solving an inverse problem. In this work, we adopt a Bayesian approach (detailed in Section III) that leads to the estimation of M parameters of interest for the nth scanned position, denoted $\boldsymbol{\theta}_n = (\theta_{n,1}, \ldots, \theta_{n,M})$, together with a measure of their uncertainties $\boldsymbol{\epsilon}_n = (\epsilon_{n,1}, \ldots, \epsilon_{n,M})$, where both $\boldsymbol{\theta}_n$ and $\boldsymbol{\epsilon}_n$ are extracted from the estimated parameter posterior distributions. These estimates are only available on scanned points N_s, which represent a subset of the high-resolution grid composed of N pixels. To obtain a high-resolution ROI map, the obtained sub-sampled estimates should be spatially extended to cover more space in the HR grid. This leads to an inpainting problem, in which the parameters and uncertainties of an nth unavailable pixel (i.e., unscanned pixels or scanned pixels without photon detection) can be inferred from scanned neighbours pixels, as follows $\boldsymbol{\theta}_n = h_p(\boldsymbol{\theta}_{\psi_n})$, and $\boldsymbol{\epsilon}_n = h_u(\boldsymbol{\epsilon}_{\psi_n})$ where ψ_n represents the $3^{\text{Wind}} \times 3^{\text{Wind}}$ window of neighbours of the nth location, and h_p, h_u represent non-linear inpainting operators on the parameters and uncertainties, respectively. In this paper, we are interested in fast operators allowing efficient AS iterations, and thus adopt the median filter for both h_p, h_u in what follows. Other advanced optimization or learning based algorithms could be considered as in [27]–[29], which is beyond the scope of this paper. The resulting HR estimates and uncertainties will be used to build the region of interest (ROI) map, as described in the following section.

B. Regions of Interest

In this subsection the algorithm generates a probability map $\boldsymbol{m} \in \mathbb{R}^N$ (with $m_n \geq 0$ and $\sum_n m_n = 1$) of regions of interest, where regions with higher values will be scanned more frequently. This ROI is closely related to the targeted task and the nature of the parameters of interest. Based on an estimation task, the method in [20] aims to improve depth estimates and defined the ROI based on depth gradients, while the method in [31] introduced additional features based on the distance between scanned points. In a target detection scenario, where a target is defined as the presence of a reflective surface in the single-wavelength case or an object with a specific spectral signature in the multispectral case, the AS procedure would focus on scanning regions having a target and spending less resources on other regions in the scene. This paper assumes the presence of uncertainty measures $\boldsymbol{\epsilon}$ in addition to parameter estimates $\boldsymbol{\theta}$, and proposes to build a ROI probability map $\boldsymbol{m}$ such as $\boldsymbol{m} = h_r(\boldsymbol{\theta}, \boldsymbol{\epsilon})$, where h_r is a chosen nonlinear function. The latter formulation allows the combination of multiple tasks, and to take advantage of the available uncertainty measures. For example, if we are interested in reconstructing the depth profile of an object with a specific spectral signature. In this case, the parameters $\boldsymbol{\theta}$ could gather a depth estimate and a spatial label that classifies pixels based on their spectral signatures, and the ROI map should highlight pixels belonging to that object and give a particular interest to pixels with high depth uncertainty. Finally, it should be noted that pixels with no estimates after the inpainting process have the highest uncertainty and will receive a high probability for sampling, in contrast to pixels that have reached the maximum acquisition time and which would be excluded from the next sampling iteration.

C. Generation of New Locations and Acquisition Times

The proposed strategy is based on sampling N_s pixels at each adaptive sampling iteration. The locations of scanned points are subject to physical constraints due to the sampling scenario, i.e., pixel-wise scanning or array based sampling as shown in Fig. 2. Single-photon pixel-wise scanning has been widely used

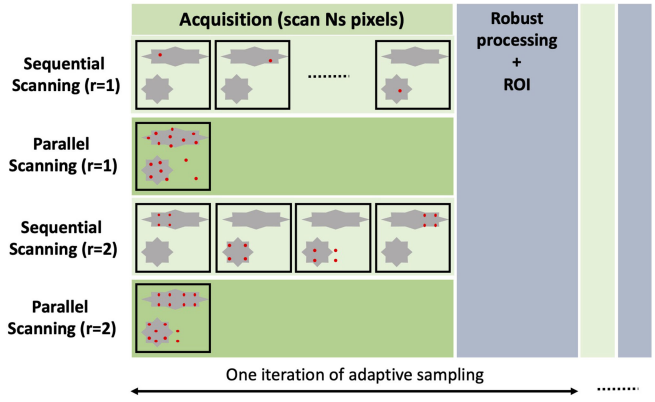


Fig. 2. Illustration of different acquisition scenarios where gray objects represent regions of interest. N_s pixels (red dots) are scanned in each iteration of adaptive sampling ($N_s = 16$ in this example) using sequential or parallel scanning, with different array sizes ($r = 1$ and $r = 2$). Some pixels are outside the ROI to explore all parts of the scene.

as it allows eye-safe imaging even at km-range distances [19], [32], [33]. With recent technological advances, array scanning is possible by moving a detector array spatially to scan a large field of view [18], [26], [34], [35]. In this paper, we study both scanning cases, and investigate the scenarios of sequential scanning (i.e., when scanning one pixel/array at a time) and parallel scanning when scanning all pixels in one shot (e.g., using a binary mask to select pixels or by activating electronically some pixels of an array of detectors, as illustrated in Fig. 2). Given the ROI probability map, we can select the next locations to scan (N_s/r^2 locations when scanning using $r \times r$ array, i.e., N_s locations in pixel-wise scanning) by selecting those with the highest probabilities. As stated in [20], this strategy prevents the exploration of the full scene which might lead to missing small or dynamic objects. Consequently, given the ROI probability map, the new locations will be randomly sampled according to the distribution $\boldsymbol{m}$ using the Metropolis-Hastings algorithm [36] with a uniform proposal distribution on the HR grid. It is worth noting that array scanning requires selecting N_s/r^2 array positions together with the associated array sizes, i.e., distance between points that mimic an array zoom-in or zoom-out. These array sizes are fixed using multi-scale information of the ROI map $\boldsymbol{m}$. Finally, it can be seen from Fig. 2 that the dwell time depends on the scenario and the acquisition time per shot. For example, consider a fixed acquisition time per step t_0, the sequential array scanning will require $N_s t_0/r^2$, while parallel scanning could be done in one shot requiring acquisition time t_0. Each new location (to be sampled from N_s or N_s/r^2) will be assigned an acquisition time ranging from t_0 to ct_0, where t_0 is a user defined acquisition time step and c is the importance level of a shot. During the iterative process, the defined acquisition time step is updated to ensure that the proportion of scanned pixels with photon detections is between $[0.7, 0.9]$. This will avoid working with too small time steps leading to low photon detections, or wasting resources by using unnecessary long time steps.

D. Stopping Criteria

Many stopping criteria can be considered for Algo. 1. The first criterion compares the two last depth parameter estimates and stops the algorithm if their root mean square error is smaller than a given user-defined threshold, that is: $\text{RMSE}(d^{(t+1)}, d^{(t)}) \leq \xi$, where $\text{RMSE}(d^{(t+1)}, d^{(t)})$ denotes the root mean square error between the depth estimates at the iteration $t+1$ and t. Other criteria can be considered such as reaching a pre-defined maximum acquisition time-per-pixel, a maximum number of scanned points and/or a maximum number of iterations.

III. Hierarchical Bayesian Model for Classification Strategy

The proposed adaptive sampling strategy is optimized for a predefined task. In this paper, we are interested in a target detection task based on a known object spectral signature. More precisely, we propose a spatial classification algorithm that labels pixels based on their spectral signatures, which then allows concentrating scanning samples on pixels of interest. Due to data sub-sampling during acquisition, the resulting histograms are often sparse and some pixels might be empty, leading to an underdetermined problem to perform the task. The latter is solved using a Bayesian approach that performs the classification task on multispectral LiDAR data. Adopting a Bayesian framework, the unknown parameters will be assigned prior distributions that will allow inclusion of additional information and regularization to the ill-posed problem. In addition to parameter estimates, this approach will also provide uncertainty measures regarding the estimated parameters, as required by the AS strategy. The next subsections introduce the likelihood summarizing data statistics, and the considered parameter prior distributions, which are combined to obtain the parameters joint posterior distribution. Note that the parameter's posterior distributions are evaluated in a pixel-wise fashion only on the scanned pixels, i.e., for $n \in \{1, \ldots, N_s\}$.

A. Likelihood

We consider a 3-dimensional cube of histograms $\boldsymbol{Y}$ of LiDAR photon counts of dimension $N \times L \times T$, where N, L and T are the number of pixels, spectral wavelengths and time bins, respectively. Let $\boldsymbol{Y}_n = [\boldsymbol{y}_{n,1}, \boldsymbol{y}_{n,2}, \ldots, \boldsymbol{y}_{n,L}]^T$ be an $L \times T$ matrix where $\boldsymbol{y}_{n,l} = [y_{n,l,1}, y_{n,l,2}, \ldots, y_{n,l,T}]^T$. According to [22], [37], each photon count $y_{n,l,t}$, where $n \in \{1, \ldots, N\}$, $l \in \{1, \ldots, L\}$ and $t \in \{1, \ldots, T\}$, is assumed to follow a Poisson distribution given by:

$$y_{n,l,t} | r_{n,l}, d_n, b_{n,l} \sim \mathcal{P}[r_{n,l}\, g_l(t - d_n) + b_{n,l}], \qquad (1)$$

where $\mathcal{P}(.)$ denotes a Poisson distribution, $r_{n,l} \geq 0$ is the reflectivity observed at the lth wavelength, $d_n \in \{1, 2, \ldots, T\}$ represents the position of an object surface at a given range from the sensor, $b_{n,l} \geq 0$ is the constant background level associated with dark counts and ambient illumination and $g_l(.)$ is the system impulse response function (IRF), whose shape can differ between wavelength channels (see Fig. 3), assumed to be known from a calibration step and normalized $\sum_{t=1}^{T} g_l(t) = 1$. An equivalent model can be considered as in [38] using the signal-to-background ratio (SBR), defined as the ratio of the useful detected photons $r_{n,l}$ and the total number of background photons in the histogram $b_{n,l}T$, i.e.: $w_{n,l} = \frac{r_{n,l}}{b_{n,l}T}$ with $w_{n,l} \geq 0$.

