A Bayesian Based Unrolling Approach to Single-Photon Lidar Imaging through Obscurants

Jakeoung Koo, Abderrahim Halimi, Aurora Maccarone, Gerald S. Buller and Stephen McLaughlin
School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh UK

***Abstract*—In this paper, we propose a deep learning model for 3D single-photon Lidar imaging through obscurants, i.e., in the presence of a high and non-uniform background. The proposed method unrolls the iterative steps of a Bayesian based-algorithm into the layers of a deep neural network. To deal with imaging through obscurants, the method first unmix signal and background photons in a pre-processing step. Following this, the method builds on multiscale information to improve robustness to noise and uses the attention framework for scale selection within the network. Experimental results on simulated and real underwater data demonstrate that our method can estimate accurate depth maps in challenging situations with a high non-uniform background. Compared to state-of-the-art deep learning methods, the proposed method enables an estimation of parameters uncertainties, suitable for decision making.**

***Index Terms*—3D reconstruction, single-photon imaging, Lidar, obscurants, algorithm unrolling, attention**

I. Introduction

Single-photon Lidar is a powerful depth imaging technique when reconstructing 3D scenes in challenging low-illumination situations and has a wide range of applications [1], [2]. Single-photon Lidar systems illuminate 3D scenes with laser pulses and count reflected photons, obtaining a histogram of photon counts with respect to the Time-of-Flight (ToF). The histogram data contains information on the depth and reflectivity of targets and reconstructing a scene from the data has recently been the subject of study by many researchers. One challenging situation is the presence of non-uniform background noise in data due to obscurants [3]–[7] such as fog. From such noisy data, robustly reconstructing 3D information is a requirement in many applications such as autonomous driving [1].

Existing reconstruction methods can be divided into two broad categories: statistical methods and deep learning methods. Statistical methods [8]–[10] model single-photon Lidar data with some prior information such as spatial correlation. These methods provide good interpretability, but their solutions depend on some imposed priors and user-defined parameters. Current deep learning methods learn features from the data by adopting conventional architecture such as U-net [11], [12] or a non-local block to consider long-range correlations [13]. Such methods can generalize to unseen photon-sparse or noisy data, but lack interpretability and often suffer from over-smoothing effects around surface boundaries.

This work was supported by the UK Royal Academy of Engineering under the Research Fellowship Scheme (RF/201718/17128) and EPSRC Grants EP/T00097X/1,EP/S000631/1,EP/S026428/1.

This paper suggests a hybrid approach using a statistical model to design a deep learning architecture for the reconstruction of challenging Lidar data with non-uniform background. Our hybrid method unfolds an underlying Bayesian method [14] into an interpretable neural network, where weighted median filtering and a soft-shrinkage operator in [14] are replaced by attention-based network layers. We generalize the recently proposed method [15] to imaging through obscurants presenting non-uniform background. Our proposed method has three main benefits. First, our unrolling approach leads to an efficient neural network with less parameters and with fast inference time. Second, the proposed architecture is interpretable and produces uncertainty information on the final output, a distinct advantage lacking in previous deep learning methods for single-photon Lidar [11], [13]. Third, the proposed method introduces an efficient scheme to remove non-uniform background photons which often degrade previously reported deep learning methods. These advantages are demonstrated with experiments on synthetic and real underwater data in the presence of non-uniform background.

II. Approximate multiscale observation model

A Lidar system obtains a histogram of counts $y_{n,t}$ at the n-th pixel and the t-th time bin which follows a Poisson likelihood $y_{n,t} \sim \mathcal{P}(s_{n,t})$ [8] with the mean $s_{n,t}$ given by

