Supplementary Information:

Adaptive Filtering and Sequence Estimation for High-Precision Single-Photon 3D Imaging

Zihao Pei[†, ‡, ¶, §], **Haitao Guan**[†, ‡, ¶, §], **Bowen Wang**[†, ‡, ¶, *], **Sheng Li**[†, ‡, ¶], **Min Zeng**[†, ‡, ¶], **Qian Chen**[¶], **and Chao Zuo**[†, ‡, ¶, *]

[†]Smart Computational Imaging Laboratory (SCILab), School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, Jiangsu Province 210094, China.
[‡]Smart Computational Imaging Research Institute (SCIRI) of Nanjing University of Science and Technology, Nanjing, Jiangsu Province 210019, China.
[¶]Jiangsu Key Laboratory of Spectral Imaging Intelligent Sense, Nanjing, Jiangsu Province 210094, China.
[*]wangbowen@njust.edu.cn; zuochao@njust.edu.cn
[§]They contributed equally to this work

ABSTRACT

This document provides additional information on “Adaptive Filtering and Sequence Estimation for High-Precision Single-Photon 3D Imaging”. We present an advanced single-photon imaging technique that combines photon sequence estimation with adaptive Gaussian filtering to mitigate pile-up effects.

Contents:

S1. Basic principle and limitations of Coates's estimator

In practical applications, the incident photons recorded by the TCSPC module include echo photons from the target and background photons from the environment. Specifically, the incident photon sequence r_i received in the i^{th} (i=1, 2,..., I) time bin can be expressed as

$$r_i = r_{sig,i} + r_{bkg} \tag{S1}$$

where $r_{sig,i}$ and r_{bkg} represent signal photons in the i^{th} time bin and average background noise, respectively. Meanwhile, the probability of received photons in the i^{th} time bin follows the Poisson distribution[1,2], expressed as

$$P(k) = \frac{r_i^k}{k!} e^{-r_i} \tag{S2}$$

where k represents the number of photons recorded in the i^{th} time bin. Given that a single photon can trigger an electron avalanche in the SPAD, it is crucial to determine the probability q_i that at least one photon is detected in the i^{th} time bin, which can be derived as Eq. S3.

$$q_i = 1 - P(k=0) = 1 - e^{-r_i} \tag{S3}$$

Once a photon is detected, the SPAD immediately enters dead time t_d, during which no further photons can be detected. Consequently, the probability of detecting the first photon in the i^{th} time bin depends not only on the probability q_i of a photon being present in the i^{th} time bin but also on the absence of any photon detection in the preceding n_{td} bins, where n_{td} represents the number of bins occupied by the dead time t_d. Therefore, the probability of detecting the first photon in the i^{th} time bin can be expressed as

$$p_i = q_i \prod_{j:j<i} (1 - q_j) \tag{S4}$$

In addition, the probability of failing to acquire photons during a laser period is denoted as

$$p_0 = 1 - \sum_{1}^{I} p_i = \exp(-\sum_{1}^{I} r_i) \tag{S5}$$

The likelihood function corresponding to the detected histogram can be expressed as[3]

$$\mathscr{L}(q_1, ..., q_I) = \prod_{i=1}^{I+1} q_i^{N_i} (1 - q_i)^{D_i - N_i} \tag{S6}$$

where the sequence D_i and N_i represent the total number of exposures and the total number of counted photons in the i^{th} time bin, respectively. By taking the partial derivative of Eq. S6, we obtain an estimate of q_i

$$q_i = \frac{N_i}{D_i} \tag{S7}$$

In the photon-driven mode, when the first photon is recorded in a time bin, the subsequent n_{td} time bins will be in dead time. Consequently, the total number of exposures D_i obtained for each time bin is equal to the total sustained time bins minus the number of entries into the dead time, which can be indicated as

$$D_i = \frac{T}{I\Delta} - \sum_{j=1}^{n_{td}} N_{i\ominus j} \tag{S8}$$

where $\ominus$ denotes the cyclic subtraction operation, indicating that the pulsed laser is periodic and the dead time in the previous laser period will extend into the beginning of the next period. By substituting Eq. S3 and Eq. S8 into Eq. S7, the initial incident photon sequence r_i can be obtained by Eq. S9.

