

Fast adaptive scene sampling for single-photon 3D lidar images

Abderrahim Halimi, Philippe Ciuciu, Aongus Mccarthy, Stephen Mclaughlin, Gerald S Buller

▶ **To cite this version:**

Abderrahim Halimi, Philippe Ciuciu, Aongus Mccarthy, Stephen Mclaughlin, Gerald S Buller. Fast adaptive scene sampling for single-photon 3D lidar images. IEEE CAMSAP 2019 - International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, Dec 2019, Le Gosier (Guadeloupe), France. hal-02298998

HAL Id: hal-02298998
https://inria.hal.science/hal-02298998

Submitted on 27 Sep 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

FAST ADAPTIVE SCENE SAMPLING FOR SINGLE-PHOTON 3D LIDAR IMAGES

Abderrahim Halimi[(1)]*, Philippe Ciuciu*[(2,3)]*, Aongus McCarthy*[(1)]*,*
Stephen McLaughlin[(1)]*, Gerald S. Buller*[(1)]

[(1)] School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh U.K.
[(2)] CEA/NeuroSpin, Univ. Paris-Saclay, F-91191 Gif-sur-Yvette, France.
[(3)] Parietal Team, INRIA Saclay, Universite Paris-Saclay, Saclay, France.

ABSTRACT

Reducing acquisition time is a major challenge for single-photon based imaging. This paper presents a new approach for adaptive scene sampling allowing for faster acquisition when compared to classical uniform sampling or random sampling strategies. The approach is applied to the laser detection and ranging (Lidar) three-dimensional (3D) imaging where sampling is optimized regarding the depth image. Based on data statistics, the approach starts by achieving a robust estimation of the depth image. The latter is used to generate a map of regions of interest that informs next samples positions and their acquisition times. The process is repeated until a stopping criterion is met. A particular interest is given to fast processing to allow real-world application of the proposed approach. Results on real data show the benefits of this strategy that can reduce acquisition times by a factor of 8 compared to uniform sampling in some scenarios.

***Index Terms*—** Adaptive sampling, 3D imaging, Single-photon Lidar, Poisson statistics, robust estimation.

1. INTRODUCTION

3D Lidar Imaging using time-correlated single-photon detectors has generated significant interest in the scientific community in recent years. This is a direct consequence of its ability to provide high-resolution depth and reflectivity profiles of the observed targets. The imaging system emits picosecond duration laser pulses and records the arrival times of the reflected photons using a time-correlated single-photon counting (TCSPC) module. This operation is repeated for different beam locations to build a cube of data containing the 3D information about the target. As a result to this repetitive process, the imaging acquisition time is often high due to the large number of scanned pixels and the necessary dwell time per-pixel to collect enough photons.

Several strategies have been proposed in the literature to deal with this challenge either by optimizing data acquisition [1–3] or by proposing sophisticated post-processing algorithms to restore the collected data in the photon sparse regime [4–8]. The former strategy is considered in this paper. This includes compressive sensing methods that sample random locations and restore data using a reconstruction algorithm [2, 9], or scene-dependent sampling methods that carefully select informative locations to scan [1, 10]. In this case, the notion of informative data will mainly depend on the task to achieve and the sampling of the same scene may differ if the final task is, for example, about target detection, reconstruction, classification, etc.

This paper proposes a new scene-dependent sampling strategy that is optimized for target depth reconstruction but can be easily generalized to other tasks. The online approach operates in an adaptive manner by selecting the most informative locations to scan at each iteration, and optimizing their acquisition/dwell time. It is based on three steps, (i) robust estimation of the depth map using data statistics, (ii) generation of a map of regions of interest (ROI), and (iii) sampling the new scan positions and their acquisition times. A particular interest is given to fast processing to allow real-world application of the proposed approach. The proposed algorithm is validated on real data by studying different scenarios. A comparison with uniform sampling (US) and random sampling (RS) strategies shows a clear improvement in performance.

The paper is structured as follows. The considered problem is introduced in Section 2 including details about the Lidar observation model and parameters estimation. Section 3 presents the proposed adaptive sampling strategy. Results on real data are presented in Section 4. Conclusions and future work are finally discussed in Section 5.