$$s_{n,t} = r_n\, g\,(t - d_n) + b_{n,t}, \tag{1}$$

where r_n is a target's reflectivity, g is a system impulse response function (IRF), d_n denotes a target's depth and $b_{n,t}$ denotes the background of counts due to the environment. To deal with obscurants, one can incorporate the background term $b_{n,t}$ in the likelihood model, which might lead to a computationally expensive algorithm. Instead, our strategy is to remove the background counts through pre-processing (Section IV-D) and consider a background-free approximate likelihood model. We further assume $\sum_{t=1}^{T} g\,(t - d_n) = 1$ with the total number of bins T; the shape of the system IRF is a Gaussian function $\mathcal{N}(\cdot\,; \mu, \sigma)$ with the mean μ and the standard deviation σ; and independent observations with respect to different pixels. Under these assumptions, the likelihood function for $\boldsymbol{Y} = \{y_{n,t}\}$ is given by

$$p\,(\boldsymbol{Y} \mid \boldsymbol{d}, \boldsymbol{r}) = \prod_{n=1}^{N} \prod_{t=1}^{T} \frac{s_{n,t}^{y_{n,t}}}{y_{n,t}!} e^{-s_{n,t}}, \tag{2}$$

where $\boldsymbol{d} = \{d_n\}$ and $\boldsymbol{r} = \{r_n\}$ are $N \times 1$ vectors. The maximum likelihood estimate d_n^{ML} of the depth is given by

$$d_n^{\mathrm{ML}} = \arg\max_d \sum_t y_{n,t} \log g(t-d). \tag{3}$$

By a straightforward calculation [14], we can rewrite the likelihood function as follows:

$$\begin{aligned} p\left(\boldsymbol{y}_n \mid d_n, r_n\right) \propto Q\left(\boldsymbol{y}_n\right) \mathcal{G}\left(r_n; 1+\bar{s}_n, 1\right) \\ \times \mathcal{N}(d_n; d_n^{\mathrm{ML}}, \bar{\sigma}^2), \end{aligned} \tag{4}$$

where the symbol $\propto$ denotes the proportionality; Q is a normalization factor depending on $\boldsymbol{y}_n$; $\bar{s}_n = \sum_{t=1}^{T} y_{n,t}$ and $\mathcal{G}$ is the gamma distribution with shape and scale parameters, so that $\mathcal{G}\left(r_n; 1+\bar{s}_n, 1\right) \propto r_n^{\bar{s}_n} e^{-r_n}$.

To deal with noisy data, multiscale information has been used in many previous reported work including statistical algorithms [9], [14] and deep learning methods [11]–[13]. Since the sum of independent Poisson distributions follows a Poisson distribution, applying a low-pass filter with uniform filters to a histogram will generate the down-sampled ones which still follow Poisson distribution. We downsample the original histogram to generate L multiscale histograms $\boldsymbol{y}_n^{(\ell)}$ with $\ell \in \{1, 2, ..., L\}$, with the following likelihood

$$\begin{aligned} p\left(\boldsymbol{y}_n^{(\ell)} \mid r_n^{(\ell)}, d_n^{(\ell)}\right) \propto \mathcal{G}\left(r_n^{(\ell)}; 1+\bar{s}_n^{(\ell)}, 1\right) Q\left(\boldsymbol{y}_n^{(\ell)}\right) \\ \times \mathcal{N}\left(d_n^{(\ell)}; d_n^{\mathrm{ML}(\ell)}, \bar{\sigma}^{2(\ell)}\right), \end{aligned} \tag{5}$$

with $\bar{s}_n^{(\ell)} = \sum_{t=1}^{T} y_{n,t}^{(\ell)}$ and $\bar{\sigma}^{2(\ell)} = \sigma^2 / \bar{s}_n^{(\ell)}$. We can downsample the histogram with respect to either spatial positions or time bins. In this work, we will consider $L = 12$ and the details are explained in Section IV-C.

III. Underlying Bayesian algorithm

We review the underlying Bayesian algorithm [14] used in our framework. Here, we focus on estimating depth maps, but the method [14] can also reconstruct reflectivity.