$$r_i = \ln(\frac{1}{1-q_i}) = -\ln(1 - \frac{N_i}{\frac{T}{I\Delta} - \sum_{j=1}^{n_{td}} N_{i\ominus j}}) \tag{S9}$$

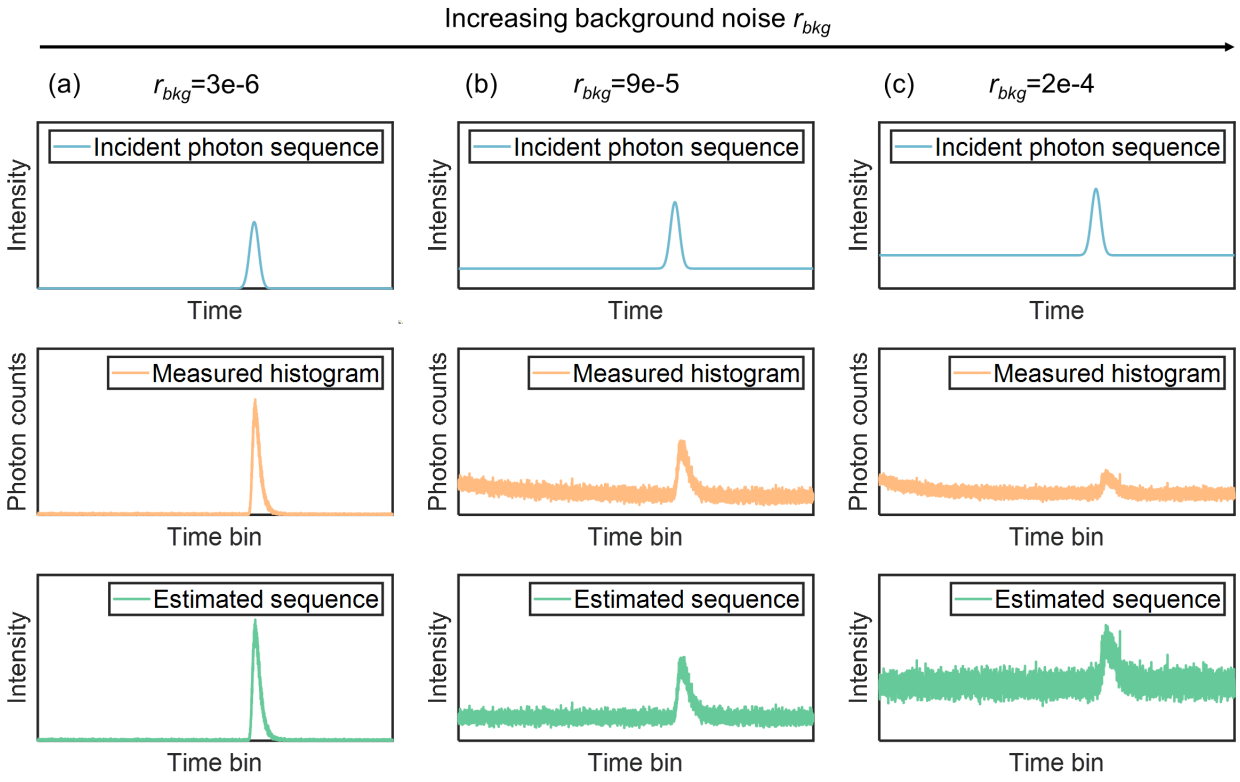

Figure S1. Effectiveness of mitigating photon pile-up via Coates's estimator under different background noise scenes. (a) In low background noise scenes, photon pile-up effects are negligible. (b) In moderate background noise scenes, photon pile-up can be observed as a characteristic exponential fall-off in the measured histogram. The sequence recovered by Coates's estimator has low peak shifts. (c) Strong background noise causes severe peak shifts in the estimated sequences, making it challenging to accurately reconstruct depth information.