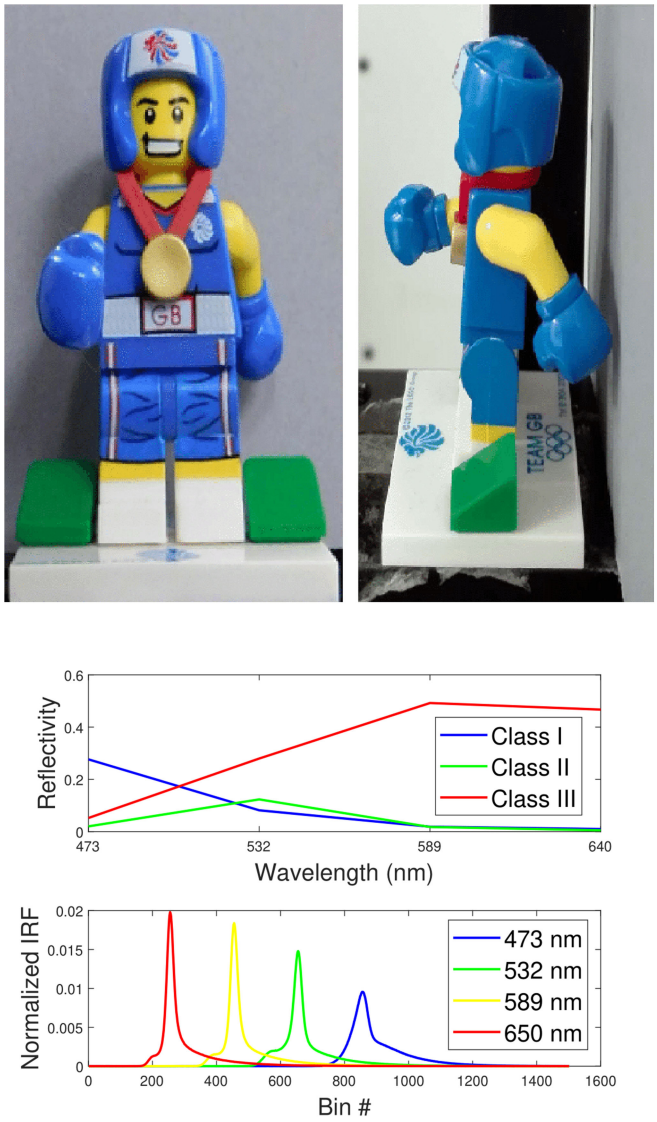


Fig. 3. (Top) The considered Lego scene used in section VI, (middle) spectral signatures of the spectral classes associated with Lego scene, (bottom) the normalized IRFs associated with the wavelengths 473, 532, 589 and 640 nm [19] used in Section VI.

Thus, (1) can be written in the following form

$$y_{n,l,t}|\omega_{n,l}, d_n, b_{n,l} \sim \mathcal{P}\{b_{n,l}\,[w_{n,l}\,T\,g_l(t-d_n)+1]\}. \quad (2)$$

This formulation allows an easy marginalization of the posterior distribution with respect to (w.r.t.) the background noise parameter as indicated in Section IV. The joint likelihood, when assuming the observed pixels, wavelengths and bins mutually independent, is then given by

$$p(\boldsymbol{Y}|\boldsymbol{\Omega}, \boldsymbol{d}, \boldsymbol{B}) = \prod_{l=1}^{L}\prod_{t=1}^{T} p(y_{n,l,t}|\omega_{n,l}, d_n, b_{n,l}) \quad (3)$$

where $\boldsymbol{d} = (d_1, \ldots, d_N)$ and $\boldsymbol{\Omega}$, $\boldsymbol{B}$ are two matrices gathering $\omega_{n,l}, \forall n, l$, and $b_{n,l}, \forall n, l$, respectively.

B. Prior Distributions

A LiDAR histogram can either result from background counts (in the absence of a target photons due to $r_n = \omega_n = 0$) or a mixture of target and background counts (when $r_n \geq 0$ or $\omega_n \geq 0$). Assuming the presence of K spectral signatures, the classification problem aims to associate a pixel with a target to one of the K spectral classes. The reflectivity prior accounts for this effect by considering a mixture of $K+1$ distributions as follows

$$P(r_{n,l}|u_n, \alpha^r_{k,l}, \beta^r_{k,l}, K) = \delta(u_n)\delta(r_{n,l}) + \sum_{k=1}^{K} \delta(u_n - k)\mathcal{G}(r_{n,l}; \alpha^r_{k,l}, \beta^r_{k,l}) \quad (4)$$

where $u_n \in \{0, 1, \ldots, K\}$ is a latent variable that indicates the absence of target if $u_n = 0$, otherwise, it indicates the label of the class, $\delta(.)$ is the Dirac delta distribution centred in 0, $\mathcal{G}(r_{n,l}; \alpha^r_{k,l}, \beta^r_{k,l})$ represents a gamma density whose shape and scale hyperparameters $(\alpha^r_{k,l}, \beta^r_{k,l})$ are fixed based on the K known spectral signatures. This prior is inspired by the spike-and-slab prior used in [38]. It accounts for $K+1$ cases, the first represents the absence of a target in the nth pixel and is obtained for $u_n = r_{n,l} = 0$, hence the use of a Dirac distribution (the spike part). The slab part accounts for the presence of one of the K signatures by using a gamma distribution. Thanks to the use of many wavelengths, this prior extends the object detection problem in [38] to a class detection problem using the spectral signature of each class. Considering the non-negativity of $b_{n,l}$, $\forall n, l$ and its continuous nature, the background level will be modelled with a gamma distribution as in [39]:

$$P(b_{n,l}|\alpha^b_l, \beta^b_l) = \mathcal{G}(b_{n,l}, \alpha^b_l, \beta^b_l) \quad (5)$$

where α^b_l and β^b_l are background hyper-parameters. Considering that only dozens of distinctive wavelengths will be used, we will consider the channels to be uncorrelated to keep the estimation strategy tractable. Since we are interested in using the model described in (2) instead of (1), assuming that the reflectivity and the background noise are independent and by applying a random variable change, the resulting joint prior distribution will yield:

$$\begin{aligned} p(\boldsymbol{\omega_n}, \boldsymbol{b_n}|u_n, \boldsymbol{\phi}, K) &= \prod_{l=1}^{L} p(\omega_{n,l}, b_{n,l}|u_n, \phi_l) \\ &= \prod_{l=1}^{L} \Big[\delta(u_n)\delta(\omega_{n,l})\mathcal{G}(b_{n,l}, \alpha^b_l, \beta^b_l) \\ &\quad + \sum_{k=1}^{K} \delta(u_n - k)C_{k,l}(\omega_{n,l}) \\ &\quad \times \mathcal{G}(b_{n,l}, \alpha^+_{l,k}, \beta^+_{l,k}(\omega_{n,l}))\Big] \end{aligned} \quad (6)$$

with

$$C_{k,l}(\omega_{n,l}) = \frac{(\beta^b_l)^{\alpha^b_l}(\beta^r_{k,l})^{\alpha^r_{k,l}} T^{\alpha^r_{k,l}}}{B(\alpha^r_{k,l}, \alpha^b_l)} \frac{\omega_{n,l}^{\alpha^r_{k,l}-1}}{\beta^+_{l,k}(\omega_{n,l})^{\alpha^+_{l,k}}}$$

$$\alpha_{l,k}^{+} = \alpha_l^b + \alpha_{k,l}^r$$

$$\beta_{l,k}^{+}(\omega_{n,l}) = \beta_l^b + \beta_{k,l}^r T\omega_{n,l}$$

where $B(.)$ is the beta function and $\boldsymbol{\phi} = (\phi_1, \phi_2, \ldots, \phi_l)$ with $\phi_l = (\alpha_{k,l}^r, \beta_{k,l}^r, \alpha_l^b, \beta_l^b)$, $k \in \{0, \ldots, K\}$.

As we assume no prior knowledge about a pixel's class, the parameter u_n is assumed to be drawn from a uniform distribution, i.e.: $p(u_n = k) = \frac{1}{K+1}$, where $k \in \{0, \ldots, K\}$. However, This non-informative class prior can be changed in the presence of additional information regarding the classes. The depth parameter d_n is assigned a non-informative uniform prior as follows:

$$p(d_n = t) = \frac{1}{T} \quad , \forall t \in \{1, \ldots, T\}. \tag{7}$$

Nonetheless, this can be modified in case of additional information regarding the target position.