A. Prior distributions

The multiscale likelihood (5) considers multiscale depth maps derived from different noise levels. From multiscale information, the goal is to estimate a single depth map $\boldsymbol{x} = \{x_n\}$ with sharp boundaries and smooth profiles within an object. To achieve this goal, a prior is imposed on the latent variable $\boldsymbol{x}$ following the Laplace distribution as follows:

$$\begin{aligned} x_n \mid d_{\nu_n}^{(1,\cdots,L)}, w_{\nu_n,n}^{(1,\cdots,L)}, \epsilon_n \sim \\ \prod_{n' \in \nu_n} \left[\prod_{\ell=1}^{L} \mathcal{L}\left(x_n; d_{n'}^{(\ell)}, \epsilon_n / w_{n',n}^{(\ell)}\right)\right], \end{aligned} \tag{6}$$

where $\mathcal{L}$ denotes the Laplacian distribution with mean and scale parameter; ν_n a spatial neighborhood around the pixel n; ϵ_n the variance of the depth x_n; and $w_{n',n}^{(\ell)}$ the pre-defined weights to guide the correlation between multiscale depths and the latent variable x_n. In [14], these weights can be obtained using guiding information such as an additional sensor.

To estimate uncertainty on the depth maps, a prior distribution is also imposed on the variance of the depth maps $\boldsymbol{\epsilon} = \{\epsilon_n\}$ following the conjugate inverse gamma distribution

$$\boldsymbol{\epsilon} \sim \prod_n \mathcal{IG}\left(\epsilon_n; \alpha, \beta\right) \tag{7}$$

with small values for α, β to obtain a non-informative prior.

B. Posterior distribution

Combining the multiscale likelihood (5) with the prior distributions (6), (7) leads to the following posterior distribution

$$p\left(\boldsymbol{x}, \boldsymbol{\epsilon}, \boldsymbol{D} \mid \boldsymbol{Y}, \boldsymbol{W}\right) \propto p\left(\boldsymbol{Y} \mid \boldsymbol{D}\right) p\left(\boldsymbol{x}, \boldsymbol{D} \mid \boldsymbol{\epsilon}, \boldsymbol{W}\right) p\left(\boldsymbol{\epsilon}\right) \tag{8}$$

where $\boldsymbol{D} = \{d_n^{(\ell)}\}$ and $\boldsymbol{W} = \{w_{n',n}^{(\ell)}\}$. A coordinate descent algorithm is considered to approximate the maximum a posterior (MAP) estimator of (8). The latent variable $\boldsymbol{x}$ is updated by a weighted median filter with the guidance weights $\boldsymbol{W}$ and its uncertainty variable $\boldsymbol{\epsilon}$ is given by an analytic expression. The multiscale depths $\boldsymbol{D}$ are updated by a generalized soft-shrinkage operator. We refer to [14] for the details.

IV. Unrolling method

We propose a deep learning model which unfolds the underlying Bayesian algorithm [14], by replacing the internal operations by neural network blocks. To support imaging through obscurants, we also incorporate the background removal scheme. Our method is summarized in Algorithm 1.

Algorithm 1 Proposed method

1: Input: Lidar data $\boldsymbol{Y}$
2: Remove background counts (Sec. IV-D)
3: Construct the initial multiscale depths (Sec. IV-C)
4: Compute the depth map $\boldsymbol{x}$ by the network (Sec. IV-A)
5: Compute the uncertainty map $\boldsymbol{\epsilon}$ (Sec. IV-B)
6: Output: $\boldsymbol{x}, \boldsymbol{\epsilon}$

A. Network

Our network consists of K stages having the same structure except for the last one, and the weights are not shared among stages. As shown in Fig. 1, each stage inputs a set of multiscale depths $\boldsymbol{d}$ and consists of feature extraction, the squeeze block and the expansion block. The features of $\boldsymbol{d}$ are extracted by three convolution layers with 3×3 filters. Throughout the network, all the convolutional layers use the 3×3 filters with LeakyReLU activation except for PAConv shown in Fig. 1.