Although Coates's estimator is proven effective in mitigating the photon pile-up, the estimated photon sequence suffers from severe peak shifts under strong background noise. As shown in Figure S1, in the

case of negligible background noise, all the photon arrival time line up with the location of the peak of the incident photon sequence. The estimated sequence is almost identical to the real incident photon sequence. As the background noise increases, the non-linear mapping from measurements to the estimate sequence significantly amplifies measurement noise at later time bins.[4] The incident photon sequence retrieved by Coates's estimator will have severe peak shifts compared with the ground truth. In addition, the effect of background noise on the estimated photon sequence becomes more severe with shorter detection times. The detection process is similar to rolling a dice, where the statistical results gradually stabilize as the number of rolls increases. Therefore, under the condition of sufficient detection time, the empirical frequency of stochastic events will asymptotically converge to their theoretical probability distribution. This also provides ideas on how we can use Markov chain-based adaptive Gaussian filtering algorithm to mitigate the limitations of Coates's estimator.

S2. Optimization of Gaussian filtering kernel

To mitigate the effects of background noise on the estimated photon sequence, we perform adaptive Gaussian filtering on the stationary distribution q_i^s. For stationary distributions at different scanning points, Gaussian kernel width is dynamically adjusted by the error metric to obtain the optimal filtering results. An error metric L, based on the stationary distribution q_i^s, is designed to optimize the Gaussian kernel width.

$$L = \frac{\sigma_{peak} + \sigma_{subpeak}}{E_{peak} - E_{subpeak}} \tag{S10}$$

We perform a weighted summation centered on the peak and subpeak of the stationary distribution q_i^s and calculate the standard deviation σ_{peak} and $\sigma_{subpeak}$ to assess the stability of the incident photon sequence. The difference value between the peak expectation E_{peak} and the subpeak expectation $E_{subpeak}$ is calculated to identify the peak of the incident photon sequence[5,6].

In photon-driven mode, each photon captured by the SPAD corresponds to a state transition of the Markov chain. Therefore, the number of transitions is equal to the total number of photons n_{photon}, which is the sum of the photon counts N_i across all time bin in the histogram. The expectation of the histogram after weighted summation by the Gaussian kernel a_i can be expressed as

$$E[\sum_{i=1}^{I} a_i N_i] = \sum_{i=1}^{I} a_i E[N_i] = n_{photon} \sum_{i=1}^{I} a_i q_i^s \tag{S11}$$

Meanwhile, the variance of the weighted summation results can be denoted as

$$D[\sum_{i=1}^{I} a_i N_i] = \sum_{i=1}^{I} \sum_{k=1}^{I} a_i a_k Cov(N_i, N_k) \tag{S12}$$

Each detected photon has a fixed probability of being assigned to a specific time bin, implying that the photon counts in each bin follows a mutually exclusive binomial distribution. q_b represents the probability that both time bins in the histogram simultaneously capture photons, which is equal to zero. Thus, the covariance $Cov(N_i, N_k)$ in Eq. S12 can be denoted as

$$Cov(N_i, N_k) = \begin{cases} n_{photon}(q_b - q_i^s q_k^s) = -n_{photon} q_i^s q_k^s & , i \neq k \\ n_{photon}(q_i^s - q_i^s q_k^s) = n_{photon} q_i^s (1 - q_k^s) & , i = k \end{cases} \tag{S13}$$

σ_{peak} and $\sigma_{subpeak}$ in Eq. S10 can be calculated by taking the square root of the Eq. S12.

S3. Comparative experiment results with the deep learning method

To illustrate the strengths and potential limitations of the proposed method, we additionally compared the performance of a deep learning method proposed by Lindell et al.[7] with our method. Lindell's method introduces a convolutional neural network (CNN) to robustly estimate depth information under a sensor fusion configuration, which utilizes an additional high-resolution intensity measurement. However, the intensity image of the target is difficult to match with the depth point cloud, which restricts the application scope of this method. On the other hand, although Lindell's method can effectively reduce the effect of background noise on depth reconstruction accuracy, excessive smoothing of the entire scene will lead to indistinct or distorted edges of objects, as shown in Figure S2a. In contrast, our method emphasizes the accuracy of depth reconstruction at each scanning point, rather than focusing on local smoothness. Specifically, the peak shift of the incident photon sequence at each scanning point is corrected according to the photon detection model to obtain accurate depth. Therefore, as shown in Figure S2d, our method is able to obtain depth information closer to the real scene compared with Lindell's method. Figure S2b,e displays the depth distribution of Models 1-3 reconstructed by Lindell's method and our method, respectively. Although the depth information recovered by Lindell's method can be smoother, the recovered depth information has larger errors compared with the ground truth. As shown in Figure S2c,f, we compared the depth errors of two methods. It is calculated that compared with Lindell's method, our method improves the reconstruction accuracy of Models 1-3 by 31.4%, 50.4% and 32.0%, respectively. These supplementary experimental results demonstrate that our method is able to better preserve edge details and avoid blurring problems.

