





Fast and Robust Restoration of Single Photon 3D Data Using Parameterized Kernel

Songmao Chen , Xiuqin Su , Zhenyang Zhang , *Graduate Student Member, IEEE*, Weihao Xu, Jie Wang, and Wei Hao 

Abstract—Single photon 3D imaging is an emerging topic for optronic sensing under extreme scenarios (e.g. spaceborne altimeter, long range imaging). However, such technique suffers from low photon counts and strong noise, which is due to either strong attenuation from the environment or reduced acquisition time. Although state-of-the-art algorithms have been proposed to achieve high resolution results from corrupted single photon 3D data, the trade-off between the restoration performance and computational complexity remains challenging. This paper presents a fast and robust restoration approach for single photon 3D data, which adaptively smooth the sparse and noisy histogram by applying a parameterized kernel and finally reconstruct the 3D image using matched filter. The implementation can be fast as the core step of the processing is generalized as a 3D convolution that can be solved by Fast Fourier Transform (FFT). The method is validated on various conditions and scenarios from Middbury dataset and real data, where the proposed method showed robust results as the competing state-of-the-art algorithms with fast implementation.

Index Terms—Single photon imaging, 3D restoration, spatial correlation, sparse signal, noisy data, computational imaging.

I. INTRODUCTION

ACTIVE 3 dimensional (3D) laser imaging is receiving growing interest from photonics community due to its rich information and broad range of applications in autonomous vehicles [1], [2], terrain mapping [3], spaceborne altimeter [4] and

non-line-of-sight imaging [5], [6]. Among the progresses, single photon 3D imaging LiDAR (Light Detection and Ranging) has been considered as a promising candidate as it offers single photon sensitivity, picosecond timing resolution and high photon efficiency. Compared to conventional LiDAR, single photon sensitivity and high photon efficiency boosted the measuring range and picosecond timing time-of-flight information brings excellent depth resolution, which indicates that accurate 3D information could be captured at kilometer range. Recent research showed that such technique could acquire high resolution 3D information up to 200 km [7], as well as highly turbid media [8], [9] and atmospheric obscurant [2], [10] scenarios.

By employing laser with picosecond duration, single photon LiDAR operates by directing abundant pulses to the target within a user defined time (known as ‘acquisition time’ [11]) and collecting the scattered photons to single photon detector. A discrete waveform (i.e., histogram) is then formed via the time-of-arrival of multiple collected photons using picosecond timing electronics, where the amplitude of peak indicates the reflectivity and the time delay infers the time-of-flight, and hence target depth. Finally, the 3D information of the target is obtained by repeating the above measurement for each pixel. In practice, the histogram can be sparse due to either long range imaging [12] or limited acquisition time [13], at the same time, due to the back scattering effects from media or the background light, the histogram can be noisy as well [14], [15]. Although better hardware has been reported, some of the problems are not suitable to be solved by hardware alone [16].

To tackle these challenges, various restoration algorithms have been proposed. The algorithms can be roughly classified into 2 classes: statistical based algorithms and deep learning based algorithms. Statistical based algorithms aim to design a model that accounts for the data statistics and the prior information, and then estimating the parameters of interest using optimization [17], [18] or stochastic simulation methods [14], [19]. Statistical based algorithms usually present good interpretability and fast implementation as the solution can be analytical, however, such methods are subject to tuning of parameters and handcraft designed prior and model, which is considered as a bottleneck of the family. Deep learning based algorithms restore the corrupted data using a well-trained network with tailored datasets [20]. Compared to statistical based algorithms, deep learning based algorithms showed powerful restoration performance, especially in the extreme cases [21], [22]. However, as only limited data can be acquired in the real world, deep

Manuscript received 18 November 2022; revised 31 March 2023; accepted 18 April 2023. Date of publication 25 April 2023; date of current version 15 December 2023. This work was supported in part by the China Postdoctoral Science Foundation under Grant 2020M683600 and in part by the Youth Innovation Promotion Association XIOPM-CAS under Grant XIOPMQCH2021008. (Corresponding author: Wei Hao.)

Songmao Chen, Xiuqin Su, and Wei Hao are with the Key Laboratory of Space Precision Measurement Technology, Xi’an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi’an 710119, China, and also with the Pilot National Laboratory for Marine Science and Technology, Qingdao 266200, China (e-mail: chensongmao@opt.ac.cn; suxiuqin@opt.ac.cn; hwei@opt.ac.cn).

Zhenyang Zhang and Jie Wang are with the Key Laboratory of Space Precision Measurement Technology, Xi’an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi’an 710119, China, also with the University of Chinese Academy of Sciences, Beijing 100049, China, and also with the Pilot National Laboratory for Marine Science and Technology, Qingdao 266200, China (e-mail: zhangzhenyang2017@opt.ac.cn; wangjie20@opt.ac.cn).

Weihao Xu is with the Key Laboratory of Space Precision Measurement Technology, Xi’an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi’an 710119, China, and also with the University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: xuweihao@opt.ac.cn).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/JSTQE.2023.3269747>.

Digital Object Identifier 10.1109/JSTQE.2023.3269747

learning based algorithms usually train the network on simulated data with preset hyper-parameters, which might lead to the phenomenon named domain shift when applied to measured data due to the inaccurate estimation of hyper-parameters [21], [23]. For another, training the network is computational-intensive, which consumes a huge amount of computing resources and time [22], [24].

The dilemma of the approaches lies in the fact of the contradiction of conciseness and preciseness of the model (or network). Several attempts have been reported for developing physical inspired network, which seems to be the breakthrough. Koo et al. unrolls a Bayesian inference into a new deep learning architecture, the results showed such approach could reduce the number of training parameters and is more robust to noise and mismodelling effects. Nonetheless, the method still requires tens of thousands parameters and takes hours of training [24].

This paper presents a new method that first localizes signals distributions and then performs 3D restoration with a Parameterized Convolution Kernel (PICK-3D). The kernel accounts for local spatial correlations and data driven, leading to a fast and robust performance.