Squeeze block. The obtained features are fed into the squeeze block which mimics the weighted median filtering in [14]. The squeeze block relies on an attention layer named PAConv to compute attention weights $\{w_n^{(\ell)}\}$ that indicate the importance of each scale within a given pixel. The squeezed depth $\boldsymbol{x}$ is obtained by selecting one scale for each pixel as follows

$$x_n = d_n^{(\ell')}, \quad \ell' = \underset{\ell \in \{1,\cdots,L\}}{\operatorname{argmax}} \; w_n^{(\ell)}, \tag{9}$$

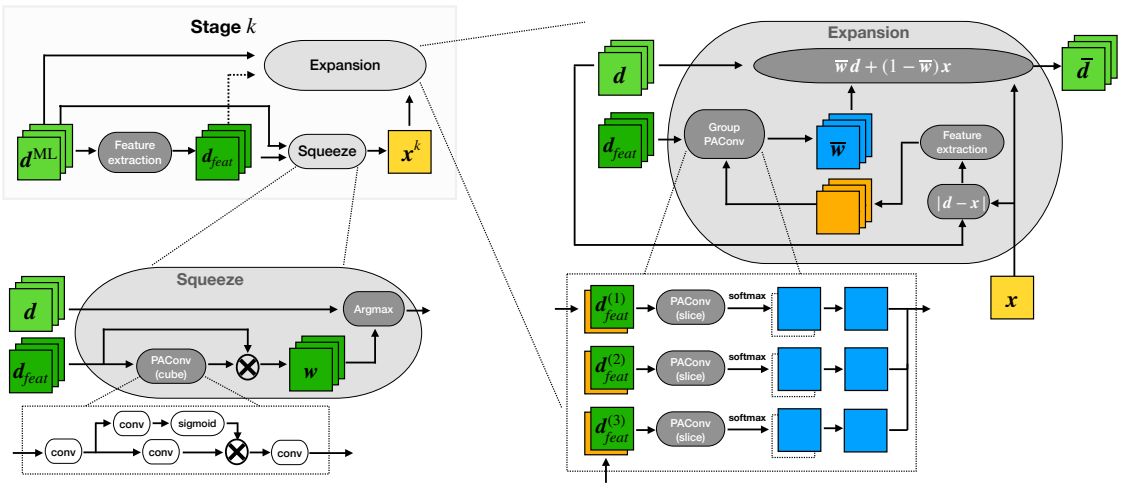


Fig. 1. The proposed network architecture for one stage k when $L = 3$. Each stage consists of three main blocks: feature extraction, squeeze block and expansion block. All the feature extraction layers consist of three convolution layers.

Expansion block. The squeezed depth $\boldsymbol{x}$, the multiscale depths $\boldsymbol{d}$ and its features are fed into the expansion block. This block corresponds to the generalized shrinkage operator in [14] and its goal is to refine the multiscale depths. To emphasize the relative difference, the block computes $|\boldsymbol{d}^{(\ell)} - \boldsymbol{x}|, \forall \ell$ whose features are used to compute another attention weights. Unlike the squeeze block, the expansion block computes attention weights slice by slice, normalizing between 0 and 1. These normalized weights $\overline{\boldsymbol{w}}$ are used to compute the refined multiscale depths as $\overline{\boldsymbol{d}}^{(\ell)} = \overline{\boldsymbol{w}}^{(\ell)}\boldsymbol{d}^{(\ell)} + (1 - \overline{\boldsymbol{w}}^{(\ell)})\boldsymbol{x}$.

The refined multiscale depths $\overline{\boldsymbol{d}}^{(\ell)}, \forall \ell$ are again used as an input of the next stage. The last stage considers the squeezed depth as the final estimated depth and has no expansion block.

B. Computing the uncertainty variable

The uncertainty variable in Eq. (7) has a closed-form solution for its MAP estimator [14]. Motivated by the analytical solution, we adopt the following formula to compute the uncertainty of our estimated depth map:

$$\epsilon_n = \frac{1}{K-1}\sum_{k=1}^{K-1}\frac{\mathcal{C}_n^k + \beta}{L+2+\alpha},\ \mathcal{C}_n^k = \sum_{\ell=1}^{L}\overline{\overline{w}}_n^{k,(\ell)}|d_n^{k,(\ell)} - x_n^K|, \quad (10)$$

where k indicates the stage, $d_n^{k,(\ell)}$ is the multiscale depth map, x_n^K is the estimated depth and $\overline{\overline{\boldsymbol{w}}}_n^{k,(:)}$ is the normalized values of $1 - \overline{\boldsymbol{w}}_n^k$ by softmax with respect to scales, where $\overline{\boldsymbol{w}}^k$ is the attention weights in the expansion block.