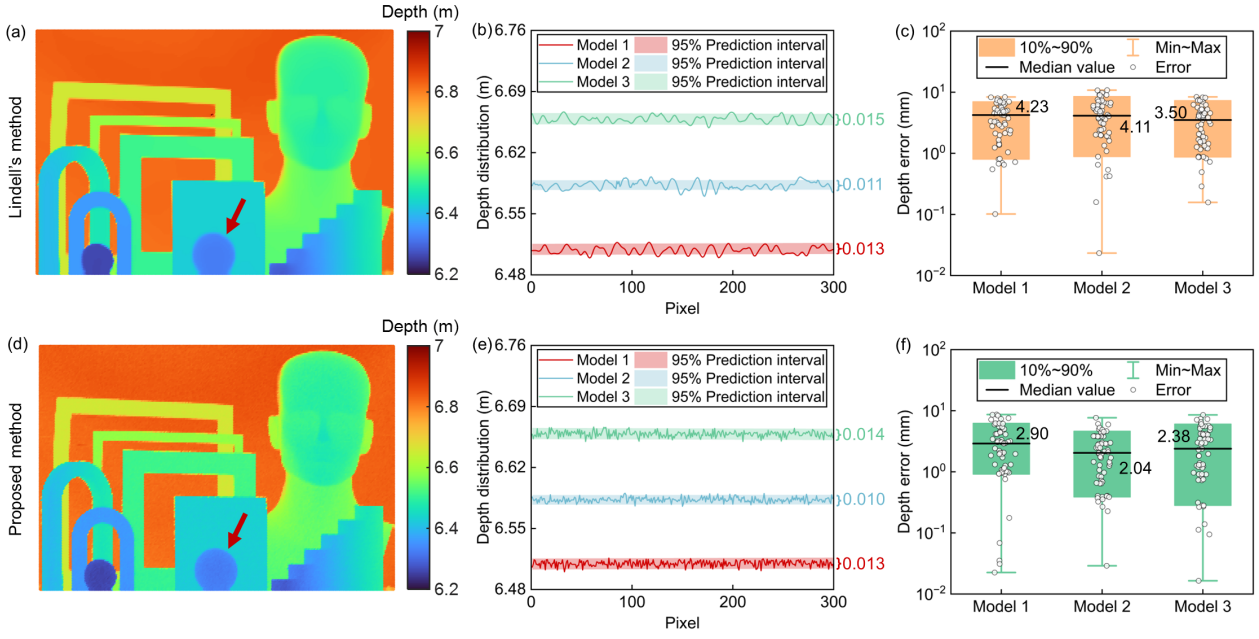

Figure S2. Comparison results of Lindell's method and our method. (a, d) The front view of depth point clouds reconstructed by Lindell's method and our method. (b, e) Depth distribution of models 1-3. (c, f) Comparison of depth errors.