In summary, the contributions of this paper are:

- Proposed a parameterized kernel associated with physically informed parameters, which can efficiently restore the corrupted data with convolution operator and data driven feature parameters.
- State-of-the-art performance is achieved with only a handful of parameters and limited computation cost, supporting on-chip implementation and other applications (e.g., super-resolution) in the future.
- Different filtering strategies are adaptively assigned for different cases, where a cascade strategy is applied for extreme cases and a selective filter for mild cases to boost the performance.

The remainder of the paper is organized as follows. Section II describes the assumptions and observation model for single-photon 3D data. Section III introduces the details of proposed method. Section IV evaluates and analyze the performance of proposed method with evaluation metrics. Conclusions and future work are discussed in Section V.

II. OBSERVATION MODEL

This section introduces the model of the recorded histogram associated with parameters of interest. In this paper, the observation model is built on following common assumptions: 1) there is exactly only one surface is present in each pixel, i.e., all the histograms contain a single peak; 2) the target is approximately static during the measurement; 3) the observation window covers all the depths and the depth distribution is concentrated in a smaller window. Let N be the number of pixels and T be the number of time bins, \mathbf{Y} of size $N \times T$ be the histogram and $y_{n,t} \in \mathbf{Y}$ be the number of photon counts in the t th bin of n th pixel, the observation is commonly assumed to be drawn from Poisson distribution [1], [18] as follows:

$$y_{n,t} \sim \mathcal{P}(s_{n,t}), \quad (1)$$

where \mathcal{P} stands for Poisson distribution and $s_{n,t}$ is given as follows:

$$s_{n,t} = r_n \mathbf{f}(t - t_n) + b_t. \quad (2)$$

In (2), $r_n > 0$ and $t_n > 0$ represents the reflectivity and the time delay of the n th pixel respectively. Depth d_n is given by $d_n = (C \times t_n)/2$, C is the speed of light. \mathbf{f} is the Impulse Response Function (IRF), which can be obtained in the calibration step. b_t is the noisy counts in t th bin, related research indicates the distribution of noise is approximately uniform in a range of scenarios, the proposed approach adopt this assumption and set $b_t = b$ for all bins [1], [17]. The goal of this paper is to estimate \mathbf{R} that gathers all r_n and \mathbf{D} that gathers all d_n from corrupted histogram \mathbf{Y} .

III. PROPOSED METHOD

This section introduces the details of the proposed PICK-3D algorithm. As shown in Fig. 1, the proposed method first estimates the feature parameters from data, then intercepts the signal intervals along the t axis. After that, the PICK-3D algorithm initializes the kernel and improve the performance by adopting a cascade filtering strategy for the cases with extremely low photon counts or strong noise (named as “extreme cases”), and a selectively filtering strategy for the cases with relatively high photon counts or mild noise (named as “mild cases”). For ordinary cases, PICK-3D algorithm applies the kernel directly to get smoothed histogram. The estimated \mathbf{D} and \mathbf{R} is finally given by the matched filter.

A. Parameter Estimation

The selection and estimation of feature parameters are essential for improving performance and data driven scheme. Here, 4 parameters are considered for further processing: photon-per-pixel (PPP) Θ , signal-to-background ratio (SBR) Φ , the size of the filtering kernel δ and the full width half maximum (FWHM) of IRF τ . As we assume uniform distribution of b (i.e. $b_t = b$), the estimation of b is given by the following equation

$$\hat{b} = \frac{\sum_{t=1}^{T_b} \sum_{n=1}^N y_{n,t}}{N \times T_b}, \quad (3)$$

where \hat{b} is the estimation of b and T_b is a user defined range for \hat{b} estimation. PPP and SBR measures the sparseness of signal and noise level respectively, which can be obtained using (4).

$$\Theta = \frac{\sum_{t=1}^T \sum_{n=1}^N y_{n,t}}{N} - (\hat{b} \times T), \quad \Phi = \frac{\Theta}{(\hat{b} \times T)}. \quad (4)$$

The FWHM of IRF is measured by accumulating enough photons, which shape depends on the jitter of the system and the timing resolution of the Time Correlated Single Photon Counting (TCSPC) module. Empirically, τ is a small value that correspond to a narrow peak, however, high timing resolution can lead to a large τ , which result in inaccurate estimation of key parameters. Therefore, τ is corrected by $\tau = \lceil \Gamma \log_{10} \tau \rceil$ when τ is large, where Γ is the FWHM of a reference IRF and $\lceil \cdot \rceil$ is the round down operator. In this paper, Γ is assumed to be 7 for all cases.

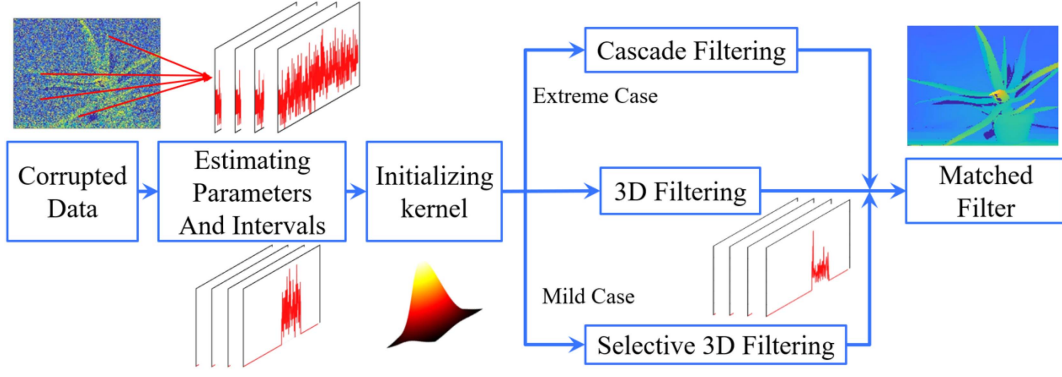


Fig. 1. Overview of the proposed method.