C. Constructing initial multiscale depths

Assuming background photons are removed, the Lidar data is cross-correlated with the system IRF g, obtaining $\overline{\boldsymbol{Y}}$. It is downsampled spatially by 4 uniform filters denoted by $F = \{m_1 \times m_1, m_2 \times m_2, m_3 \times m_3, m_4 \times m_4\}$ where we use $m_1 = 1$, $m_2 = 3$, $m_3 = 7$, $m_4 = 13$, unless explicitly mentioned. This results in 4 downsampled histograms $\boldsymbol{Y}^{(1,2,3,4)}$. To deal with high noisy data, we consider 8 additional scales, so that in total $L = 12$. The cross-correlated histogram $\overline{\boldsymbol{Y}}$ is downsampled by the 3D uniform filters of size $m_3 \times m_3 \times m_3$ and $m_4 \times m_4 \times m_4$, obtaining two additional histograms. On each of two histograms, the spatial uniform filter F is applied, obtaining 8 downsampled histograms $\boldsymbol{Y}^{(5,6,7,8,9,10,11,12)}$. For each downsampled histogram, we take the argmax operator to obtain $\boldsymbol{d}^{\mathrm{ML}(1,2,\ldots,12)}$, which is the input of the network.

D. Background removal

In Section II, the observation model is derived without considering background photons. Here, we introduce a strategy to remove background photons which are assumed to be spatially homogeneous for effiency. With this assumption, the downsampled histogram $y_{n,t}^{(L)}$ with the largest scale L can capture the structures of background photons. This downsampled histogram is used to approximate the temporal shape of obscurants as

$$\hat{b}_t^{\mathrm{shape}} = \mathrm{median}\left(y_{\sqcap_n,t}^{(L)}\right) \quad (11)$$

where $\sqcap_n$ denotes the pixel indices of the lowest 20% values of $y_{:,t}^{(L)}$ given t, assuming those pixels contain only backgrounds without overlapping to signals. This assumption holds unless the target is a flat plane with the same depth values. With few signal peaks per pixel, the background levels per pixel can be approximated as median over time bins as follows [16]:

$$\hat{b}_n^0 = \mathrm{median}\left(y_{n,1}^{(L)}, y_{n,2}^{(L)}, \cdots, y_{n,T}^{(L)}\right). \quad (12)$$

Then, the background photons can be considered as:

$$\hat{b}_{n,t} = \hat{b}_n^0 + \hat{b}_t^{\mathrm{shape}} - \mathrm{mean}(\{\hat{b}_t^{\mathrm{shape}}\}). \quad (13)$$

To avoid underestimating the background, we further add the standard deviation of $\hat{b}_{n,t}$ which is assumed to follow Poisson distribution, to obtain our background estimation as:

$$b_{n,t} = \hat{b}_{n,t} + \eta\frac{\mathrm{std}(\hat{\boldsymbol{b}})}{\mathrm{mean}(\hat{\boldsymbol{b}})}\sqrt{\hat{b}_{n,t}}, \quad (14)$$

where $\eta > 0$ is a hyerparameter, $\hat{\boldsymbol{b}} = \{\hat{b}_{n,t}\}$ and $\mathrm{std}(\cdot)/\mathrm{mean}(\cdot)$ is known as the normalized standard deviation. From the estimated background counts $b_{n,t}$, the histogram data are subtracted by $b_{n,t}$, obtaining the signal counts (whose negative values are set to zeros).