C. Joint Posterior Distribution

From the joint likelihood in (3) and the prior distributions specified in Section III-B, we can obtain the joint posterior distribution for $\boldsymbol{\omega_n}$, $\boldsymbol{b_n}$, d_n and u_n given the 3D histograms $\boldsymbol{Y_n}$ and the hyperparameters $\boldsymbol{\phi}$ and K. Using Bayes rule and assuming that d_n and u_n are independent from $\boldsymbol{\omega_n}$ and $\boldsymbol{b_n}$, the joint posterior distribution of the proposed Bayesian model can be formulated as follows:

$$p(\boldsymbol{\Theta_n}|\boldsymbol{Y_n}, \boldsymbol{\phi}, K) \propto p(\boldsymbol{Y_n}|\boldsymbol{\Theta_n}) p(\boldsymbol{\Theta_n}|\boldsymbol{\phi}, K) \tag{8}$$

where

$$\begin{aligned} \boldsymbol{\Theta_n} &= (\boldsymbol{\omega_n}, \boldsymbol{b_n}, d_n, u_n) \\ p(\boldsymbol{\Theta_n}|\boldsymbol{\phi}, K) &= p(\boldsymbol{\omega_n}, \boldsymbol{b_n}, d_n, u_n|\boldsymbol{\phi}, K) \\ &= p(\boldsymbol{\omega_n}, \boldsymbol{b_n}|\boldsymbol{\phi}, u_n, K) p(d_n) p(u_n). \end{aligned} \tag{9}$$

IV. Estimation Strategy

The posterior distribution in (8) reflects our knowledge of the unknown parameters to be estimated given the photon data and the available prior knowledge. In this paper, we are interested on estimating the depth and label parameters, i.e., $M = 2$ and $\boldsymbol{\theta} = (\boldsymbol{d}, \boldsymbol{u})$. The Bayesian estimator to be considered, both for the depth and class parameter, is the maximum a posteriori (MAP) estimator as in [13], [40]. From (8), we marginalize the background noise and signal-to-background parameters to get the joint depth and class marginal probability as follows:

$$p(u_n, d_n|\boldsymbol{Y_n}) = \int\int p(\boldsymbol{\omega_n}, \boldsymbol{b_n}, d_n, u_n|\boldsymbol{Y_n}) d\boldsymbol{b_n}\, d\boldsymbol{\omega_n}. \tag{10}$$

A. Class Estimation

The decision rule adopted to determine the pixel label is

$$H_n = \underset{k \in \{1,\ldots,K\}}{\operatorname{argmax}} \; p(u_n = k \,|\, \boldsymbol{Y_n}) \tag{11}$$

with

$$p(u_n|\boldsymbol{Y_n}) = \sum_{d_n=1}^{T} \int\int p(\boldsymbol{\omega_n}, \boldsymbol{b_n}, d_n, u_n|\boldsymbol{Y_n}) d\boldsymbol{b_n}\, d\boldsymbol{\omega_n} \tag{12}$$

where H_n represents the class of the nth pixel. Note that for $K = 1$ and $L = 1$, we end up with a target detection decision rule as in [38]. We demonstrate that the marginal probability $p(u_n|y_n)$ is :

$$\begin{aligned} p(u_n = 0|\boldsymbol{Y_n}) &= \prod_{l=1}^{L} p(u_n = 0|\boldsymbol{y_{n,l}}) \\ &= \prod_{l=1}^{L} \frac{p(u_n = 0)\Gamma(\bar{y}_{n,l} + \alpha_l^b)}{(T + \beta_l^b)^{(\bar{y}_{n,l} + \alpha_l^b)}\, \gamma_l} \end{aligned} \tag{13}$$

$$\begin{aligned} p(u_n = k|\boldsymbol{Y_n}) = \sum_{d_n=1}^{T} \int_0^{\infty} \prod_{l=1}^{L} [&p(u_n = k) p(d_n) D_{n,l,k} \gamma_l^{-1} \\ &F_{n,l,k}(\omega_{n,l}, d_n) d\omega_{n,l}] \end{aligned} \tag{14}$$

with

$$\gamma_l = \frac{\Gamma(\alpha_l^b)}{(\beta_l^b)^{\alpha_l^b}} \prod_{t=1}^{T} y_{n,l,t}!$$

$$D_{n,l,k} = \frac{\Gamma(\bar{y}_{n,l} + \alpha_l^b + \alpha_{k,l}^r)(T\beta_{k,l}^r)^{\alpha_{k,l}^r}}{\Gamma(\alpha_{k,l}^r)}$$

$$\begin{aligned} &F_{n,l,k}(\omega_{n,l}, d_n) \\ &\quad = \frac{\exp\{\sum_{t=1}^{T} y_{n,l,t} \ln[\omega_{n,l} T g_l(t - d_n) + 1]\}}{\omega_{n,l}^{1-\alpha_{k,l}^r} \{\beta_l^b + [T\,(1 + \omega_{n,l}(1 + \beta_{k,l}^r))]\}^{\alpha_{l,k}^{+} + \bar{y}_{n,l}}} \end{aligned} \tag{15}$$

where $\bar{y}_{n,l} = \sum_{t=1}^{T} y_{n,l,t}$. In the event of no target, we can see that the integral is available in its analytical form thanks to the conjugacy between the model (2) and the priors (6). The marginal distribution in (14) is, however, intractable in presence of a target. One way to simplify it is to consider that the depth captured is different across all the spectral wavelengths. This simplification improves the tractability of the marginal class probability and will transform (14) into (16) as follows:

$$\begin{aligned} p(u_n = k|\boldsymbol{Y_n}) = \prod_{l=1}^{L} \sum_{d_{n,l}=1}^{T} \Big[&p(u_n = k) p(d_{n,l}) D_{n,l,k} \gamma_l^{-1} \\ &\int_0^{\infty} F_{n,l,k}(\omega_{n,l}, d_{n,l}) d\omega_{n,l}\Big]. \end{aligned} \tag{16}$$

The resulting integral with respect to $\omega_{n,l}$ in (16) can be numerically approximated with a quadrature method. The matched filter in (15) can be computed with $\mathcal{O}(T \log T)$ using the fast Fourier transform (FFT) leading to an overall complexity of the integral per-pixel in (16) given by $\mathcal{O}(KLJT \log T)$, where K is the number of classes considered, L is the number of wavelengths, J is the computational cost of the evaluated integrand and T is the number of the temporal bins.

B. Depth Estimation

The depth estimate can be obtain as follows

$$\hat{d}_n = \underset{d \in \{1,\ldots,T\}}{\operatorname{argmax}} \; p(d_n \,|\, \boldsymbol{Y_n}) \tag{17}$$

where

$$p(d_n|\boldsymbol{Y_n}) = \sum_{k=1}^{K}\prod_{l=1}^{L}\left[p(u_n=k)p(d_n)D_{n,l,k}\gamma_l^{-1}\right.$$
$$\left.\int_0^{\infty} F_{n,l,k}(\omega_{n,l}, d_n)d\omega_{n,l}\right]. \quad (18)$$

Although ω can be integrated out from (18), this might lead to a high computational cost. In this paper, we choose to estimate the depth given the easily computed marginal map estimate $\omega_{n,l}^{\text{map}}$ using the simplified model introduced in Section IV-A, leading to

$$\hat{d}_n = \underset{d\in\{1,\ldots,T\}}{\text{argmax}}\ p(d_n \mid \boldsymbol{Y_n}, \boldsymbol{\omega}_n^{\text{map}}). \quad (19)$$

The proposed approach allows for the evaluation of the full marginal depth posterior. In addition to the depth point estimate, this distribution will allow uncertainty quantification (i.e., to quantify our confidence regarding the estimates). In this paper, we evaluate the depth uncertainty by considering the depth negative log-cumulative marginal posterior around the MAP estimate, i.e., NCD $= -\log[\sum_{\hat{d}_n-\epsilon}^{\hat{d}_n+\epsilon} p(d_n \mid \boldsymbol{Y_n}, \boldsymbol{\omega}_n^{\text{map}})]$, where ϵ is a user fixed constant. Note that a small NCD indicates a high confidence about the estimate, while a large one would be an indication of low confidence.

V. Results on Simulated Data

This section evaluates the performance of the proposed adaptive sampling framework and classification algorithm on simulated data. We first evaluate the performance of the Bayesian classification algorithm using simulated multispectral data with $L = 4$ wavelengths when varying the signal-to-background ratio (SBR) and the average signal photons per pixel. Then, we compare different adaptive sampling scenarios based on sequential and parallel scanning modes for different array sizes. Finally, we compare the proposed adaptive sampling strategy with other static and dynamic sampling schemes from the literature. Except for network based algorithms which were trained on a server equipped with an RTX 3090 GPU and 128 GB RAM, all other simulations were performed using Matlab 2020a on an Intel Core i7-8700@3.2 GHz, 16 GB RAM and results are averaged based on three Monte Carlo realisations.

A. Comparison Algorithms and Evaluation Criteria

The proposed approach (denoted AS) is compared against various sampling strategies, namely:

1) Static sampling strategies: Including uniform (US) and random (RS) sampling strategies that only scan a ratio of the N pixels (we considered 30% and 60% of pixels for random sampling). The maximum likelihood estimate (MLE) in the absence of background counts is considered to estimate the depth parameter [33], which reduces to the maximum of the correlation between the log-IRF and each histogram (denoted by Xcorr for cross-correlation). For multispectral data, the depth estimate reduces to maximizing the sum of the cross-correlations between each log-IRF and its corresponding histogram channel-wise, which is the maximum likelihood estimator in absence of background counts.
2) Fast adaptive sampling (FAS) strategy [20]: This algorithm adopts the same framework as in Section II while considering a depth estimation task. Depth is estimated from histograms using a cross-correlation with a Gaussian IRF (i.e., approximate MLE) leading to fast analytical computations. The ROI map is generated based on depth variations over pixels (i.e., depth's gradient).
3) Learning-based Adaptive Sampling (LAS) strategy using Ensemble Variance [30]: This is an iterative algorithm to perform adaptive sampling. Within each iteration (phase), few depth points are initially scanned, then depth completion is performed using a deep learning algorithm that estimates an ensemble M of depth maps to approximate depth uncertainty. The latter is then used to define the next informative pixels to sample. Note that the proposed Bayesian algorithm is applied to the histogram data to obtain the depth estimates required by this algorithm. In this paper, the LAS algorithm will not use an RGB guide, as it is not always available for real data. This algorithm was trained on a server equipped with an RTX 3090 GPU using the KITTI data suggested by the authors and described in [30] with the following parameters: 4 phases, $M = 5$, batch size of 4, an 18-layer architecture, and a maximum of 7 training epochs.