2. PROBLEM FORMULATION

Imaging or sensing collects information to help improve our understanding of a phenomenon and perform a given task. In Lidar, we aim at acquiring information about the object depth, reflectivity to perform image reconstruction or other high level tasks such as object detection and/or recognition. This paper proposes a new strategy for scene adaptive sampling of Lidar data to improve depth estimation while reduc-

This work was supported by the UK Royal Academy of Engineering under the Research Fellowship Scheme (RF/201718/17128), and EPSRC Grants EP/J015180/1, EP/N003446/1, EP/M01326X/1, EP/K015338/1, EP/S000631/1.

ing the acquisition time. The next subsections introduce the data observation model and parameters estimation which are necessary ingredients for our sampling approach.

2.1. Observation model

The TCSPC system provides time-tagged photons that can be gathered into a histogram $\boldsymbol{y}_{n,t}$, denoting the number of detected photons at pixel location $n \in \{1, \cdots, N\}$, and time-of-flight bin $t \in \{1, \cdots, T\}$. Each photon counts can be assumed to be drawn from the Poisson distribution $\mathcal{P}(.)$ as follows [4, 11]:

$$y_{n,t} \sim \mathcal{P}(s_{n,t}), \text{ with } s_{n,t} = r_n f(t - d_n) + b_n \quad (1)$$

where $d_n \geq 0$ denotes the distance of the object from the sensor (related to its depth), $r_n \geq 0$ is the reflectivity of the object surface, $b_n \geq 0$ is the background and dark counts of the detector, and f represents the system impulse response assumed to be known from a calibration step. Note that (1) assumes the presence of one peak in each pixel location.

The joint likelihood can be expressed, after assuming independence between the observed pixels, as follows:

$$P(\boldsymbol{Y}|\boldsymbol{d}, \boldsymbol{r}, \boldsymbol{b}) = \prod_n \prod_{t=1}^{T} \frac{s_{n,t}^{y_{n,t}}}{y_{n,t}!} e^{-s_{n,t}} \quad (2)$$

where $\boldsymbol{d}, \boldsymbol{r}, \boldsymbol{b}$ are $N \times 1$ vectors gathering the elements d_n, r_n, $b_n, \forall n$ (in lexicographic order), with N the number of pixels, and T is the total number of bins. Note finally that time-tagged data can be also used to simulate a reduction in acquisition time which will be useful during our simulations.

2.2. Depth and reflectivity estimation

The maximum likelihood estimates of the depth and reflectivity of the target $(\boldsymbol{d}, \boldsymbol{r})$ can be obtained by minimizing the negative log-likelihood given by (after removing unnecessary constants)

$$\begin{aligned} \mathcal{L} &= -\log\left[P(\boldsymbol{Y}|\boldsymbol{d}, \boldsymbol{r}, \boldsymbol{b})\right] \\ &= \sum_n \sum_{t=1}^{T} \left[s_{n,t} - y_{n,t} \log\left(s_{n,t}\right)\right] + cst \end{aligned} \quad (3)$$

To obtain fast estimates, it is common to assume the absence of the background level, i.e., $b_n = 0$ (see [11, 12]) and that the observation window is larger than the target's depth and the width of the impulse response, in which case $c_1 = \sum_{t=1}^{T} f(t - d_n)$ is a constant for all realistic depths d_n. Under these assumptions, the depth of each pixel can be estimated using a log-matched filter of the measured histograms, and the reflectivity is given by $r_n^{\text{ML}} = \frac{1}{c_1}\left(\sum_{t=1}^{T} y_{n,t}\right)$.

3. DEPTH BASED ADAPTIVE SAMPLING

The proposed strategy is summarized in Alg. 1. Given a high-resolution (HR) grid of N pixel positions, the algorithm performs a first scan of a reduced number of pixels N_s (whose indices are denoted by ν_{N_s}) with a low acquisition time-per-pixel $\boldsymbol{t}_n^{\text{acq}} = t_0$ for $n \in \nu_{N_s}$, which will provide a first estimate of the scene. The measured time-tagged photons are then loaded to the processing unit to build fast and robust estimates of the depth and reflectivity images. These are then used to build a map of regions of interest that controls the position and acquisition times of the new sampled points. The process is iteratively repeated until convergence. The next subsections describe in more details each of these steps.