Recent advances suggested that \mathbf{Y} is sparse regarding to signal photons [25], leading to inefficient processing or manually interception as pre-processing. Akin to [26], under the assumptions in Section II, the range of the signal intervals could be adaptively given by employing Adaptive Gating based Noise Reduction (AGNR) algorithm. The AGNR algorithm first summing all the histograms in the spatial domain (i.e. $\sum_{n=1}^N y_{n,t}$) and then iteratively estimating the signal intervals by analysing the summed histogram and the model. In this paper, AGNR is applied to \mathbf{Y} to form a dense matrix \mathbf{Y}^g of size $N \times T_g$, where $T_g < T$ is the width of the interval. As the rest of the processing are based on \mathbf{Y}^g , Θ and Φ is updated as Θ^g and Φ^g accordingly using (4).

δ determines the number of involved pixels in filtering, a larger δ indicates a stronger denoising performance but also reduced image details [18]. Although multi-scale information are suggested to promote smoothness of the signal [18], [24], PICK-3D only select the best δ for filtering as it is more computational efficient, unless cascade filtering strategy is used (see Section III-C for details). Intuitively, δ should be inverse proportional to Θ and Φ , and proportional to τ as a smaller τ implies a concentrated signal distribution [17]. Considering a $\delta \times \delta$ kernel, by denoting ‘max’ as the maximize operator and $\lceil \cdot \rceil$ as the round up operator, δ is empirically given as follows

$$\delta = \left\lceil \sqrt{\max \left[\left(\frac{2 \times \tau}{\Phi^g} \right), \left(\frac{2 \times \tau}{\Theta^g} \right) \right]} \right\rceil. \quad (5)$$

Note that (5) plays a role of initial estimation that usually close to the optimal choice, better performance might be achieved by fine-tuning δ manually (see the discussion in Section IV-C).

B. Kernel

The structure of the kernel is the key factor for the restoration performance. Classical choice is the uniform kernel that considers the neighbours share the same information [17]. Uniform kernel suppresses the noise by “borrowing” photons from neighbours, however, the assumption of neighbours share the same information degrades the image details at the same time. It can be proved that the spectrum of the uniform kernel is a low-pass filter, such property leads to a multi-scale approach for better performance [18].

The proposed PICK-3D algorithm aims to model a finer spatial correlations with computational efficient form. Although different models can be chosen, a large number of research suggested that 2D Gaussian model fits for depicting the spatial correlations between neighbour pixels [27]. Filtering with a Gaussian based kernel equivalent to summing the histogram with Gaussian based weights, where higher weights are assigned to closer neighbours. Furthermore, the kernel considers a constant Φ^g that accounts for the uniform distributed noise. The kernel is expressed as follows

$$\mathcal{G} = \frac{1}{\sqrt{2\pi}\sigma} \left[\exp \left(\frac{-\Omega^2}{2\sigma^2} \right) + \Phi^g \right], \quad (6)$$

where Ω and σ^2 are the weights and variance of the 2D surface, Φ^g is vectorized Φ^g of size $\delta \times \delta$. By defining $\Omega_{(i,j)}$ to be the index of the kernel and considering a stronger spatial correlation for sparse case, Ω and σ^2 are given by

$$\Omega = \left\{ \Omega_{(i,j)} - \left\lfloor \frac{\delta}{2} \right\rfloor \right\}_{(i,j) \in [1,\delta]}, \quad \sigma = \frac{\tau}{2 \times \Theta^g}. \quad (7)$$

Finally, \mathcal{G} is normalized for further processing using (8).

$$\mathcal{G} = \frac{\mathcal{G}}{\sum_{i=1}^{\delta} \sum_{j=1}^{\delta} \mathcal{G}}. \quad (8)$$

C. Filtering Strategy

Once the kernel and the cropped histogram are obtained, a straightforward choice is to apply the kernel \mathcal{G} on \mathbf{Y}^g directly to smooth the histogram, and hence smoothed \mathbf{D} and \mathbf{R} . Specifically, let \mathbf{Y}^s and \otimes be the smoothed histogram and cross correlation operator respectively, \mathbf{D} and \mathbf{R} is given by (9) and (10), where (10) is also known as matched filter.

$$\mathbf{Y}^s = \mathcal{G} \otimes \mathbf{Y}^g, \quad (9)$$

$$[(D_i, R_i) \leftarrow \arg \max_{\mathbf{D}, \mathbf{R}} (\mathbf{Y}_i^s \otimes \mathbf{f})]_{\forall i \in [1, \dots, N]}. \quad (10)$$

The ‘max’ operator in (10) returns the value and its index of the histogram.

However, for extreme cases, a single filter might failed to show satisfied results. Inspired by the idea of mutli-scale processing

Algorithm 1: PICK-3D.

```

1: Input:  $\mathbf{Y}$ ,  $\mathbf{f}$ ,  $\tau$ ,  $\Gamma$ 
2: Initialize:  $\Theta$  and  $\Phi$  using (4).
    $\delta$  using (5).
3:  $\mathbf{Y}^g \leftarrow \text{AGNR}(\mathbf{Y})$  [26]
4: Update  $\Theta^g$  and  $\Phi^g$  using  $\mathbf{Y}^g$  and (4)
5: Initialize  $\Omega$  and  $\sigma$  using (7)
6:  $\mathcal{G} = \frac{1}{\sqrt{2\pi\sigma}} [\exp(\frac{-\Omega^2}{2\sigma^2}) + \Phi^g]$ 
7:  $\mathcal{G} = \frac{\mathcal{G}}{\sum_{i=1}^{\delta} \sum_{j=1}^{\delta} \mathcal{G}}$ 
8: if  $2 < \delta < 3\tau$  then
9:    $\mathbf{Y}^s = \mathcal{G} \otimes \mathbf{Y}^g$ 
10: else
11:   obtain  $\hat{\mathbf{Y}}^g$  using (11)–(13).
12:   if  $\delta \geq 3\tau$  then
13:      $\mathbf{Y}^s = \mathcal{G} \otimes \hat{\mathbf{Y}}^g$ .
14:   else
15:      $\mathbf{Y}^s = \hat{\mathbf{Y}}^g$ 
16:   end if
17: end if
18: for  $i = 1 : N$  do
19:    $\{(D_i, R_i) \leftarrow [\max(\mathbf{Y}_i^s \otimes \mathbf{f})]\}$ 
20: end for
21: Output:  $\mathbf{D}$ ,  $\mathbf{R}$ 