It is worth noting that depth completion (i.e., filling unscanned pixels) is performed using a median operator h_p when considering US, RS, FAS and the proposed AS approach, while LAS used a sparse-to-depth approach developed in [27].

To evaluate the performance of the proposed algorithm against the approaches stated above, we use the root mean square error (RMSE) and the accuracy to evaluate the depth estimation and class estimation, respectively. The RMSE is defined as RMSE $= \sqrt{\frac{1}{N}||\boldsymbol{d}^{\text{ref}} - \widehat{\boldsymbol{d}}||^2}$, where $\boldsymbol{d}^{\text{ref}}$ is obtained from sampling the whole scene with the maximum acquisition time and under a negligible background illumination (see the second column of Fig. 4). The accuracy in percentage is defined as ACC $= \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}}$, where TP, TN, FP and FN represent: true positive, true negative, false positive and false negative probabilities, respectively.

B. Datasets

One simulated scene and two real single-photon data sets have been used to assess the performance of the proposed algorithm. The first synthetic scene is the cluttered Art scene Fig. 4 (top-left image) extracted from the Middlebury dataset.[1] This scene is commonly used as a standard scene for algorithms evaluation in many LiDAR imaging experiments [15], [22]. It contains $N = 123 \times 155$ depth and grayscale images that are used to simulate histograms of photon counts according to the model in (1), when considering a real impulse response function, $T = 164$ bins and a 16 ps time bin resolution. Note that the scene is truncated in depth to simulate pixels without target returns as represented in

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

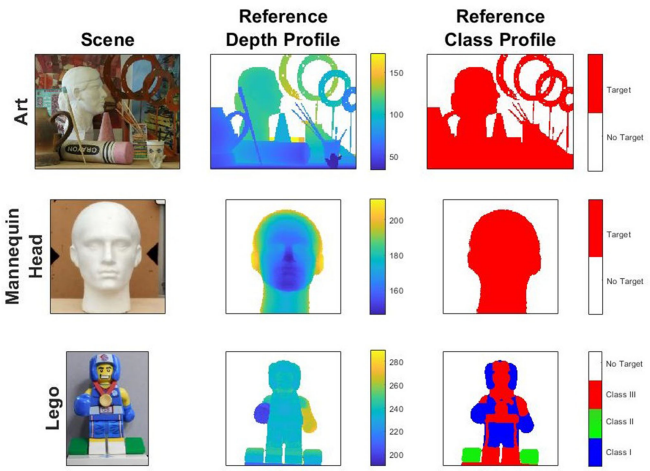


Fig. 4. From top to bottom: The Art scene, the Mannequin head and the Lego scene. From left to right column: the scene of interest, reference depth estimation (in time bins) and class detection map with maximum acquisition time per-pixel in absence of illumination.

Fig. 4. The second scene is the life-sized mannequin head shown in Fig. 4 (middle-left image). This data was acquired at a standoff distance of 40 m using a time-of-flight scanning sensor, based on a time-correlated single-photon counting system (TCSPC) (the reader is referred to [32], [33] for more information regarding the transceiver system and data acquisition hardware used for this work). The spatial, spectral and temporal dimensions for the mannequin head are $N = 142 \times 142$ pixels, $L = 1$ wavelength and $T = 191$ bins (bin width of 16 ps). Each pixel was acquired for 30 ms acquisition per-pixel in a dry and clear sky environmental conditions leading to an SBR around $\omega = 70$. In all simulations using the Art scene and the mannequin head data, we consider $\boldsymbol{\phi}_l = \boldsymbol{\phi} = (\alpha^r, \beta^r, \alpha^b, \beta^b) = (2, \frac{2}{r^M}, 1, \frac{T}{r^M})$ with r^M being the average number of signal photons per pixel. These hyper-parameter values are relatively non-informative for both the reflectively and the background noise parameters.

The third scene is the Lego data depicted in Fig. 3 and bottom-left of Fig. 4 (the reader is referred to [19] for more detail). The object, of size 42 mm tall and 30 mm wide, was scanned at a standoff distance of 1.8 m using a TCSPC module for an acquisition time per-pixel of 160 ms (40 ms acquisition time per pixel using four wavelengths where the system IRFs $g_l(.), \forall l$ are shown in Fig. 3-bottom). The size of the spatial, spectral and temporal dimension of the single-photon Lego data are, respectively, $N = 200 \times 200$ pixels, $L = 4$ wavelengths and $T = 1500$ bins with a timing bin size of 2 ps. Two versions of the Lego data are used in the experimental section. The first version was acquired in absence of background illumination (SBR $= \omega = 66$) and the second one was acquired with presence of ambient illumination leading to an SBR of $\omega = 1.3$. In all simulations, the Lego exhibits three classes of interest ($K = 3$) whose spectral signatures (related to $\boldsymbol{\alpha^r}$ and $\boldsymbol{\beta^r}$) are extracted from pixels acquired considering a negligible background contribution and after maximum acquisition time per-pixel (see signatures in Fig. 3). α_l^b and β_l^b are relatively non-informative such that $(\alpha_l^b, \beta_l^b) = (1, \frac{T}{r_l^M})$ with r_l^M being the average number of signal photons per pixel for the l th spectral wavelength.

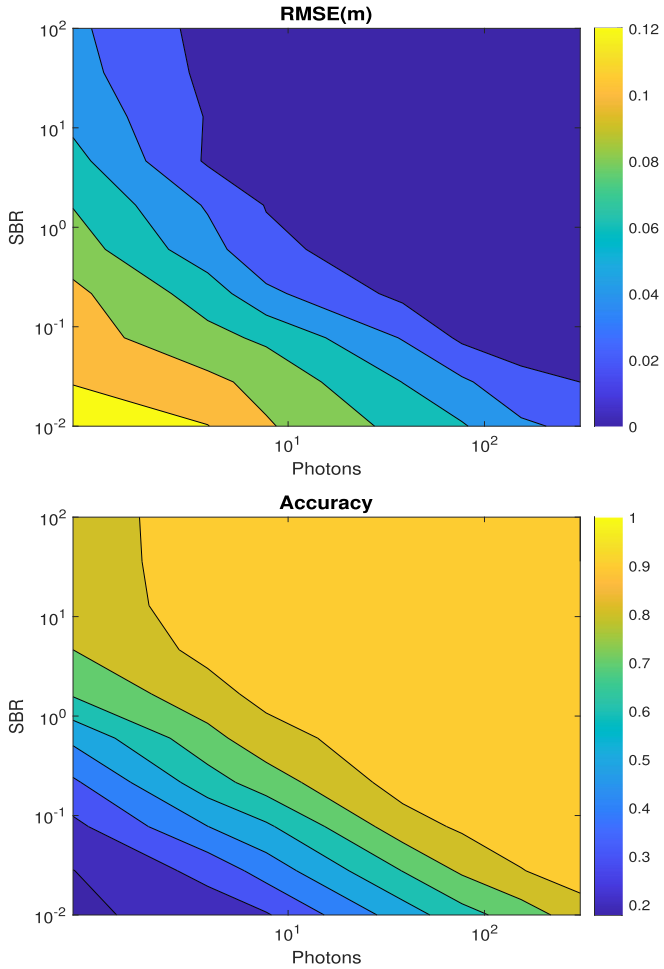
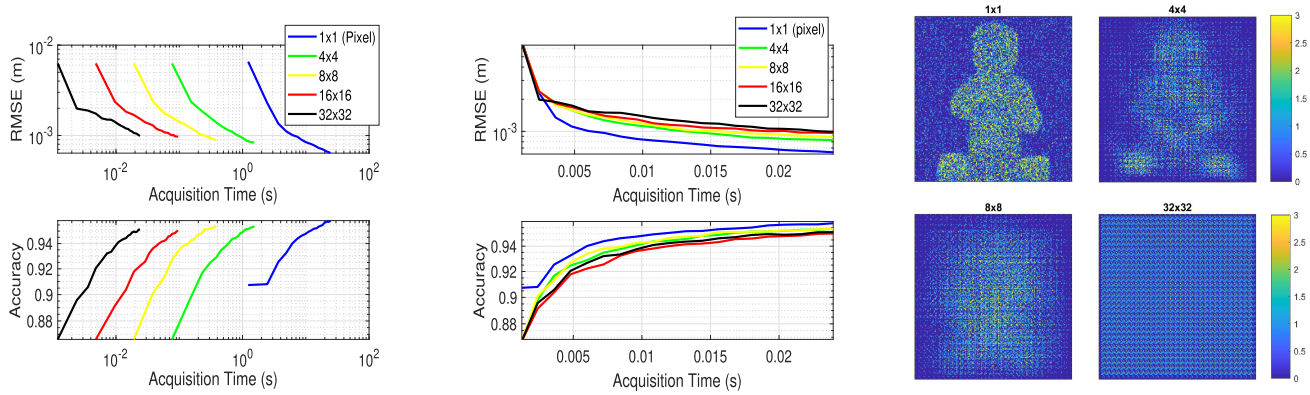


Fig. 5. (Top) Depth RMSEs and (bottom) Classification accuracy of the sub-sampled Lego data w.r.t. signal-to-background ratio (SBR) and the average signal photons.