E. Training procedures

For training data, we simulate Lidar measurements based on the Poisson model (1) and choose 30 scenes from [17], [18] for depth and reflectivity images. We generated data with four different scenarios of average Photons-Per-Pixel (PPP) and Signal-Background-Ratio (SBR): (PPP=1,SBR=1), (64,1), (1,64), (64,64). Due to memory requirements, we extract patches of size 256×256 with stride 48 for initial multiscale depths. We employ the ℓ_1-loss between the squeezed depths x^k with $k = 1, ..., K$ and the ground truth depth map. Our network is trained using the ADAM optimizer [19], 200 epochs and the learning rate 0.0001 which is decreased by half at the 100th epoch.

V. Experimental results

A. Synthetic data

We evaluate the proposed method on simulated data with non-uniform background. Lidar data is simulated from the 555×695 depth and intensity images of the Art scene from [17]. As shown in Fig. 2, we impose both uniform and non-uniform background photons with different levels of PPP and SBR. To generate a non-uniform background, we use a gamma-shaped function $t^{1.2}e^{-0.02t}$. In this experiment, the hyperparameter η in Eq. (14) is set to 0.1.

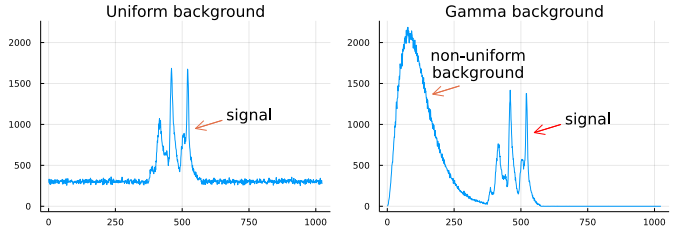


Fig. 2. Simulated histogram summed over all pixels when PPP= 1 and SBR= 0.25 with uniform (left) and gamma-shaped (right) background.

We compare our method to the classical matched filter (Classic), the underlying Bayesian method [14] and the state-of-the-art deep learning method (Peng) [13]. Fig. 3 shows examples of estimated point clouds under non-uniform background noise, which highlight the robustness of the proposed method. The last column shows the estimated depth uncertainty map by the proposed algorithm. This map is a distinct advantage of our method, compared to the previous learning-based methods [11], [13]. Fig. 4 shows the Depth Absolute Error (DAE) of different algorithms, i.e., the average absolute difference between the ground truth and estimates. The proposed method is more robust than [14], and shows better performance than Peng's method for non-uniform background, when PPP is larger than 16 and SBR is low.

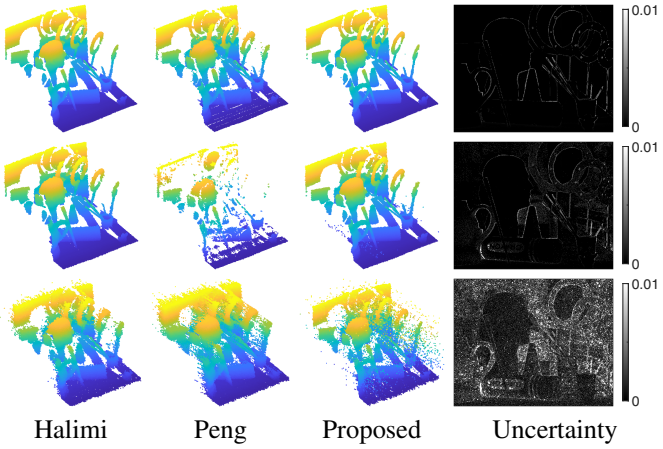


Fig. 3. Reconstructed depth maps from data with non-uniform background when PPP=64, SBR=4 (Top), PPP=64, SBR=0.25 (Middle) and PPP=1, SBR=1 (Bottom). The last column shows the uncertainty maps by the proposed method.