```

that already showed powerful performance, PICK-3D algorithm adopts a cascade processing strategy that detecting the corrupted pixels with a threshold and replacing the corrupted histogram with the filter smoothed histogram before applying (9) and (10). As the corrupted pixels usually show low photon counts and small SBR, the set of all corrupted pixels Λ is given by

$$\Lambda \leftarrow \left[\mathbf{Y}^g < \left(\rho \times \hat{b} \times T_g \right) \right], \quad (11)$$

where ρ is a hyper-parameter. After that, only the corrupted histograms are filtered

$$\{\hat{\mathbf{Y}}_i^s\}_{i \in \Lambda} = \mathcal{G} \otimes \{\mathbf{Y}_i^s\}_{i \in \Lambda}, \quad (12)$$

and the updated $\hat{\mathbf{Y}}^g$ is formed by replacing the corrupted histogram using $\{\hat{\mathbf{Y}}_i^s\}_{i \in \Lambda}$.

$$\hat{\mathbf{Y}}^g = \{\hat{\mathbf{Y}}_i^s\}_{i \in \Lambda} + \{\mathbf{Y}_i^g\}_{i \notin \Lambda}. \quad (13)$$

Finally, \mathbf{D} and \mathbf{R} are obtained using (9) and (10) by replacing \mathbf{Y}^g as $\hat{\mathbf{Y}}^g$.

Similarly, for mild case, PICK-3D algorithm only process the corrupted information instead of processing the full image, and hence avoid degrading the image details. The corrupted pixels and the smoothed histograms are obtained using (11)–(13), \mathbf{D} and \mathbf{R} are given by (10) by replacing \mathbf{Y}_i^s as $\hat{\mathbf{Y}}_i^g$. In this paper, for the cases of $\delta \geq 3\tau$ are considered to use the cascade strategy and cases of $\delta \leq 2$ (i.e., δ reaches the smallest size) will only replacing the corrupted pixels. Otherwise the kernel will be applied directly using (9) and (10).

As the cross correlation can be solved efficiently using Fast Fourier Transform (FFT) which has a computational complexity

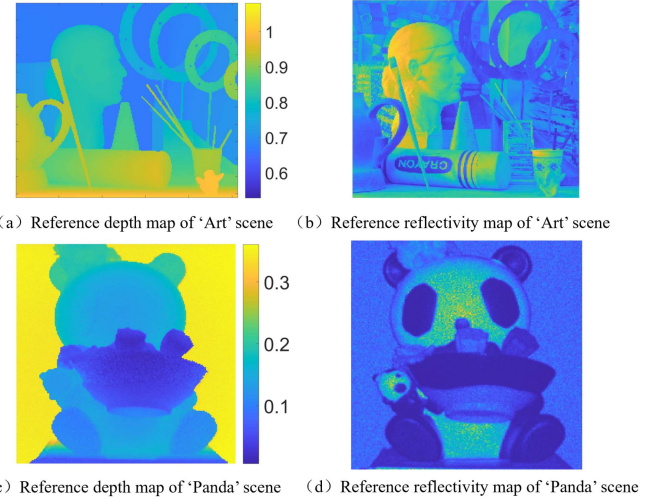


Fig. 2. Reference image of “Art” scene (first row) and “Panda” scene (second row), the depth image are shown in left column and the reflectivity image is shown in right column.

of $N \log N$, the computational complexity of PICK-3D algorithm should be $(NT_g) \log(NT_g)$ as the most computational intensive step consists of T_g times cross correlation operations, which is much lower than state-of-the-art approaches. Compared to mutli-scale approach, the proposed cascade strategy shows lower computational cost and the processing is performed in latent space (i.e. histogram) instead of images [18]. The proposed PICK-3D algorithm is summarized in Algorithm 1.

IV. EXPERIMENTAL RESULTS

This section exhibits and discuss the results of the proposed method and the state-of-the-art approaches. Quantitative analysis are discussed for various conditions to illustrate the relative advantages. Note that a work station with EXOS 3090 GPU is employed to meet the requirement of deep learning based approaches while the statistical based approaches are implemented on a laptop which has an i7-10750H CPU and 32 G RAM.

A. Dataset

All the methods are validated on both simulated data and real data. The simulated data are generated using Middlebury stereo dataset [28] and (1) and (2), where the depth information is given by the scene while different r_n and b are used to simulate different imaging conditions. Specifically, the *Art* scene of size $224 \times 256 \times 800$ ($r \times c \times T$, $r \times t = N$) and *Bowling* scene of size $276 \times 312 \times 1200$ are used for simulation, the reference image of *Art* scene is shown in Fig. 2(a) and (b). The width of the time bin and FWHM is assumed to be 16 ps and 7 respectively, other specifications are listed in Tables I and II. Particularly, the ‘Art’ scene is tested for different conditions, and the ‘Bowling’ scene is designed for studying the performance at different SBR with fixed PPP. Note that as the paper assume there is exactly only one surface is present in each pixel, the outliers in the original depth map are inpainted by neighbour pixels before generating the data.

TABLE I
QUANTITATIVE RESULTS OF ART SCENES WITH DIFFERENT CONDITIONS

PPP=1,SBR=0.05						
Methods	RMSE	RSNR		Absolute Error		IT
		Depth	Reflectivity	DAE	RAE	
Classic	0.0260	3.5568	-11.6503	0.4157	3.4635	0.27
OPN3DR [18]	0.0095	23.2356	9.9901	0.0230	0.2852	11.62
UA [17]	0.0084	26.5118	9.3197	0.0249	0.3311	18.61
BU [24]	0.0147	14.9648	-	0.0440	-	0.03 [†]
PICK-3D	0.0071	26.6387	10.6187	0.0131	0.2606	0.28
PPP=3,SBR=0.3						
Classic	0.0191	6.9200	-0.2449	0.1955	0.9325	0.28
OPN3DR [18]	0.0052	29.8927	12.2115	0.0082	0.2144	9.23
UA [17]	0.0061	30.7700	15.4639	0.0125	0.1385	11.09
BU [24]	0.0060	27.2853	-	0.0067	-	0.03 [†]
PICK-3D	0.0051	30.2527	14.0501	0.0072	0.1726	0.27
PPP=10,SBR=0.5						
Classic	0.0117	12.9447	6.5926	0.0518	0.4217	0.27
OPN3DR [18]	0.0049	31.4404	12.7750	0.0064	0.2030	9.65
UA [17]	0.0057	33.8653	17.9515	0.0092	0.1067	10.64
BU [24]	0.0040	32.4734	-	0.0037	-	0.03 [†]
PICK-3D	0.0041	33.1477	16.2985	0.0048	0.1324	0.26

The best result is marked as bold. GPU results are marked with [†].