C. Evaluation of the Classification Algorithm

In this section, we will evaluate the performance of the classification algorithm (according to metrics described in Section V-A) by simulating data based on the real multispectral single photon Lego data described in Section V-B. We consider a spatially sub-sampled data to analyse the behavior of the algorithm w.r.t. SBR and photons levels. The subsampled data has $N = 40 \times 40$ pixels, $L = 4$ wavelengths and $T = 1500$ time bins (bin width of 2 ps), and is corrupted so that the SBR varies from 0.01 to 100. Fig. 5 represents the RMSE in meters (top) and the class accuracy (down) w.r.t SBR and the average signal photons. These two figures provide the user with the required number of useful photons (which is proportional to the scanning time) needed to have a given depth precision and accuracy for different SBR levels. Table I depicts the confusion matrix for an SBR of magnitude $\omega = 0.6$ and after 5 ms of acquisition time per-pixel (which corresponds to signal photons per-pixel ≈ 42). In this table, the bold values exhibit the number of pixels of each predicted class, and bellow them their percentages w.r.t the total number of pixels (here 40^2 pixels). The last horizontal and vertical lines of this table represent the precision and the



(a) Sequential acquisition (b) Parallel acquisition (c) Sampling pattern

Fig. 6. Depth RMSEs and classification accuracy w.r.t. acquisition times for different SPAD array resolution when the acquisition mode is done (a) sequentially, or (b) in parallel. (c) Depicts the sampling pattern for different array sizes.

TABLE I
CONFUSION MATRIX OF THE SUB-SAMPLED LEGO DATA (40×40 PIXELS) FOR SBR $= 0.6$ AND 5 MILLISECONDS OF DWELL TIME PER-PIXEL AND PER WAVELENGTH (SIGNAL AVERAGE PHOTONS PER-PIXEL ≈ 42)

	Predicted classes				Recall
True classes for $\omega = 0.6$	**223** 13.9%	**9** 0.6%	**6** 0.38%	**1** 0.06%	93.3%
	0 0%	**63** 3.94%	**1** 0.06%	**2** 0.13%	95.5%
	7 0.44%	**7** 0.44%	**238** 14.88%	**6** 0.4%	92.3%
	0 0%	**10** 0.63%	**0** 0%	**1027** 64.19%	99%
Precision	97.4%	98.1%	83.1%	82.3%	96.9%

recall, respectively. This table highlights a good classification accuracy for these SBR and photon levels. Considering previous parameters and for maximum acquisition time, the average computational time per-pixel of the proposed classification algorithm is ≈ 55 ms, while it should be noted that per-pixel operations are independent allowing parallel processing.

D. Pixel and Array Based AS

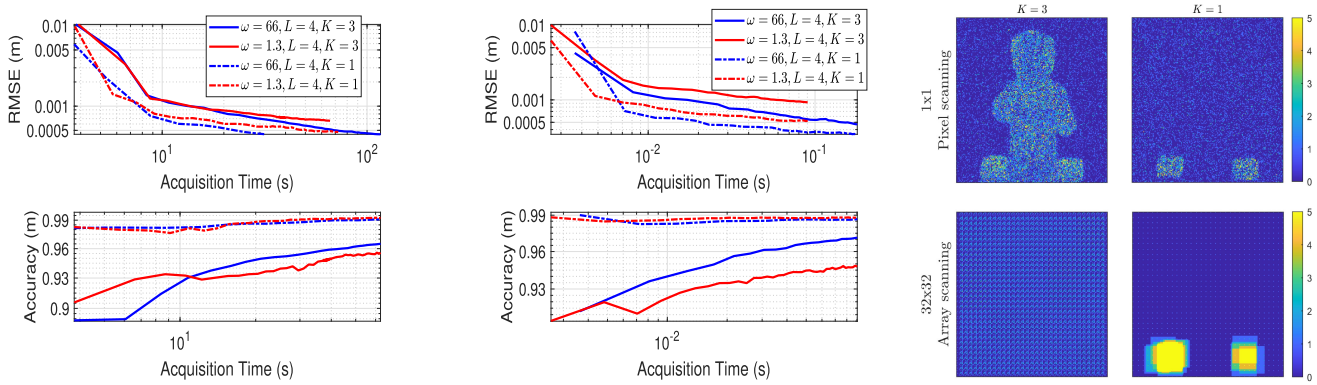
This section analyses the performance of the proposed AS strategy when considering different scanning scenarios as indicated in Fig. 2. More precisely, we study sequential and parallel sampling, when considering different array sizes. For all cases, we evaluate the depth and class accuracy with respect to acquisition time. As indicated in Section II-C, the acquisition time per one iteration of AS will depend on the considered sampling scenario (e.g., equal to $N_s t_0 / r^2$ for sequential scanning and to t_0 for parallel scanning, when assuming t_0 dwell time per shot and $r \times r$ array). The parameters considered for this section are $N_s = 32^2$, $t_0 = 300\mu s$.

First, we fix SBR$=$ 66 and study the effect of varying array sizes as follows $r \in \{1, 4, 8, 16, 32\}$. Fig. 6(a) and (b) shows the variation of depth RMSE and classification accuracy w.r.t. acquisition time. As expected, these figures show faster convergence for parallel acquisition reaching millisecond levels as opposed to seconds in sequential scanning. Considering sequential scanning, Fig. 6(a) highlights faster convergence for larger array systems as they acquire r^2 samples in parallel. For parallel scanning, large arrays impose spatial constraints on the sampled locations, hence, better performance is obtained by small arrays as they better approximate the scene features. This effect is highlighted in Fig. 6(c) showing the sampled points, where smaller arrays allow fine scene scanning by locating more points on the Lego shape (i.e., dense and focused scanning pattern) than larger arrays.

Second, we evaluate the effect of SBR and the size of the object of interest within the field of view, when considering a sequential acquisition and $r \in \{1, 32\}$. To change the target size, we compare a large ROI obtained when targeting the full Lego (i.e., the $K = 3$ classes are the target) and the case where the target of interest are the green pixels in Fig. 3 (i.e., the target is only one class $K = 1$). Fig. 7(a) and (b) show depth and accuracy performance for different SBR$\in \{1.3, 66\}$ and object sizes. As expected, the proposed AS approach shows faster convergence for smaller regions of interest (i.e., dashed lines better than continuous lines), and/or for larger SBR values (i.e., blue curves are similar or slightly better than red ones). Fig. 7(c) shows the sampled patterns for the different cases, indicating finer scanning for $r = 1$ compared to coarse results when considering $r = 32$.

E. Comparison With Static and Dynamic Sampling Algorithms

This section evaluates the performance of the proposed framework on the simulated Art scene when considering static (US, RS) and dynamic (FAS, LAS) sampling strategies. In both cases, performance is compared by evaluating depth RMSE w.r.t. average scanned photons. For this section, we consider a sequential scanning system with $r = 1$ and the sampling parameters $N_s = 475$ pixels, and $t_0 = 900\mu s$. Fig. 8 (left) shows results when comparing AS to static sampling strategies for high and low SBR. It highlights the benefit of the proposed algorithm as it converges faster than other approaches regardless of the scenario considered. While the factor of improvement is around two in absence of illumination, the potential of the proposed approach appears when the level of background is high. For example, at SBR$=$ 0.79 and for an RMSE$=$ $0.02\tilde{}m$, the proposed AS algorithm outperforms static sampling strategies by a factor of



(a) 1×1 pixel scanning (b) 32×32 array scanning (c) Scanning Pattern

Fig. 7. Depth RMSEs and classification accuracy w.r.t. acquisition times for two SBR levels for multispectral target detection ($K = 1$) and spectral classification ($K = 3$) using a (a) 1×1 pixel scanning system, (b) 32×32 array scanning system. (c) Depicts the sampling pattern obtained for $r \in \{1, 32\}$ when considering the full Lego as a target (i.e., $K = 3$) or only the green region (i.e., $K = 1$). Similar patterns are observed for high or low SBR.

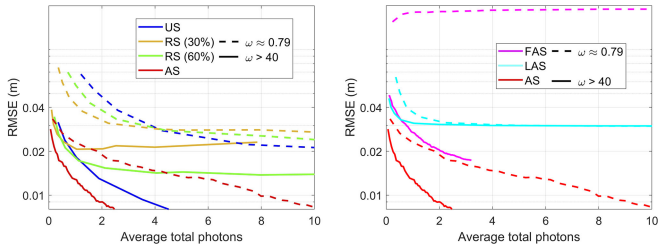


Fig. 8. Comparison of depth RMSEs using the proposed approach w.r.t. average total photons against: (left) static sampling strategies, (right) Dynamic sampling strategies. Solid and dashed lines represent: SBR > 40 (absence of illumination) and SBR $= 0.79$, respectively. Results are obtained with 3 Monte-Carlo realizations.

≈ 5, highlighting the efficiency of the proposed statistical model and sampling strategy. Fig. 8 (right) compares the proposed AS to dynamic sampling algorithms. This figure highlights the robustness of the AS approach when compared to FAS which assumes the absence of background counts in the data. The LAS is more robust than FAS, however, it shows limited performance and seems to converge to a relatively high constant RMSE level. This could be due to differences between the training and test data. This is illustrated in Fig. 12 (top-row) which shows the obtained depth maps at low SBR when considering the different sampling strategies. This figure confirms best results with the proposed AS algorithm and highlights the sensitivity of FAS to background counts as it shows the worst depth map.