S4. Comparative experimental results under different system parameters

To investigate the effects of different SBR on reconstruction accuracy, we further compared the reconstruction results under SBR=0.3, 1, and 50. As shown in Figure S3a, the test target comprises 8 blocks designed for depth assessments, each with a side length of 20 mm and depth increments from 3 mm to 24 mm in steps of 3 mm. The depth reconstruction results for different SBR are shown in Figure S3c-e. It can be observed that the number of outliers in the depth image decrease as the SBR increases, indicating that the impacts of ambient noise on depth estimation are gradually reduced. The depth distribution of the red line is displayed in Figure S3b. Notably, the grooves between neighboring blocks are indistinguishable in the depth reconstruction for SBR=0.3, highlighting the significant impacts of low SBR on reconstruction. As shown in Figure S3f, due to the high reflectivity of white models, the photons returned from the grooves will be diffusely reflected several times on the sides of the neighboring blocks. Therefore, grooves are reconstructed deeper than the model edges. Error bars for 8 blocks under different SBR are presented in Figure S3g-i. To minimize the impacts of block edges on depth estimation, only the central 8 pixels×8 pixels of each block are considered for error analysis. Calculated results show that the root mean square error (RMSE) of depth information is minimized to 2.37 mm for SBR=50. In contrast, the RMSE values for SBR=0.3 and SBR=1 are 10.87 mm and 4.16 mm, respectively. It implies that the reconstruction accuracy for SBR=50 is approximately 4.59 and 1.76 times that of SBR=0.3 and SBR=1, respectively. Due to the background noise and counting error, the maximum depth error for SBR=0.3 reaches 50.04 mm, while for SBR=50, the maximum error is limited to 8.58 mm. The comparative results demonstrate that increasing SBR can significantly improve the depth reconstruction accuracy.

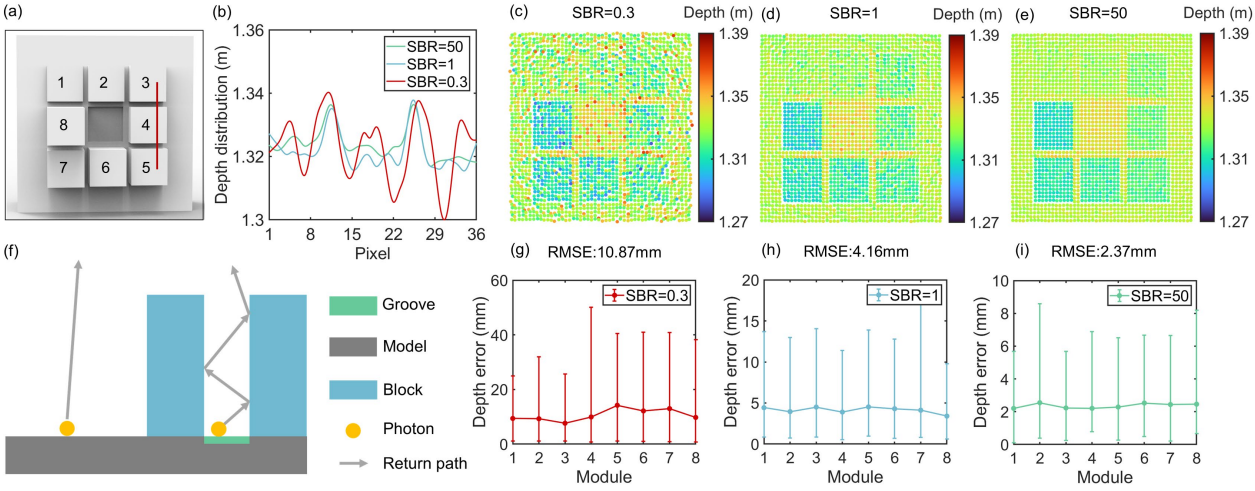

Figure S3. Comparative results reconstructed by our method under different SBR. (a) The test target consisted of 8 blocks with depths ranging from 3 mm to 24 mm in 3 mm increments and the red line used to compare the depth distribution. (b) Depth distributions of the red line. (c-e) Depth point clouds for SBR=0.3, 1 and 50, respectively. (f) Diagram of diffuse reflection at different positions. (g-i) Depth errors of 8 blocks at different SBR values.

In addition, comparative experiments were supplemented to evaluate the influences of other system

parameters on depth reconstruction accuracy, mainly including exposure time, bin width, laser pulse width, and laser spot size. As shown in Figure S4a1-a6, when the exposure time is set to 0.01 s, the blocks are barely discernible, with significant variations in depth within the same plane. As the exposure time increases, more photons are captured at each scanning position, which facilitates the extraction of the peak value of the echo signal. Consequently, block edges under the exposure time=1 s become sharper and depth estimation errors are reduced. RMSE for the exposure time=0.01 s and 0.1 s are 4.11 and 1.37 times that for the exposure time=1 s, respectively. Since the exposure time=0.1 s is sufficient to effectively capture the echo signal, further increasing the exposure time by an order of magnitude has limited improvement in reconstruction accuracy. Therefore, the exposure time=0.1 s is selected as the optimal parameter to ensure effective acquisition of echo signals while maintaining acquisition efficiency.