TABLE II
QUANTITATIVE RESULTS OF BOWLING SCENES WITH DIFFERENT SBR

PPP=2,SBR=0.005						
Methods	RMSE	RSNR		Absolute Error		IT
		Depth	Reflectivity	DAE	RAE	
Classic	0.0388	0.7461	-19.3691	2.9962	7.6729	0.71
OPN3DR [18]	0.0066	28.6904	7.5329	0.0087	0.3456	23.46
UA [17]	0.0137	22.9885	12.0013	0.0570	0.2286	248.37
BU [24]	0.0143	15.1799	-	0.0432	-	0.03 [†]
PICK-3D	0.0058	31.1739	9.9596	0.0063	0.2618	1.97
PPP=2,SBR=0.05						
Classic	0.0342	1.9103	-9.5931	2.2826	2.4911	0.73
OPN3DR [18]	0.0046	33.2736	10.5588	0.0048	0.2439	19.26
UA [17]	0.0047	34.2395	16.7809	0.0061	0.1255	21.28
BU [24]	0.0057	28.7097	-	0.0045	-	0.03 [†]
PICK-3D	0.0039	35.8467	13.1680	0.0038	0.1808	0.56
PPP=2,SBR=0.2						
Classic	0.0284	3.6946	-4.1287	1.5364	1.3286	0.74
OPN3DR [18]	0.0052	33.5143	12.3390	0.0048	0.1945	24.74
UA [17]	0.0039	36.7187	20.7478	0.0046	0.0681	14.51
BU [24]	0.0054	33.2049	-	0.0026	-	0.03 [†]
PICK-3D	0.0041	35.9324	13.4515	0.0038	0.1750	0.59

The best result is marked as bold. GPU results are marked with [†].

The real data was acquired in March 2022, on Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences with dark lab condition. The employed coaxial imaging system is broadly similar to that described in [11], [29]. Specifically, the system employs a 532 nm laser with emitting power of approximately 100 μ W and repetition rate of 1 MHz, as well as an silicon based single-photon avalanche diode (Si-SPAD) detector, where the photon detection efficiency is 50% @ 532 nm. Timing resolution of 4 ps is achieved by using a PicoQuant PicoHarp 300 TCSPC module. The FWHM of IRF is measured to be 160, and τ is calibrated to be 15 by assuming $\Gamma=7$ and $\tau = \lfloor \Gamma \log_{10} \tau \rfloor$. A panda model is used as target, Fig. 2(c) and (d) visualized the reference image of the scene (named as *Panda* scene in the following analysis) that acquired by the system described above. The stand-off distance of the target is around 2 m, and the reference image is obtained by matched filter and

the data with acquisition time of 100 ms. Akin to [18], [30], the collected data allows the construction of histogram associated with reduced acquisition time. In this section, the evaluation is performed on acquisition time of 10 μ s, 100 μ s and 10 ms, and the size of the *Panda* scene is $200 \times 200 \times 1600$.

B. Comparison Algorithms and Evaluation Metric

The proposed PICK-3D algorithm is compared with state-of-the-art approaches that include both statistical based algorithms and deep learning based algorithms. Explicitly, the following methods are considered for comparison:

- *Classic*: Performing pixel-wise cross correlation between the histogram and the IRF, also known as matched filter.
- *Unmixing Algorithm(UA)* [17]: Considers a penalized maximum likelihood estimator that accounts for local spatial correlations and a tailored windowing strategy.
- *Optimization based Non-local 3D Restoration (OPN3DR)* [18]: Exploiting non-local spatial correlations as prior for restoration, which is achieved by introducing multi-scale analysis and 3D low-pass filtering.
- *Bayesian-based Unrolling(BU)* [24]: Unrolling a statistical Bayesian algorithm into a deep learning network, where the iterative updates are done by neural network layers. Note that BU only estimates depth information and was implemented on GPU. This paper uses the pre-trained model instead of retraining the model on the tested data."

By denoting \hat{D} and \hat{R} be the estimated depth and reflectivity, D and R be the reference depth and reflectivity, the results are comprehensively analyzed using the following metrics [9], [21], [31]:

- *RMSE*: Root mean squared error, which measures the difference between estimated depth image and reference depth image, defined as:

$$\text{RMSE}(D, \hat{D}) = \sqrt{\frac{\|D - \hat{D}\|_2^2}{N}}, \quad (14)$$

where $\|\cdot\|_2^2$ stands for the $L - 2$ norm.

- *RSNR*: Restoration Signal-to-Noise Ratio, which evaluate both depth and reflectivity, defined as:

$$\begin{aligned} \text{RSNR}(D, \hat{D}) &= 10 \times \log_{10} \left(\frac{\|D\|_2^2}{\|D - \hat{D}\|_2^2} \right), \\ \text{RSNR}(R, \hat{R}) &= 10 \times \log_{10} \left(\frac{\|R\|_2^2}{\|R - \hat{R}\|_2^2} \right). \end{aligned} \quad (15)$$

- *AE*: Absolute Error, which evaluate both depth (DAE) and reflectivity (RAE), defined as:

$$\begin{aligned} \text{DAE}(D, \hat{D}) &= \frac{\|D - \hat{D}\|_1}{N}, \\ \text{RAE}(R, \hat{R}) &= \frac{\|R - \hat{R}\|_1}{\|R\|_1}, \end{aligned} \quad (16)$$

where $\|\cdot\|_1$ stands for the $L - 1$ norm.