VI. Results on Real Data

This section evaluates the proposed strategy on two experimental data sets, namely the monochromatic mannequin head and the multispectral Lego scene. In both cases, we compare the proposed adaptive sampling strategy against static and dynamic sampling strategies and highlight its benefits for range estimation, both in absence and presence of ambient illumination cases. Finally, the system that will be used is a pixel-wise SPAD system ($r = 1$) and the acquisition mode is sequential.

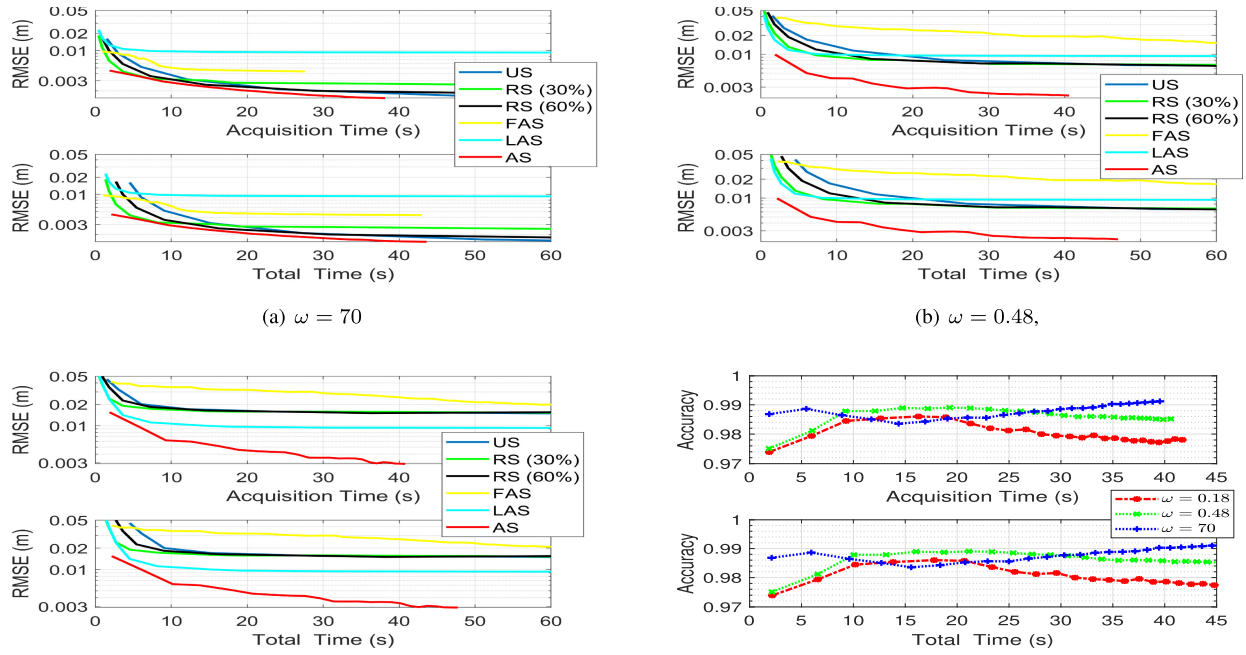
TABLE II
Confusion Matrix of the Lego Scene (200×200 Pixels) in (Top) $\omega = 66$ and 3 Seconds of Acquisition Time Per Wavelength (≈ 26 Signal Photons Per-Pixel) and (Bottom) $\omega = 1.3$ and 2.55 Seconds of Acquisition Time Per-Pixel and Per Wavelength (≈ 22 Signal Photons Per-Pixel and Per Wavelength))

	Predicted classes				Recall
True classes for $\omega = 66$	**4950** 12.38%	**45** 0.11%	**679** 1.7%	**102** 0.26%	85.7%
	22 0.06%	**1548** 3.87%	**75** 0.19%	**23** 0.06%	92.8%
	557 1.39%	**49** 0.12%	**6286** 15.72%	**45** 0.11%	90.6%
	332 0.83%	**352** 0.88%	**264** 0.66%	**2476** 61.68%	72.3%
Precision	84.5%	77.6%	86.1%	99.3%	93.9%

True classes for $\omega = 1.3$	**5022** 12.56%	**29** 0.07%	**594** 1.49%	**131** 0.33%	87%
	24 0.06%	**1530** 3.82%	**93** 0.23%	**21** 0.05%	91.7%
	616 1.54%	**48** 0.12%	**6231** 15.58%	**42** 0.11%	89.8%
	262 0.66%	**395** 0.99%	**310** 0.78%	**24652** 61.63%	96.2%
Precision	84.8%	76.4%	86.2%	99.2%	93.6%

A. Evaluation of AS on the Mannequin Head

This subsection compares the proposed AS approach to static and dynamic sampling strategies when considering a real monochromatic dataset. We consider the mannequin head scene and focus on the particular case where the number of wavelength and classes of interest are respectively $L = 1$ and $K = 1$. This reduces the classification algorithm to a target detection algorithm. The proposed algorithm estimates the parameters of interest (depth and class) in a pixel-wise fashion, thus we report pixel-wise values as the processing can be parallelized. We compare for two SBR levels, a high SBR$= 70$ with almost no background light in Fig. 9(a), and a noisy case with SBR $= 0.48$ and SBR $= 0.18$ in Figs. 9(b) and (c), respectively. The accuracy of all SBR levels are reported in Fig. 9(d). In any of the above figures,



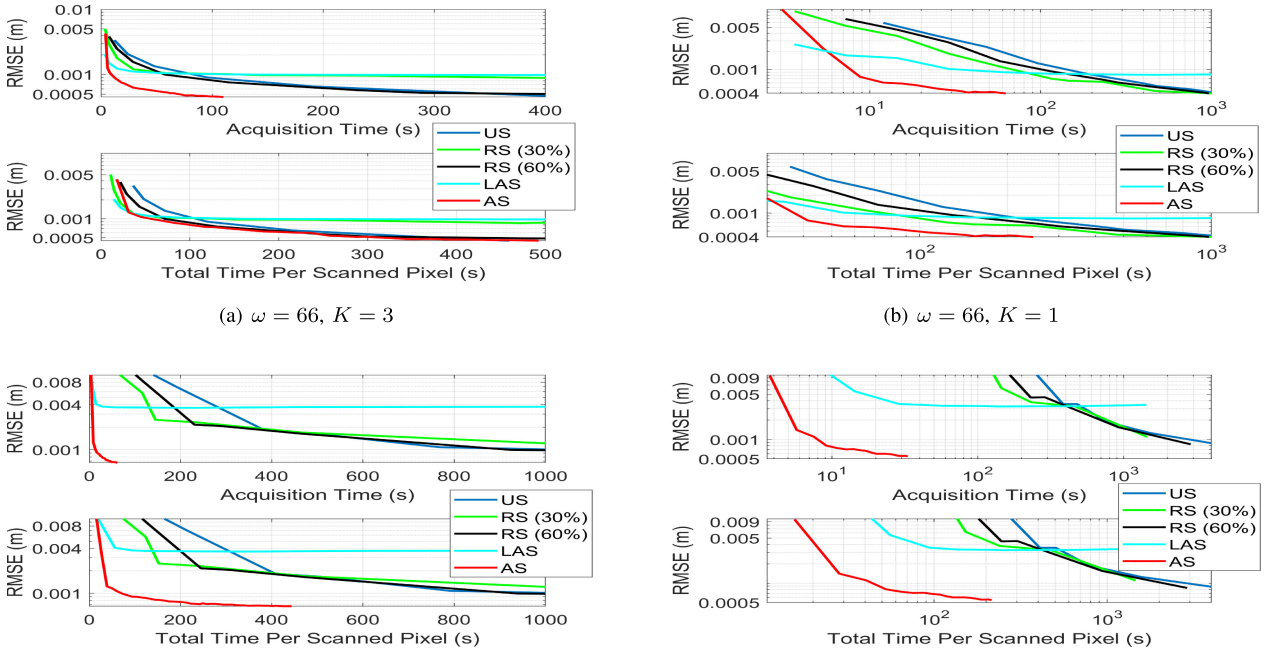
(a) $\omega = 70$

(b) $\omega = 0.48$,

(c) $\omega = 0.18$

(d) Target accuracy for different ω

Fig. 9. Depth RMSEs of the full mannequin head target for different sampling strategies w.r.t. (top) dwell time and (bottom) total time when (a) $\omega = 70$, (b) $\omega = 0.48$, (c) $\omega = 0.18$. (d) Target accuracy of the full mannequin head target for different sampling strategies w.r.t. (top) dwell time and (bottom) total time for three SBR levels.



(a) $\omega = 66$, $K = 3$

(b) $\omega = 66$, $K = 1$

(c) $\omega = 1.3$, $K = 3$

(d) $\omega = 1.3$, $K = 1$

Fig. 10. Results of the proposed algorithm on the Lego data w.r.t acquisition time (dwell time) and total time compared to static sampling strategies for number of classes and two SBR levels targeting (a) 3 classes and (b) 1 class in the absence of background illumination, (c) 3 classes and (d) 1 class under the presence of background illumination.

the top subplots evaluate performance w.r.t. the acquisition time (dwell time) used to scan the data. We also evaluate performance while accounting for the total time of the AS process which includes: the dwell time, the per-pixel processing time to perform the task, the time to build the ROI map and the time to move the scanning mirrors (one move is approximated by 150μ s). Under high SBR, Figs. 9(a) show that depth performance are similar for the different sampling strategies, although the proposed algorithm provides additional classification information. The benefit of the proposed framework becomes clear in the noisy case,

where an improvement factor of $\simeq 100$ is observed compared to other sampling strategies using Xcorr (at depth RMSE equal to 15 mm in Fig. 9(c)). Note that the FAS algorithm is sensitive to background and performs poorly in noisy cases. Note also that LAS converges to approximately the same RMSE value for different noise levels, which indicates its robustness to noise. Fig. 9(d) shows a target accuracy higher than 97% at the very beginning of the sampling process regardless of the SBR level indicating the robust restoration strategy performed by h_p, h_u using the median operator. Finally, Fig. 12 (middle-row) shows the mannequin depth maps obtained with the different sampling strategies indicating best performance with the proposed AS framework.