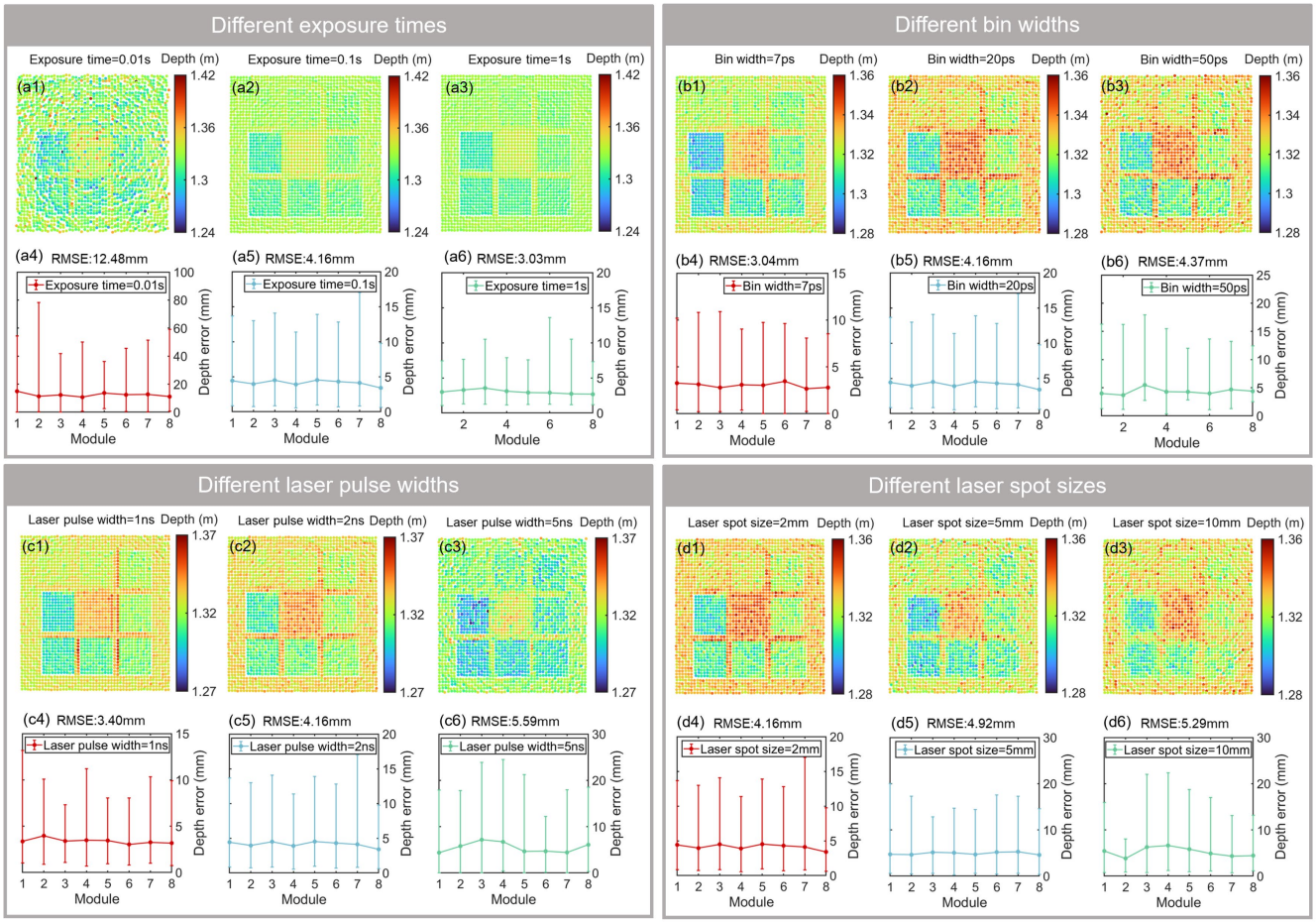

Figure S4. Comparison of depth reconstruction results under different conditions. (a1-a6) Depth point clouds and errors for the exposure time=0.01 s, 0.1 s and 1 s, respectively. (b1-b6) Depth point clouds and errors for the bin width=7 ps, 20 ps and 50 ps, respectively. (c1-c6) Depth point clouds and errors for the laser pulse width=1 ns, 2 ns and 5 ns, respectively. (d1-d6) Depth point clouds and errors for the laser spot size=2 mm, 5 mm and 10 mm, respectively.