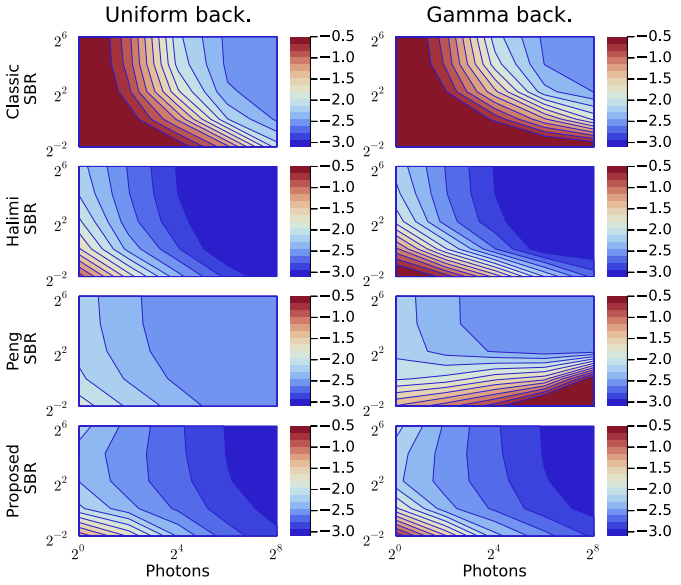
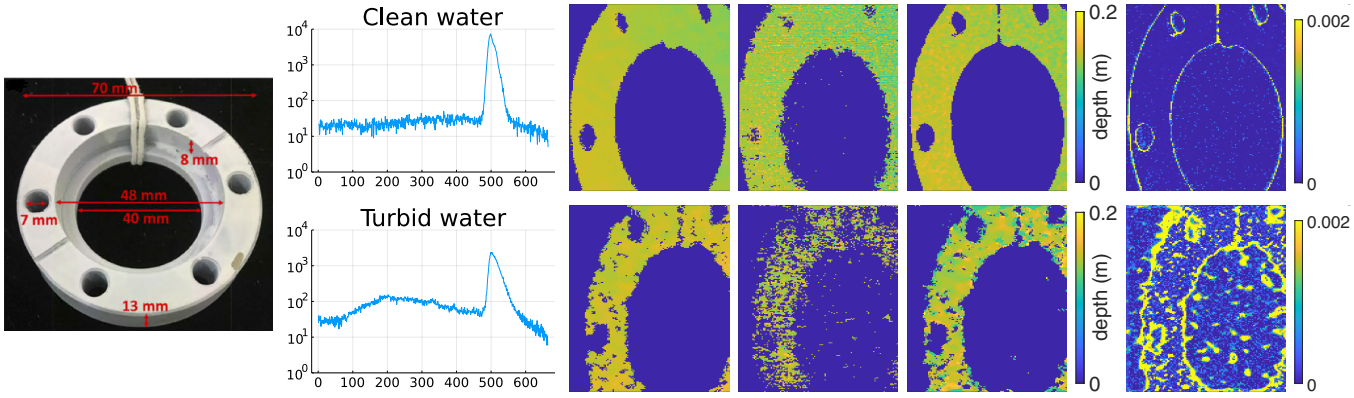


Fig. 4. Depth Absolute Error (DAE) in log scale for different levels of SBR and PPP with uniform (left) and gamma-shaped background (right). From top to bottom: Classical, Halimi [14], Peng [13] and proposed algorithms.

B. Real underwater data

We evaluate the proposed method on a moving metal target underwater, shown in Fig. 5 (a). The target is placed in a water tank at a distance of 1.7 metres from the sensor. To test varying scattering levels, water is combined with different quantities of a commercial antacid drug called Mallox, resulting in turbid water. As turbid water can scatter incoming illumination, it leads to a non-uniform background in the histogram data observed in Fig. 5 (b). The size of the Lidar data cube is 128×192×701.

The proposed method is used with smaller filters when generating initial multiscale depths, setting $m_1 = 1$, $m_2 = 3$, $m_3 = 5$, $m_4 = 7$, and we set $\eta = 0.04$. As shown in Fig. 5 (c-e), in clean water, Peng's method did not preserve the circular



(a) Target (b) Histogram (c) Halimi (d) Peng (e) Proposed (f) Uncertainty

Fig. 5. Underwater imaging experiment. (a) picture of the target. (b) The summed histogram over all pixels from data under clean water (top) and turbid water (bottom). The peaks on the right correspond to the target. (c-e) Reconstructed depth maps by Halimi [14], Peng [13] and the proposed method. (f) Uncertainty map by the proposed method.

holes, while Halimi's and proposed methods perform well. In turbid water, Peng's reconstruction result could not capture the target well leading to many disconnected parts, while the underlying Bayesian method [14] and our method yield better results. Fig. 5 (f) illustrates the uncertainty map by our method, showing higher uncertainty values in the case of turbid water and around the object's edges. We also provide the running time and the number of parameters of the networks in the below table. The proposed method is faster than other methods on GPU and our network requires an order of magnitude less parameters than Peng's network [13].