B. Evaluation of AS on the Multispectral Lego Scene

This subsection evaluates the proposed strategy on the full multispectral Lego data described in Section V-B (see Fig. 3 (top)). The approach is compared to different sampling strategies while considering two tasks: (i) classification based on spectral signature, (ii) signature based object detection.

1) AS for MS Classification: The first study consider a classification task, where the sensing aims to detect pixels with target, and to classify them into $K = 3$ classes based on known spectral signatures (see Fig. 3). Fig. 10(a) and (c) shows depth RMSEs for negligible ($\omega = 66$) and high background illumination ($\omega = 1.3$), respectively. In this case, the region of interest is large and occupies most of the HR image, which reduces the benefit of the proposed targeting approach. Nonetheless, Fig. 10(a) and (c) show a noticeable improvement by the proposed AS strategy especially for low SBR (e.g., In Fig. 10(c) and for RMSE $= 1.25$ mm, we observe an improvement factor of 80 and 18 based on dwell or total times, respectively). Table II shows the classification confusion matrix for the considered Lego scene for an average acquisition time of $75\mu s$ per-pixel and per wavelength (less than 3 seconds of total acquisition time per wavelength). This table highlights good classification results with an accuracy ≥ 0.9 even at low SBR. Finally, Fig. 11 illustrates the visual performance of the algorithm on the Lego data ($\omega = 1.3$) with respect to iterations. This figure indicates that most samples are located in the Lego region as promoted by the ROI map, and that both depth and classification maps improve with iterations.

2) AS for Signature Based Object Detection: Being a scene dependent approach, adaptive sampling is sensitive to the structure and the distribution of objects of interest in a particular scene. So far, the objects of interest that we tested occupy an important portion of scene (where the ROI map occupies more than 50% of the HR grid). To simulate a real world scenario, where targets are usually occupying a smaller portion of the scene, the algorithm is run using the samples Lego data, but targeting a single class corresponding to the green blocks on both sides of the Lego (see top sub-figures of Fig. 3). In addition to their small size, the green blocks have low reflectivity, hence present high uncertainty. Thanks to the ROI map, the algorithm will be able to scan this region and reduce the uncertainty measure by scanning more around that region. This task is

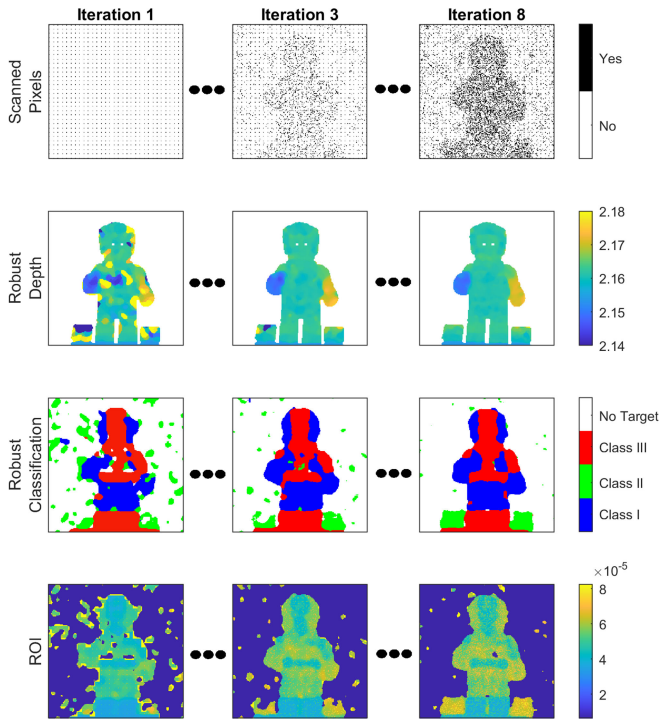


Fig. 11. Illustrative results of the AS algorithm with respect to iterations. The columns represent the process at a particular iteration and the rows represent from top to bottom: scanned samples, robust depth estimation using spatial correlation between pixels, robust class detection using spatial correlation between pixels and the ROI map where yellow regions indicate important regions to sample ($\omega = 1.3$, $N_s = 32^2$, $t_0 = 250\mu s$, $T = 1500$).

akin to target detection where the target is a single class of interest. Figs. 10(b) and (d) show the resulting depth RMSEs for $\omega = 66$ and $\omega = 1.3$, respectively. The latter figure reflects an interesting and challenging real-world scenario as we are most likely to be interested in locating: i) objects that occupy a small portion of a scene, ii) with a specific spectral signature, and iii) within an adverse environment (low SBR). In that case, AS shows a significant improvement in acquisition time when compared to other strategies, reaching an improvement factor of 10 and 200 at RMSE $=$ 1mm for high and low SBR, respectively. Finally, Fig. 12 (bottom-row) compares the proposed AS to other algorithms on the Lego scene. This figures confirms best performance with the proposed AS framework.

VII. Conclusion

This paper has presented a task optimized adaptive sampling framework for multispectral single-photon LiDAR data. The iterative approach samples new points to reduce the uncertainties of the parameters of interest and hence improve their estimates. We demonstrated the framework when considering two tasks: (i) a classification task where the goal is to improve the labeling of pixels based on their spectral signatures, and (ii) signature based object detection. In both cases, pixels only containing background photons are ignored to concentrate on informative pixels with target reflections. A new Bayesian model was proposed to perform these tasks by providing the parameter posterior distributions, which contain parameter estimates together

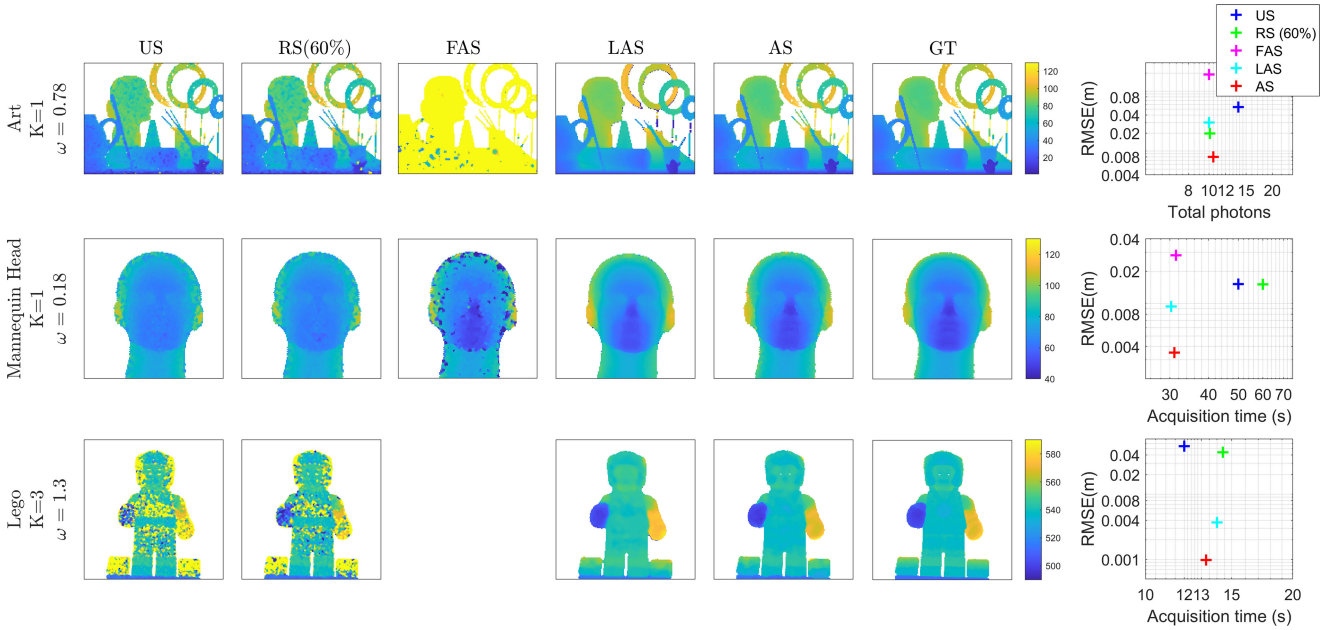


Fig. 12. Reconstructed depth maps using the studied sampling strategies for different scenes. The FAS algorithm is not applied to the multispectral Lego data as it assumes monochromatic data.

with a measure of their uncertainties. Several experiments on simulated and real data were performed showing the clear benefit of the proposed framework when compared to static sampling strategies. More precisely, we demonstrated faster convergence of the depth estimate especially in realistic scenarios involving a high background, and/or spatially small targets of interest. We have also studied the use of different detector array sizes, together with scanning modes that can be sequential or parallel. As expected, it was observed that parallel scanning allows for faster acquisition. While large arrays allows faster scanning, it was also shown that small arrays allow finer sampling and hence the targets shape. Future work will investigate a new Bayesian formulation based on photon events to avoid building histograms of counts as in the current approach. Generalizing the proposed Bayesian model to account for multiple peaks per pixel and for recursive parameter prior distributions will also be the subject of future work.