The bin width in TCSPC determines the theoretical minimum accuracy achievable without external interference. Figure S4b1-b6 illustrates the effects of bin width on depth reconstruction. As the bin width

narrows, depth errors in the reconstruction results decrease. Compared with the bin width=20 ps and 50 ps, the reconstruction accuracy for the bin width=7 ps can be improved by 36.8% and 43.75%, respectively.

To investigate the influences of laser pulse width on depth reconstruction, we performed comparative experiments with pulse width=1 ns, 2 ns, and 5 ns, which can be seen in Figure S4c1-c6. The calculated RMSE values are 3.40 mm, 4.16 mm, and 5.59 mm, respectively. It indicates that the narrower pulse width result in more accurate depth reconstruction in our experiments.

Finally, we discuss the effects of laser spot size on depth reconstruction, as shown in Figure S4d1-d6. It is observed that the edge information of reconstructed blocks becomes increasingly blurred when the spot size increases. Conversely, smaller laser spot sizes can achieve more accurate depth estimation in our method.

S5. Analysis of the advantages of our method in depth reconstruction

In this section, we will discuss the advantages of our method over previous methods, focusing on the physical characteristics of instruments and the statistical properties of the acquired data. Currently, a major hardware limitation in single-photon imaging is detector dead time, a period of insensitivity following photon detection during which arriving photons cannot be registered. In typical TCSPC applications, a laser with periodic pulses is used, and a detection histogram relative to the most recent illumination is constructed. As shown in Figure S5, without ambient light, the distribution of detected incident photons shows a linear correlation with the time distribution of the laser pulse (Gaussian distribution). However, in the sunlight environment, the early arrival of sunlight photons prevents the detection of subsequent signal photons, distorting the measured histogram towards earlier time bins and causing a non-linear distortion called photon pile-up.