- *IT*: Implementation Time, which is given by the computer when running the algorithm.

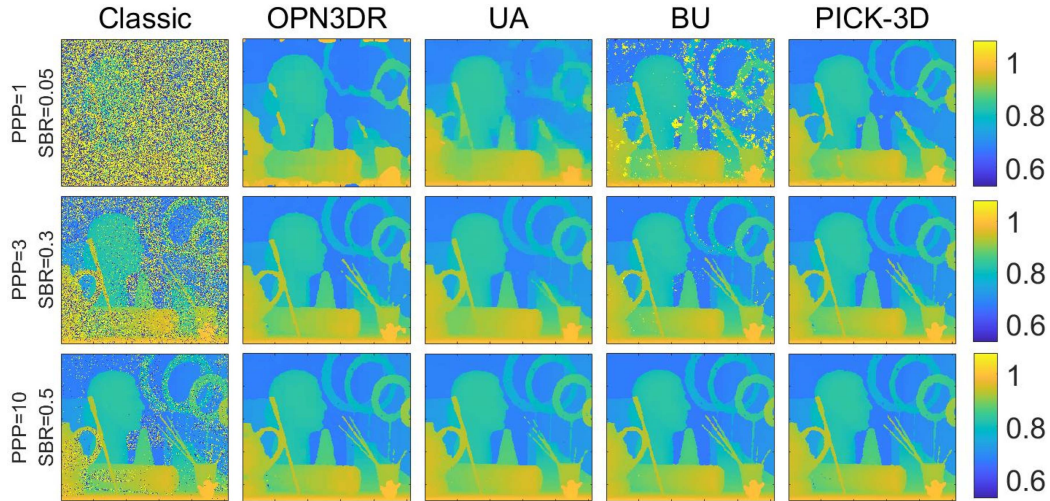


Fig. 3. Depth maps of 'Art' scene with different conditions and methods. The rows and columns corresponds to different conditions and methods respectively.

Lower RMSE and AE, or higher RSNR indicates a better restoration performance, lower IT means higher restoration efficiency. Note that RMSE and DAE is in meters, RNSR and IT are in 'dB' and seconds respectively.

C. Results and Analysis

Fig. 3 shows the depth maps with different conditions and methods on *Art* scene. It's clear that all other methods significantly improved the depth result compared to classic method, especially in the most challenging case. OPN3DR algorithm tends to show the outline of the scene, however, suffers from distortion in extreme cases. UA is robust to sparseness and noise but over smoothed the image. BU preserves the image details even in extreme cases, but also include unexpected outliers. PICK-3D recovers plenty of the details with limited distortion for the most challenging case. To compare the result quantitatively, numerical results are exhibited in Table I. Most of the results of OPN3DR are between UA and BU, and UA showed good results in RSNR, which is close to the result of PICK-3D. However, note that UA consumes much longer implementation time although it employs parallel computing tools. BU failed to show good results in RSNR, but it shows nice results in RMSE and DAE thanks to the preserved details. The proposed PICK-3D algorithm showed best results in the most challenging case and second best results for all other cases, which highlight the performance of PICK-3D. Note that BU is implemented on a work station with EXOS 3090 GPU while other algorithms are tested on an ordinary laptop and PICK-3D showed similar implementation time as classic method thanks to the algorithmic gating and low computational complexity.

Similarly, the Quantitative results of *Bowling* scene are shown in Table II. PICK-3D algorithm showed best depth results in the $SBR = 0.005$ and $SBR = 0.05$ scene, and second best for other metrics. OPN3DR showed better depth results than UA in $SBR = 0.005$ cases while UA has the best improvement on reflectivity as the scene has relatively fewer details than *Art* scene. BU wins on DAE of $SBR = 0.2$ case but lower values in other evaluation

metrics and cases, which might due to the different parameters of the scene and lead to domain shift [21]. Note that PICK-3D has longer implementation time as cascade filtering strategy is applied in $SBR = 0.005$ case. The size of the kernel that used in the experiment is $[9,4,3]^T$ and $[25(17),9,5]^T$ for *Art* and *Bowling* scene respectively (cascade filter size of 25 and 17 for the first case of *Bowling* scene). And the estimation using (5) gives $[11,5,3]^T$ and $[27,9,5]^T$, which verified the effectiveness of the equation.

For another, the *Panda* scene with different acquisition time is also validated. As the measurement was conducted in dark lab conditions, the PPP and SBR are $[0.74,6.7,673]^T$ and $[462,193,177]^T$ for $10 \mu s$, $100 \mu s$ and $10 ms$ cases. Note that the scene is challenging as the reflectivity of the scene has a huge difference (approximately 100 times) and associated with rich angles that cause strong scattering. The classic result in Fig. 4 showed the problem that the empty pixels (no received photons) are grouped (e.g. eyes, ears) instead of randomly distributed as other scenes. This phenomenon leads to some noisy pixels (on the bowl) for $10 ms$ cases although the PPP and SBR are high.

As shown in Fig. 4, OPN3DR smooth the depth maps and UA over smooth the depth for $10 \mu s$ and $100 \mu s$ cases, BU smooth the depth and keep the details at the same time. PICK-3D balances the distortion and smoothness, thus giving relatively better result. For the $10 ms$ case, all the methods showed nice results. Table III suggests that PICK-3D showed best quantitative results for $10 \mu s$ case and best results on RMSE and depth RSNR for $100 \mu s$ and $10 ms$ cases, which showed the superior performance in the challenging real world scenario. BU showed better results for high acquisition time cases, which might caused by inaccurate estimation of parameters for the low acquisition time cases. Different from the simulated scenes, It's worthy noting that the reference of *Panda* scene is from real data, which means the reference reflectivity image includes Poisson noise instead of a clean image. The restoration algorithm smooth the reflectivity and thus resulting in lower values in evaluation metrics on reflectivity for higher acquisition time cases. Particularly, PICK-3D applies selective filtering strategy