Algorithm 1 Scene based adaptive sampling

1: Initialization
2: Initialize: N_s, N, K, c, t_0, conv=0
3: **while** conv$= 0$ **do**
4: Scan N_s points with acquisition times $\boldsymbol{t}_{\nu_{N_s}}^{\text{acq}}$
5: Update t_0 depending on pixels with detected photons
6: Robust estimation of depth and reflectivity
7: Computation of a probability map of ROI
8: Generation of new positions (using a MH algorithm) and new acquisition times $\boldsymbol{t}_{\nu_{N_s}}^{\text{acq}}$
9: conv$= 1$, if the stopping criteria is satisfied.
10: **end while**

3.1. Fast and robust parameter estimation

Our strategy requires a fast update of the scanning points, thus, fast parameter estimation. Although the approach in Section 2.2 is relatively fast, it relies on building histograms and solving a maximization problem which spends a large computational cost preventing real time processing. Therefore, our strategy further assumes a Gaussian approximation for the instrument impulse response $f(t - t_n) = c_1 \exp^{-\frac{(t-t_n)^2}{2\sigma^2}}$ as in [12, 13] which leads to the following analytical maximum likelihood estimates for the reflectivity $r_n^{\text{ML}} = \frac{1}{c_1}\left(\sum_{t=1}^{T} y_{n,t}\right)$ and the depth $t_n^{\text{ML}} = \frac{\left(\sum_{t=1}^{T} t y_{n,t}\right)}{c_1 r_n^{\text{ML}}}$ as reported in [12]. Note that these estimates can be computed using time-tagged photons and simple operations which allow fast processing. Obviously, these estimates are of poor quality especially in the interesting limit of: (i) very low photon counts (i.e., fast or long-range imaging [14]) or (ii) when few pixels are scanned. However, with some modifications, the estimates will provide enough information allowing the generation of the sampling ROI map. To deal with these limitations, we rely on spatial correlation between pixels [4–7] to improve the estimates and retrieve missing information. More precisely, each pixel is associated with a graph of neighbours, i.e., a window of $3^K \times 3^K$ pixels from the HR grid. The empty pixels (i.e., non-scanned pixels or pixels with no-detection) are then filled using the counts from the closest observed neighbours. A high sampling probability is assigned to pixels when all neighbours are empty. Using the reconstructed data, a full depth and reflectivity maps can be generated. Note that the neighbours grid can be pre-computed to allow fast

processing. Note also that we considered a uniform HR grid in this paper, however, the algorithm can also be used with a non-uniform grid without changes. Note finally that other robust estimation strategies can also be considered provided that the estimation is fast and robust [15].

3.2. Regions of interest

This section generates a probability map $\boldsymbol{m} \in [0,1]^{N\times 1}$ containing informative regions in the HR grid. Obviously, this map is related to the targeted task, for example, for a classification (respectively tracking) task it delivers high probabilities around class boundaries (resp. the object of interest). In this paper, we are interested in improving depth estimates and we build the map for this specific task. Several features can be defined to capture the important information from the previously estimated depth image [10]. The latter can include the spatial gradients of the depth to catch surface variations, local variances to catch the object structure, etc. We considered horizontal and vertical gradients in our implementation which showed satisfactory performance for the considered simulations (see Section 4). The resulting map will be normalized by the sum of its values to represent a probability density function on the HR grid. Note that empty pixels/regions will receive the highest probability value to ensure they will be sampled at the next step, while pixels that reached maximum acquisition times will be assigned a zero probability to avoid sampling their locations.