References

[1] A. M. Wallace, A. Halimi, and G. S. Buller, "Full waveform LiDAR for adverse weather conditions," *IEEE Trans. Veh. Technol.*, vol. 69, no. 7, pp. 7064–7077, Jul. 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 Process. Mag.*, vol. 37, no. 4, pp. 62–71, Jul. 2020.

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

[4] H. Tan *et al.*, "Deep learning based single-photon 3D imaging with multiple returns," in *Proc. Int. Conf. 3D Vis.*, 2020, pp. 1196–1205.

[5] Z.-P. Li *et al.*, "Super-resolution single-photon imaging at 8.2 kilometers," *Opt. Exp.*, vol. 28, no. 3, pp. 4076–4087, 2020.

[6] 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, 2009.

[7] Z.-P. Li *et al.*, "Single-photon computational 3D imaging at 45 km," *Photon. Res.*, vol. 8, no. 9, pp. 1532–1540, 2020.

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

[9] Z. Li *et al.*, "Multi-beam single-photon-counting three-dimensional imaging LiDAR," *Opt. Exp.*, vol. 25, no. 9, pp. 10 189–10 195, 2017.

[10] A. Maccarone *et al.*, "Underwater depth imaging using time-correlated single-photon counting," *Opt. Exp.*, 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. Imag.*, vol. 3, no. 3, pp. 472–484, Sep. 2017.

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

[13] Y. Altmann, A. Maccarone, A. Halimi, A. McCarthy, G. Buller, and S. McLaughlin, "Efficient range estimation and material quantification from multispectral LiDAR waveforms," in *Proc. Sensor Signal Process. Defence*, 2016, pp. 1–5.

[14] 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. Imag.*, vol. 6, pp. 138–152, 2020.

[15] A. Halimi, A. Maccarone, R. A. Lamb, G. S. Buller, and S. McLaughlin, "Robust and guided Bayesian reconstruction of single-photon 3D lidar data: Application to multispectral and underwater imaging," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 961–974, 2021.

[16] I. Gyongy *et al.*, "High-speed 3D sensing via hybrid-mode imaging and guided upsampling," *Optica*, vol. 7, no. 10, pp. 1253–1260, Oct. 2020.

[17] A. W. Bergman, D. B. Lindell, and G. Wetzstein, "Deep adaptive LiDAR: End-to-end optimization of sampling and depth completion at low sampling rates," in *Proc. IEEE Int. Conf. Comput. Photography*, 2020, pp. 1–11.

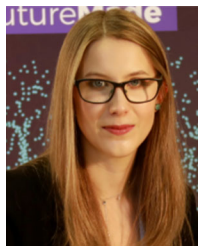
[18] D. B. Phillips *et al.*, "Adaptive foveated single-pixel imaging with dynamic supersampling," *Sci. Adv.*, vol. 3, no. 4, 2017, Art. no. e1601782.

[19] R. Tobin *et al.*, "Comparative study of sampling strategies for sparse photon multispectral LiDAR imaging: Towards mosaic filter arrays," *J. Opt.*, vol. 19, no. 9, 2017, Art. no. 094006.

[20] A. Halimi, P. Ciuciu, A. Mccarthy, S. Mclaughlin, and G. S. Buller, "Fast adaptive scene sampling for single-photon 3D LiDAR images," in *Proc. IEEE 8th Int. Workshop Comput. Adv. Multi-Sensor Adaptive Process.*, 2019, pp. 196–200.

[21] 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. Imag.*, vol. 6, pp. 208–220, 2020.
[22] J. Rapp and V. K. Goyal, "A few photons among many: Unmixing signal and noise for photon-efficient active imaging," *IEEE Trans. Comput. Imag.*, vol. 3, no. 3, pp. 445–459, Sep. 2017.
[23] A. Halimi *et al.*, "Restoration of intensity and depth images constructed using sparse single-photon data," in *Proc. 24th Eur. Signal Process. Conf.*, 2016, pp. 86–90.
[24] J. Rapp, Y. Ma, R. M. Dawson, and V. K. Goyal, "High-flux single-photon LiDAR," *Optica*, vol. 8, no. 1, pp. 30–39, 2021.
[25] Z. Tasneem, C. Adhivarahan, D. Wang, H. Xie, K. Dantu, and S. J. Koppal, "Adaptive fovea for scanning depth sensors," *Int. J. Robot. Res.*, vol. 39, no. 7, pp. 837–855, 2020.
[26] S. Li *et al.*, "Spatially adaptive retina-like sampling method for imaging LiDAR," *IEEE Photon. J.*, vol. 11, no. 3, Jun. 2019, Art. no. 6901216.
[27] F. Ma and S. Karaman, "Sparse-to-dense: Depth prediction from sparse depth samples and a single image," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2018, pp. 4796–4803.
[28] F. Ma, G. V. Cavalheiro, and S. Karaman, "Self-supervised sparse-to-dense: Self-supervised depth completion from lidar and monocular camera," in *Proc. Int. Conf. Robot. Automat.*, 2019, pp. 3288–3295.
[29] Y. Xu, X. Zhu, J. Shi, G. Zhang, H. Bao, and H. Li, "Depth completion from sparse LiDAR data with depth-normal constraints," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 2811–2820.
[30] E. Gofer, S. Praisler, and G. Gilboa, "Adaptive LiDAR sampling and depth completion using ensemble variance," *IEEE Trans. Image Process.*, vol. 30, pp. 8900–8912, 2021.
[31] G. D. P. Godaliyadda, D. H. Ye, M. D. Uchic, M. A. Groeber, G. T. Buzzard, and C. A. Bouman, "A framework for dynamic image sampling based on supervised learning," *IEEE Trans. Comput. Imag.*, vol. 4, no. 1, pp. 1–16, Mar. 2018.
[32] 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, May 2016.
[33] A. McCarthy *et al.*, "Kilometer-range, high resolution depth imaging via 1560 nm wavelength single-photon detection," *Opt. Exp.*, vol. 21, no. 7, pp. 8904–8915, 2013.
[34] G. M. Martín *et al.*, "High-speed object detection using SPAD sensors," *Proc. SPIE*, vol. 11693, 2021, Art. no. 116930L.
[35] E. Wade, R. Tobin, A. McCarthy, and G. S. Buller, "Sub-pixel micro scanning for improved spatial resolution using single-photon LiDAR," *Proc. SPIE*, vol. 11721, 2021, Art. no. 1172106.
[36] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, vol. 2. Springer, 1999.
[37] S. Hernandez-Marin, A. M. Wallace, and G. J. Gibson, "Bayesian analysis of LiDAR signals with multiple returns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 12, pp. 2170–2180, Dec. 2007.
[38] J. Tachella, Y. Altmann, S. McLaughlin, and J.-Y. Tourneret, "On fast object detection using single-photon LiDAR data," *Proc. SPIE*, vol. 11138, 2019, Art. no. 111380T.
[39] Y. Altmann, X. Ren, A. McCarthy, G. S. Buller, and S. McLaughlin, "Robust Bayesian target detection algorithm for depth imaging from sparse single-photon data," *IEEE Trans. Comput. Imag.*, vol. 2, no. 4, pp. 456–467, Dec. 2016.
[40] Y. Altmann *et al.*, "Robust spectral unmixing of sparse multispectral LiDAR waveforms using gamma Markov random fields," *IEEE Trans. Comput. Imag.*, vol. 3, no. 4, pp. 658–670, Dec. 2017.

Mohamed Amir Alaa Belmekki (Student Member, IEEE) received the Eng. degree in electronics and the M.Sc. degree in signal and communications from the National Polytechnic School of Algiers, Algeria, in 2019. He is currently working toward the Ph.D. degree as a teaching and Research Assistant with the Institute of Sensors, Signals and System's (ISSS) of the School of Engineering and Physical Science (EPS), Heriot-Watt University, Edinburgh, U.K. His research interests include computational imaging and statistical signal and image processing, with a particular interest in Bayesian inverse problems on single-photon applications such as depth imaging and remote sensing (multispectral imaging).



Rachael Tobin received the B.Sc. (Hons.) and Ph.D. degrees in physics from Heriot Watt University, Edinburgh, U.K., in June 2015 and November 2019. Her main research interests include single-photon depth imaging in challenging conditions and remote sensing. In 2019, she won the annual MacFarlane Prize for Heriot-Watt University's best Ph.D. thesis. In 2021, Rachael was awarded a Royal Academy of Engineering (RAEng) Research Fellowship to investigate single-photon imaging and sensing.

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 with Heriot-Watt University. In 2002, he co-founded Helia Photonics Ltd., based in Livingston, U.K. From 2015 to 2021, he was an EPSRC Established Career Fellowship in Quantum Technology to investigating 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 U.K. Institute of Physics, and the Optical Society of America.

Stephen McLaughlin (Fellow, IEEE) received the B.Sc. degree from the University of Glasgow, Glasgow, U.K., in 1981, and the Ph.D. degree from the University of Edinburgh, Edinburgh, U.K., in 1990. From 1981 to 1986, he was a Development Engineer in industry. In 1986, he joined the Department of Electronics and Electrical Engineering with the University of Edinburgh and ultimately held a Chair in Electronic Communication Systems. In October 2011, he joined Heriot-Watt University, as a Professor of signal processing. Prof. McLaughlin is a Fellow of the Royal Academy of Engineering, of the Royal Society of Edinburgh, of the Institute of Engineering and Technology, of the IEEE and is a EURASIP Fellow.

Abderrahim Halimi (Senior Member, IEEE) 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 Postdoctoral 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 (EPS) in Heriot-Watt University, Edinburgh, U.K. His research interests include statistical signal and image processing, with a particular interest in Bayesian inverse problems with applications to single-photon depth imaging, remote sensing (hyperspectral imaging, satellite altimetry) and medical imaging.