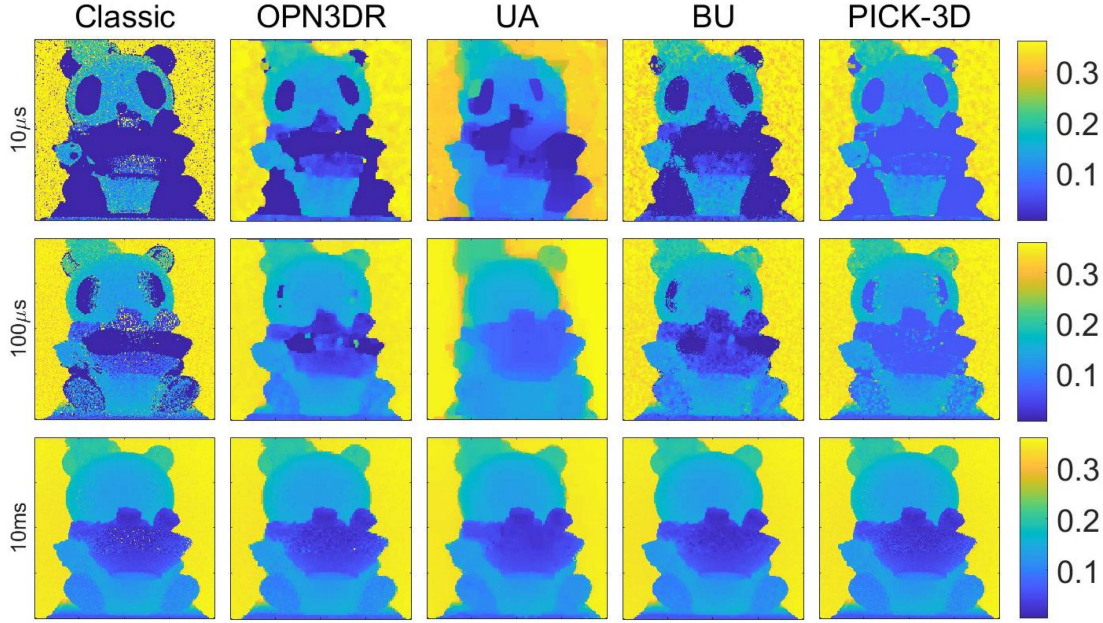


Fig. 4. Depth maps of 'Panda' scene with different acquisition time and methods. The rows and columns corresponds to different conditions and methods respectively.

TABLE III
QUANTITATIVE RESULTS OF PANDA SCENES WITH DIFFERENT ACQUISITION TIME

Acquisition time = 10 μ s						
Methods	RMSE	RSNR		Absolute Error		
		Depth	Reflectivity	DAE	RAE	IT
Classic	0.0167	5.6659	10.3521	0.0789	0.3362	1.10
OPN3DR [18]	0.0134	10.8508	10.2144	0.0343	0.3091	10.40
UA [17]	0.0125	12.2254	5.5667	0.0385	0.3804	10.62
BU [24]	0.0199	6.6220	-	0.0675	-	0.03 [†]
PICK-3D	0.0106	14.3825	11.8078	0.0270	0.2671	2.06
Acquisition time = 100 μ s						
Classic	0.0129	7.8908	25.7364	0.0419	0.0552	1.09
OPN3DR [18]	0.0109	14.6953	11.5515	0.0152	0.2270	12.29
UA [17]	0.0109	15.3943	11.5252	0.0246	0.2367	8.83
BU [24]	0.0103	15.5636	-	0.0181	-	0.03 [†]
PICK-3D	0.0079	18.1103	16.0153	0.0161	0.1366	1.18
Acquisition time = 10 ms						
Classic	0.0071	18.9201	47.3903	0.0059	0.0046	1.10
OPN3DR [18]	0.0049	23.4271	12.6422	0.0056	0.2083	3.55
UA [17]	0.0052	21.2141	13.3627	0.0067	0.1862	26.35
BU [24]	0.0042	23.4865	-	0.0052	-	0.03
PICK-3D	0.0037	26.6202	45.2848	0.0050	0.0058	1.18

The best result is marked as bold. GPU results are marked with [†].

and the numerical results verify the benefits of the proposed strategy.

V. CONCLUSION

In this paper, we have proposed a fast and robust restoration algorithm for corrupted single photon 3D data. The proposed PICK-3D algorithm takes the advantage of physically informed kernel and parameters, as well as tailored filtering strategy, leading to a faithful restoration performance with high computation efficiency. Compared to statistical based algorithm, PICK-3D algorithm conducts the analysis in latent space instead of 2D images which thoroughly utilize the desired information. For the

deep learning approaches, PICK-3D algorithm requires much less computational resources, gains higher computation efficiency and better interpretability, thus shed light on real-world practice. The future study includes the application in other challenging real-world scenarios and supporting other applications that work on photon-starved regime.

REFERENCES

- [1] 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.
- [2] 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.
- [3] J. White et al., "Evaluating the capacity of single photon lidar for terrain characterization under a range of forest conditions," *Remote Sens. Environ.*, vol. 252, 2021, Art. no. 112169.
- [4] Y. Nan, Z. Feng, B. Li, and E. Liu, "Multiscale fusion signal extraction for spaceborne photon-counting laser altimeter in complex and low signal-to-noise ratio scenarios," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 1000105.
- [5] C. Wu et al., "Non-line-of-sight imaging over 1.43 km," *Proc. Nat. Acad. Sci.*, vol. 118, no. 10, 2021, Art. no. e2024468118.
- [6] W. Xu, S. Chen, Y. Tian, D. Wang, and X. Su, "Fast non-line-of-sight imaging based on product-convolution expansions," *Opt. Lett.*, vol. 47, no. 18, pp. 4680–4683, Sep. 2022.
- [7] Z.-P. Li et al., "Single-photon imaging over 200 km," *Optica*, vol. 8, no. 3, pp. 344–349, Mar. 2021.
- [8] A. Maccarone, F. M. D. Rocca, A. McCarthy, R. Henderson, and G. S. Buller, "Three-dimensional imaging of stationary and moving targets in turbid underwater environments using a single-photon detector array," *Opt. Exp.*, vol. 27, no. 20, pp. 28437–28456, Sep. 2019.
- [9] 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.
- [10] 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, 2021, Art. no. 11236.