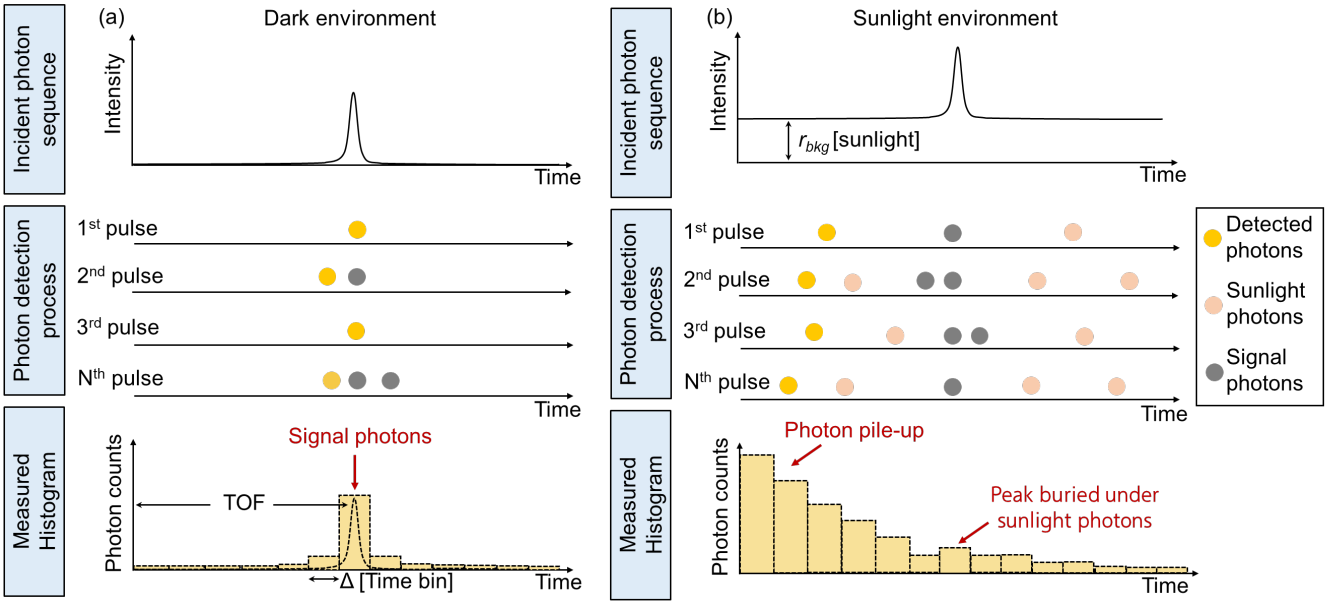

Figure S5. The effect of ambient light on photon detection. (a) In the dark environment, there is a linear mapping between the distribution of detected incident photons and the temporal distribution of the laser pulse. (b) In the sunlight environment, background photons arriving before the laser pulse skew the shape of the histogram, leading to nonlinear distortions.

Single-photon imaging data are mainly divided into two types: One case is that in low photon flux scenes, such as remote imaging, the number of photons received per laser period is typically less than one photon. The acquired histogram is almost unaffected by dead time and approximates a scaled version of the received temporal waveform (Gaussian distribution). The primary challenge in data processing for these scenes lies in effectively separating the echo photon signal from the background noise signal. The other is high photon flux scenes, where the number of photons received per laser cycle is much more than one photon, and the histogram formation procedure causes severe non-linear distortions due to ambient light and dead time. To mitigate these distortions, one common strategy is to limit the acquisition flux so

that photons are detected in at most 5% of illumination periods. However, attenuating the incident flux on the detector leads to inefficient utilization of the light reflected from the scene and slows down the acquisition of sufficient photons for accurate ranging. Allowing higher incident flux and compensating for the distortions would enable faster acquisition without loss of accuracy.

In our method, we firstly employ Coates's estimator to suppress photon pile-up and achieve the incident photon sequence at high photon flux. Subsequently, we converge the estimated photon sequence to the stationary distribution of a Markov chain to reduce ranging error. To mitigate the effects of background noise, depth estimation can be performed by adaptive Gaussian filtering on the stationary distribution. In addition, we also design an error metric for dynamically adjusting the Gaussian kernel width to obtain the optimal filtering results. The corrected photon sequence with our method will be closer to the temporal distribution of the laser pulse and the depth reconstruction will be more accurate.

As a comparison method, the traditional MLE ignores the effect of dead time in the modeling process. Meanwhile, Lindell's method and BM3D focus on spatial smoothness of depth images and do not have the support of the photon detection model in the data processing, leading to the distortion of complex textures and edge information. In addition, the performance of Wavelet Transform relies on the choice of wavelet basis functions. If the wavelet basis function fails to match the features of the incident photon sequence, the improvement of depth reconstruction accuracy will be limited. In consideration of the aforementioned discussions, our method can provide more accurate depth reconstruction.

References

1. Consul, P. C. & Jain, G. C. A generalization of the poisson distribution. *Technometrics* **15**, 791–799 (1973).

2. Lounis, B. & Orrit, M. Single-photon sources. *Rep. Progr. Phys.* **68**, 1129 (2005).

3. Gupta, A., Ingle, A. & Gupta, M. Asynchronous single-photon 3d imaging. In *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 7909–7918 (2019).

4. Gupta, A., Ingle, A., Velten, A. & Gupta, M. Photon-flooded single-photon 3d cameras. *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.* 6770–6779 (2019).

5. Devriendt, K., Martin-Gutierrez, S. & Lambiotte, R. Variance and covariance of distributions on graphs. *SIAM Rev.* **64**, 343–359 (2022).

6. Yager, R. R. Mean, variance and covariance of joint measure based uncertain variables. *Inform. Fus.* **55**, 245–250 (2020).

7. Lindell, D. B., O'Toole, M. & Wetzstein, G. Single-photon 3d imaging with deep sensor fusion. *ACM Trans. Graph.* **37**, 113 (2018).