	Halimi	Peng	Proposed
Runtime in CPU	6.33 sec		8.41 sec
Runtime in GPU		9.78 sec	1.42 sec
Parameters		568,298	53,136

VI. Conclusions

This paper proposes a method to reconstruct depth maps from single-photon Lidar data with non-uniform background photons. Through the connection to the underlying Bayesian method, it estimates accurate depth information in challenging conditions and provides uncertainty estimation, an advantage compared to previous deep learning methods for single-photon Lidar reconstruction. In future, the reflectivity information will be incorporated into the network as an additional input.

References

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

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

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

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

[5] A. Maccarone, F. Mattioli Della Rocca, A. McCarthy, R. Henderson, and G. S. Buller, "Three-dimensional imaging of stationary and moving targets in turbid underwater environments using a single-photon detector array," *Opt. Express*, vol. 27, no. 20, p. 28437, 2019.

[6] A. Halimi, A. Maccarone, A. McCarthy, S. McLaughlin, and G. S. Buller, "Object depth profile and reflectivity restoration from sparse single-photon data acquired in underwater environments," *IEEE Trans. Comput. Imaging*, vol. 3, no. 3, pp. 472–484, 2017.

[7] R. Tobin, A. Halimi, A. McCarthy, P. J. Soan, and G. S. Buller, "Robust real-time 3D imaging of moving scenes through atmospheric obscurant using single-photon LiDAR," *Sci Rep*, vol. 11, no. 1, p. 11236, 2021.

[8] D. Shin, A. Kirmani, V. K. Goyal, and J. H. Shapiro, "Photon-Efficient Computational 3-D and Reflectivity Imaging With Single-Photon Detectors," *IEEE Trans. Comput. Imaging*, vol. 1, no. 2, pp. 112–125, 2015.

[9] 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, 2017.

[10] A. Halimi, R. Tobin, A. McCarthy, J. Bioucas-Dias, S. McLaughlin, and G. S. Buller, "Robust Restoration of Sparse Multidimensional Single-Photon LiDAR Images," *IEEE Trans. Comput. Imaging*, vol. 6, pp. 138–152, 2020.

[11] D. B. Lindell, M. O'Toole, and G. Wetzstein, "Single-photon 3D imaging with deep sensor fusion," *ACM Trans. Graph.*, vol. 37, no. 4, 2018.

[12] A. Ruget, S. McLaughlin, R. K. Henderson, I. Gyongy, A. Halimi, and J. Leach, "Robust super-resolution depth imaging via a multi-feature fusion deep network," *Opt. Express*, vol. 29, no. 8, p. 11917, 2021.

[13] J. Peng, Z. Xiong, X. Huang, Z.-P. Li, D. Liu, and F. Xu, "Photon-Efficient 3D Imaging with A Non-local Neural Network," in *European Conference on Computer Vision (ECCV)*, 2020.

[14] A. Halimi, A. Maccarone, R. 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. on Comput. Imaging*, vol. 7, pp. 961–974, 2021.

[15] J. Koo, A. Halimi, and S. McLaughlin, "A Bayesian based deep unrolling algorithm for single-photon Lidar systems," *IEEE Journal of Selected Topics in Signal Processing*, 2022. doi.org/10.1109/JSTSP.2022.3170228

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

[17] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.

[18] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, "A naturalistic open source movie for optical flow evaluation," in *European Conference on Computer Vision (ECCV)*, 2012.

[19] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *International Conference on Learning Representations (ICLR)*, 2015.