- [11] 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.
- [12] Z.-P. Li et al., "Single-photon computational 3D imaging at 45 km," *Photon. Res.*, vol. 8, no. 9, pp. 1532–1540, Sep. 2020.
- [13] J. Tachella et al., "Real-time 3D reconstruction from single-photon lidar data using plug-and-play point cloud denoisers," *Nat. Commun.*, vol. 10, no. 1, pp. 1–6, 2019.
- [14] 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.
- [15] W. Shu-Chao et al., "A time-correlated single photon counting signal denoising method based on elastic variational mode extraction," *ACTA Physica Sinica*, vol. 70, no. 17, pp. 159–168, 2021.
- [16] S. Chen, W. Hao, X. Su, Z. Zhang, and W. Xu, "Research progress on photon counting imaging algorithms," *Laser Optoelectron. Prog.*, vol. 58, no. 18, pp. 236–251, 2021.
- [17] R. Joshua and K. G. Vivek, "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.
- [18] S. Chen et al., "Learning non-local spatial correlations to restore sparse 3D single-photon data," *IEEE Trans. Image Process.*, vol. 29, pp. 3119–3131, 2020.
- [19] J. Tachella et al., "Bayesian 3D reconstruction of complex scenes from single-photon lidar data," *SIAM J. Imag. Sci.*, vol. 12, no. 1, pp. 521–550, 2019.
- [20] J. Peng et al., "Boosting photon-efficient image reconstruction with a unified deep neural network," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 4, pp. 4180–4197, Apr. 2023.
- [21] Y. Chen, G. Yao, Y. Liu, and Y. P. Pan, "Deep domain adversarial adaptation for photon-efficient imaging," *Phys. Rev. Appl.*, vol. 18, Nov. 2022, Art. no. 054048.
- [22] J. Peng et al., "Photon-efficient 3D imaging with a non-local neural network," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 225–241.
- [23] A. Ruget et al., "Robust super-resolution depth imaging via a multi-feature fusion deep network," *Opt. Exp.*, vol. 29, no. 8, pp. 11917–11937, Apr. 2021.
- [24] J. Koo, A. Halimi, and S. McLaughlin, "A Bayesian based deep unrolling algorithm for single-photon lidar systems," *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 4, pp. 762–774, Jun. 2022.
- [25] G. Yao et al., "Dynamic single-photon 3D imaging with a sparsity-based neural network," *Opt. Exp.*, vol. 30, no. 21, pp. 37323–37340, 2022.
- [26] S.-M. Chen et al., "Noise reduction and 3D image restoration of single photon counting LiDAR using adaptive gating," *ACTA Physica Sinica*, vol. 71, no. 10, pp. 222–233, 2022.
- [27] C. Zhao, J. Lv, S. Du, and Y. Deng, "Geometric error modeling and monitoring of the 3D surface by Gaussian correlation model," in *Proc. IEEE Int. Conf. Ind. Eng. Eng. Manage.*, 2019, pp. 1022–1025.
- [28] D. Scharstein and C. Pal, "Learning conditional random fields for stereo," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2007, pp. 1–8.
- [29] K. Hua et al., "Detection efficiency for underwater coaxial photon counting lidar," *Appl. Opt.*, vol. 59, pp. 2797–2809, 2020.
- [30] 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.
- [31] A. Halimi et al., "Restoration of intensity and depth images constructed by using sparse single-photon data," in *Proc. Eur. Signal Process. Conf.*, 2016, pp. 86–90.



Songmao Chen received the B.Eng. degree in artificial intelligence from Xidian University, Xi'an, China, in 2014, and the Ph.D. degree in signal and information processing from the University of Chinese Academy of Sciences, Beijing, China, in 2020. During 2017–2018, he was a joint training Ph.D. Student with Single-Photon Group, Heriot-Watt University, Edinburgh, U.K. He is currently an Assistant Professor with Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Science, Beijing. His research interests include single-photon imaging LiDAR, image processing, and computational imaging.



Xiuqin Su received the B.Eng. degree in radio techniques from Xi'an Jiaotong University, Xi'an, China, and the M.Eng. degree in optical instrument from the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences (CAS), Beijing, China, in 1986 and 1989 respectively. During 1998–1999, she was a Visiting Scholar with the National Research Council of Canada, Information and Technology Group. She is currently a Professor with the Xi'an Institute of Optics and Precision Mechanics, CAS. Her research interests include single-photon imaging, non-line-of-sight imaging, and lensless imaging.



Zhenyang Zhang (Graduate Student Member, IEEE) received the B.Eng. degree in automation in 2017 from the Beijing University of Technology, Beijing, China, where he is currently working toward the Ph.D. degree in signal and information processing with the University of Chinese Academy of Sciences, Beijing. His research focuses on underwater single photon imaging.



Weihao Xu received the B.Eng. degree in automation from Southeast University, Nanjing, China, in 2017. He is currently working toward the Ph.D. degree in signal and information processing with the University of Chinese Academy of Sciences, Beijing, China. His research focuses on non-line-of-sight imaging.



Jie Wang received the B.Sc. degree in information security from the Xi'an University of Posts and Telecommunications, Xi'an, China, in 2018. He is currently working toward the Ph.D. degree in signal and information processing with the University of Chinese Academy of Sciences, Beijing, China. His research focuses on underwater single photon imaging.



Wei Hao received the B.Eng. degree from Shenyang Aerospace University, Shenyang, China, in 2001, and the M.Eng. and Ph.D. degrees from the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences (CAS), Beijing, China, in 2004 and 2007 respectively. He is currently a Professor and also the Deputy Head of Xi'an Institute of Optics and Precision Mechanics, CAS. His research interests include single-photon imaging, image processing, and precision measurement technology.