3.3. Generation of new locations and acquisition times

Given the probability map $\boldsymbol{m}$, one can sample the N_s new locations by selecting those with highest probabilities (or those providing maximum error reduction as in [10]). We found that this strategy tends to concentrate the samples in small regions while reducing the exploration of the full scene which might prevent the detection of small objects or dynamic effects. Therefore, we propose to randomly sample the new locations using the probability map $\boldsymbol{m}$. This is achieved using a Metropolis-Hastings algorithm [16] with a uniform proposal distribution on the HR grid. Other sophisticated strategies can instead be used to improve sampling in presence of a small or multiple regions of interest such as Hamiltonian Monte Carlo [17]. Each generated location will then be assigned an acquisition time proportional to it's probability value. For this, we sorted the N_s probability values into c decreasing levels and assigned decreasing acquisition time ranging from ct_0 to t_0, where t_0 is a user defined acquisition time step. The latter is updated during iterations to ensure that the proportion of scanned points with detections is between $[0.7, 0.9]$. This will avoid working with small time steps (low detections) or very high steps (high detections).

3.4. Stopping criteria

Several stopping criteria are considered for the iterative algorithm 1. The first criterion compares the two last parameter estimates (depth and reflectivity) and stops the algorithm if the average relative difference is smaller than a given threshold ξ, i.e., $|\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}| \leq \xi|\boldsymbol{x}^{(t)}|$, where $|.|$ denotes the absolute value and $\boldsymbol{x}$ denotes $\boldsymbol{d}$ or $\boldsymbol{r}$. A maximum acquisition time-per-pixel can also be provided to stop the algorithm once reached for all pixels. The algorithm can also be stopped by specifying a maximum number of scanned points or equivalently a maximum number of iterations.

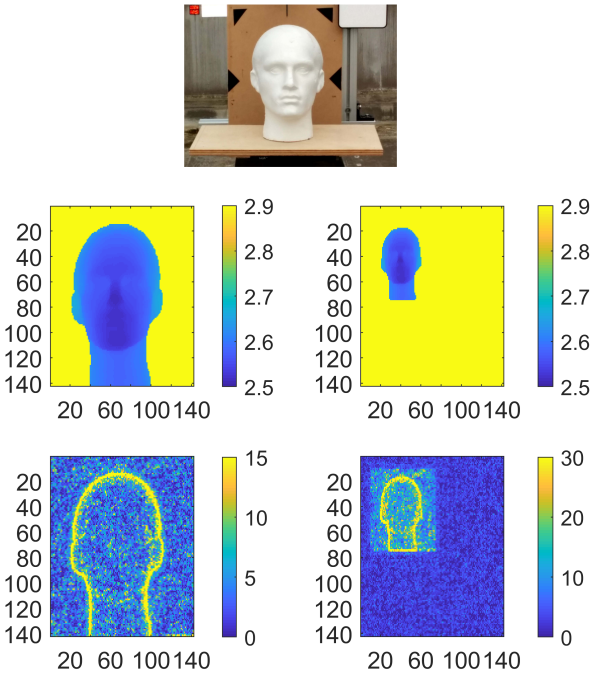


Fig. 1. (Top) The considered mannequin face scene, (middle-left) Reference depth image (3 ms acquisition time per-pixel), (middle-right) Reference of a simulated depth image of a small target, (bottom-left) Number of scanned samples using the proposed algorithm for a large target, (bottom-right) Number of scanned samples using the proposed algorithm for a small target.

4. RESULTS ON REAL DATA

The performance of the proposed algorithm is evaluated on a real single-photon data of a life-sized mannequin head acquired at a distance of 40m in November 2014 on the Edinburgh Campus of Heriot-Watt University. The data contains $N = 142 \times 142$ pixels representing the HR grid, and we consider a maximum acquisition time of 3ms using a time-of-flight scanning sensor, based on TCSPC (see Fig. 1 (top)). The transceiver system and data acquisition hardware used for this work are described in [11, 18]. As previously described, the TCSPC delivers time-tagged photons where each detected photon is associated with its arrival time. Thus, one can select the detected photons associated with a given acquisition times, which allows the simulation of the proposed

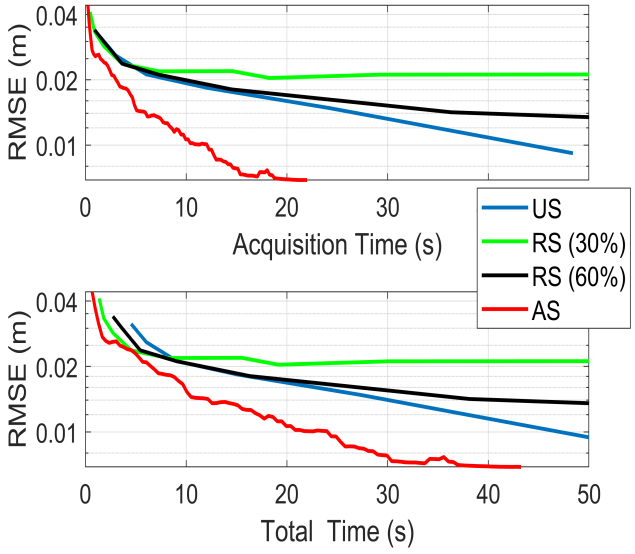


Fig. 2. Depth RMSEs of the full FOV target for different sampling strategies w.r.t. (top) dwell time, (bottom) total time.

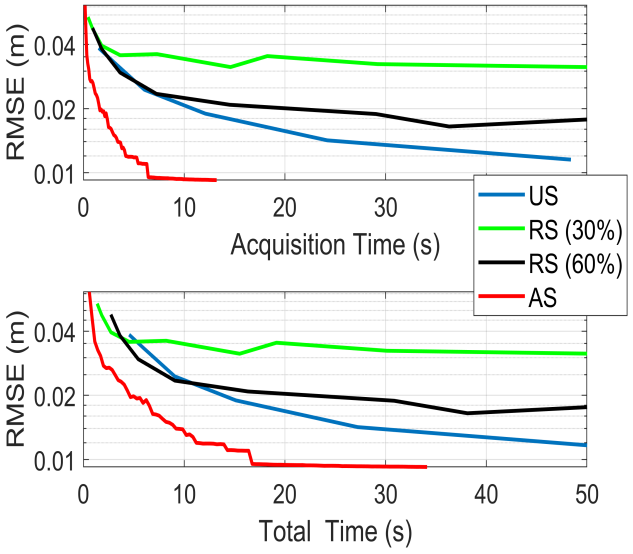


Fig. 3. Depth RMSEs of the small target for different sampling strategies w.r.t. (top) dwell time, (bottom) total time.

adaptive sampling (AS) strategy. The proposed approach is compared with the uniform sampling (US) strategy, and the random sampling (RS) strategy that only scan a ratio of the N pixels. Empty pixels of all methods are restored using the strategy described in Section 3.1. To evaluate the performance of the different sampling strategies, depth maps are estimated using the maximum likelihood estimator of Section 2.2 and the results are quantitatively evaluated using the mean square error defined by RMSE $= \sqrt{\frac{1}{N}||\boldsymbol{d}^{\text{ref}} - \widehat{\boldsymbol{d}}||^2}$, where $\boldsymbol{d}^{\text{ref}}$ is obtained from the US data at 3ms (see Fig. 1 (middle-left)). The proposed AS method considers $N_s = 32^2$ pixels, $t_0 = 75\mu$s, $c = 3$, $K = 2$, 100 maximum iterations, and RS is studied for 30% and 60% of pixels. Fig. 2 (top) shows the RMSE curves of the different sampling strategies w.r.t. the acquisition or dwell time (top figure), while Fig. 2 (bottom) shows the RMSE w.r.t. the total time that includes acquisition/dwell time, the time to move the scanning mirrors and the computational cost of Alg. 1 (bottom figure). This figure highlights the interest of the proposed strategy that presents an improvement in acquisition time by a factor of ≈ 5 compared with US and > 5 w.r.t. RS for large acquisition times. The improvement factor is obviously reduced when considering the total time due to the processing time of Alg. 1, but the results are still better than US and RS. In Fig. 1 (middle-left) the target mannequin face is ideally positioned while occupying most of the field-of-view. This is an ideal scenario that is not generally obtained in real-world scenarios where the target of interest might only occupy a small proportion of the field-of-view. To simulate such scenario, we down-sampled the mannequin face and create a simulated scene containing a large backboard with a small target as shown in Fig. 1 (middle-right), where the target occupies 64×64 pixels of the HR grid. Fig. 3 shows the clear benefit of the AS strategy in this scenario reaching an improvement factor of 8 w.r.t. US. Note that the scanned samples are concentrated on the mannequin face features (i.e., face edges, ears) to improve its depth estimate, as shown in Fig. 1 (bottom row). Note finally that one iteration of AS costs 50ms in average (with a Matlab implementation) and that we approximate the time of one mirror move by 150μs in Figs. 2 and 3.

5. CONCLUSIONS

This paper has presented a new depth adaptive sampling strategy for single-photon 3D Lidar data. The approach is based on a fast estimation of the depth map and the use of its features to build a probability map of the regions of interest. A Metropolis-Hastings algorithm is used to generate the new scan positions and the probabilities are used to control their acquisition times. The algorithm was validated on real Lidar data showing a clear improvement in performance when compared to other sampling strategies. Despite good performance, several points can still be improved including the generalization to other tasks such as object tracking or classification, and the use of other features to generate robust probability maps. Considering sophisticated sampling strategies such as Metropolis-adjusted Langevin algorithm, Hamiltonian Monte Carlo or algorithm [19] is important and will help with scenes showing small or multiple targets. Finally, developing faster implementations to further reduce the computational time of the approach is important in order to facilitate real time processing.

6. REFERENCES

[1] D. B. Phillips, M.-J. Sun, J. M. Taylor, M. P. Edgar, S. M. Barnett, G. M. Gibson, and M. J. Padgett, "Adaptive foveated single-pixel imaging with dynamic supersampling," *Science Advances*, vol. 3, no. 4, 2017.

[2] R. Tobin, Y. Altmann, X. Ren, A. McCarthy, R. A. Lamb, S. McLaughlin, and G. S. Buller, "Comparative study of sampling strategies for sparse photon multispectral lidar imaging: towards mosaic filter arrays," *Journal of Optics*, vol. 19, no. 9, p. 094006, aug 2017.

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

[4] 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.

[5] J. Tachella, Y. Altmann, X. Ren, A. McCarthy, G. S. Buller, J.-Y. Tourneret, and S. McLaughlin, "Bayesian 3d reconstruction of complex scenes from single-photon lidar data," in *ArXiv e-prints*, Oct. 2018.

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

[7] A. Halimi, X. Ren, A. McCarthy, J. Bioucas-Dias, S. McLaughlin, and G. S. Buller, "Robust restoration of sparse multidimensional single-photon LiDAR images," *IEEE Trans. Comput. Imaging*, 2019, in press.

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

[9] M. F. Duarte, M. A. Davenport, D. Takbar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, March 2008.

[10] G. M. 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. Imaging*, vol. 4, no. 1, pp. 1–16, March 2018.

[11] Y. Altmann, X. Ren, A. McCarthy, G. S. Buller, and S. McLaughlin, "Lidar waveform based analysis of depth images constructed using sparse single photon data," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 1935–1946, Mar. 2015.

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

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

[14] 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.

[15] J. Tachella, Y. Altmann, N. Mellado, A. McCarthy, R. Tobin, G. S. Buller, J.-Y. Tourneret, and S. McLaughlin, "Real-time 3d reconstruction of complex scenes using single-photon lidar: when image processing meets computer graphics," in *ArXiv e-prints*, May 2019.

[16] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. New York: Springer-Verlag, 1999.

[17] S. Brooks, A. Gelman, G. L. . Jones, and X.-L. Meng, *Handbook of Markov chain Monte Carlo*. ser. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. Taylor & Francis, 2011.

[18] A. McCarthy, X. Ren, A. D. Frera, N. R. Gemmell, N. J. Krichel, C. Scarcella, A. Ruggeri, A. Tosi, and G. S. Buller, "Kilometer-range depth imaging at 1550 nm wavelength using an InGaAs/InP single-photon avalanche diode detector," *Opt. Express*, vol. 21, no. 19, pp. 22 098–22 113, Sep 2013.

[19] C. Lazarus, P. Weiss, N. Chauffert, F. Mauconduit, L. El Gueddari, C. Destrieux, I. Zemmoura, A. Vignaud, and P. Ciuciu, "Sparkling: variable-density k-space filling curves for accelerated t2*-weighted mri," *Magnetic Resonance in Medicine*, vol. 81, no. 6, pp. 3643–3661, 2019.