An Adaptive Photon Count Reconstruction Algorithm for Sparse Count and Strong Noise Count Data With Low Signal Background Ratio

Meijun Chen , Zhendong Shi , Wei Chen, Fangjie Xu , Yong Jiang, Yijiang Mao, Shiyue Xu, Bowen Chen, Yalan Wang, Zecheng Wang, and Jie Leng

Abstract—Single-photon lidar detection data in applications can show different characteristics: sparse count data and strong noise count data with low signal-to-background ratio (SBR), making it difficult to accurately reconstruct depth and intensity information. The existing statistical-based algorithms can achieve reconstruction, but they may lack compatibility for sparse counting and strong noise counting cases which will switch to each other in practical applications. In this paper, an adaptive photon count reconstruction algorithm for sparse count and strong noise count data with low SBR is proposed based on the difference in temporal distribution characteristics between the echo and noise count data. The aggregation characteristic of echo count data in time dimension is proposed to adaptively separate the echo and noise regions in the histogram to reduce the noise interference, and based on the relative difference between count levels in the time neighborhood, an objective function is constructed to reconstruct depth and intensity using optimization. The reconstruction results based on simulated and experimental data confirm that the reconstruction accuracies under both sparse counting and strong noise counting cases are effectively improved under low SBR conditions. Compared with the state-of-the-art algorithms, the depth absolute error is reduced by nearly 50%, the edge error is reduced by an order of magnitude and the proportion of correctly reconstructed pixels reaches 90% when SBR = 0.1. It shows the potential of the proposed algorithm for improving target recognition ability and all-day imaging.

Index Terms—Photon count reconstruction, low SBR, sparse counting, strong noise, single-photon LiDAR.

I. INTRODUCTION

A S A non-contact active detection technique, single-photon LiDAR can obtain the depth and intensity information of the target and has been widely used in various detection scenarios, including natural environment monitoring [1], geographic mapping [2], [3], and autonomous driving [4], [5]. A single-photon LiDAR system functions by emitting laser pulses and capturing the photons reflected from the target using a combination of a time-correlated single photon counting (TCSPC) module and a single-photon detector, such as a single-photon avalanche diode (SPAD). This system is capable of responding to photon-level energy and recording the arrival time of each detected photon. By accumulating detection results from multiple laser pulses, a photon response count histogram is generated. The positions and amplitudes of the target echo counts in the histogram correspond to the target's depth and reflectivity. By extracting these positions and amplitudes for each pixel, depth and intensity images of the detected scene can be reconstructed.

However, single-photon LiDAR often encounters challenging low signal-to-background ratio (SBR) scenarios in real-world applications. These include the sparse counting case—where over 95% of histogram time bins are empty, and some pixels lack any echo counts [6]—commonly observed when acquisition times are reduced [7] or in long-range, low-reflectivity imaging [8], [9], [10]. Another scenario is the strong noise counting case, where echo counts are submerged by numerous noise counts [11], often due to high ambient light levels in daytime detection [12], [13]. In such cases, echo counts are insufficient while noise counts dominate the histogram, making it challenging to accurately reconstruct depth and intensity information [14], [15]. Traditional methods, such as cross-correlation [16] and maximum likelihood (ML) [17], are capable of reconstructing target depth information but become ineffective under low SBR conditions, as they do not account for spatial correlation [18]. Consequently, several photon-efficient algorithms have been developed to enable reconstruction using only a few photons.

For sparse counting case, Kirmani et al. [17] proposed the early first-photon imaging algorithm, which enables image reconstruction using only the first detected photon at each pixel, indicating that sparse reconstruction can be solved by introducing local spatial correlation between pixels. Morris et al. [44] developed and demonstrated imaging techniques capable of functioning with a minimal number of echo photons,

underscoring the feasibility and practicality of achieving efficient imaging even with a limited photon count. Shin et al. [19] proposed a sparse regularization-based reconstruction algorithm based on the first photon imaging algorithm and achieved imaging using the SPAD array when the average photon per pixel is 1, which provides critical theoretical insights that have advanced sparse reconstruction methods. Altmann et al. [20] integrated Bayesian estimation into image reconstruction, modeling pixel correlations as prior distributions and proposing a hierarchical Bayesian reconstruction framework based on the photon Poisson statistical model, which achieves robust reconstructions under challenging conditions with average echo counts per pixel below one. Tachella et al. [6], [21] developed a Bayesian-based sparse reconstruction algorithm using a reversible jump MCMC approach that incorporates an area interaction prior to enhance local spatial correlations, further pushing the boundaries of sparse counting reconstruction. Recently, neural network-based sparse reconstruction techniques have emerged rapidly. Lindell et al. [16] explored a 3D imaging method that fuses data from single-photon sensors with deep learning. Tachella et al. [45] proposed a real-time imaging algorithm called RT3D. Peng et al. [46], [47] developed a photon-efficient 3D imaging method using non-local neural networks, specifically tailored for 3D imaging under low-photon-count conditions. These neural network-based methods are characterized by their ability to rapidly reconstruct scenes.

For strong noise counting case, Rapp et al. [22] introduced the unmixing algorithm (UA), which uses a sliding window to identify echo count clusters in the histogram, effectively separating echo and noise regions and minimizing noise interference. By incorporating spatial correlations between depth and intensity, the UA algorithm achieves reconstruction in strong noise cases with an SBR of 0.04, establishing it as one of the leading reconstruction methods [23]. Halimi et al. [18], [24] proposed the MNR3D algorithm, which models multiscale correlations across time and spatial domains as prior information and employs a hierarchical Bayesian strategy for reconstruction in strong backscatter noise environments. Wang et al. [25] developed a photon count imaging algorithm based on empirical depth error correction, enabling high-flux imaging. Gupta et al. [43] introduced a technique for improving depth measurement accuracy in high ambient noise environments by asynchronously collecting photons to average pile-up distortions across multiple acquisition cycles. Lee et al. [48] proposed an innovative multi-photon cooperative processing technique, which significantly improves imaging quality and resolution under strong noise conditions by processing photon information from multiple pixels simultaneously.

These reconstruction algorithms for sparse counting rely on the sparsity of count data and do not actively suppress noise counts during echo position estimation. As a result, their efficiency and accuracy significantly decline in strong noise conditions due to interference from abundant noise counts. In contrast, reconstruction algorithms for strong noise cases typically suppress noise counts before estimating the echo position and incorporate prior information about spatial correlations and noise level distributions, enabling joint optimization of depth and

intensity, which improves reconstruction in noisy environments. However, in sparse counting scenarios, where only a few photons are detected and echo photons are scarce or absent, noise suppression methods may fail to effectively separate echo and noise regions in the histogram. This issue is particularly pronounced when the signal-to-background ratio (SBR) is low and the count difference between echo and noise is minimal, making accurate echo position estimation challenging. Consequently, these challenges make accurate echo position estimation difficult under varying detection conditions, and existing algorithms struggle to be compatible with both sparse counting and strong noise counting scenarios.

In general, these algorithms typically rely on the impulse response function (IRF) and Poisson statistical model of photon arrivals for image reconstruction, while accounting for spatial correlations between pixels [19], [20], [26]. Although effective, these methods assume the count data closely follow the Poisson distribution and require a significant intensity difference between target and background to avoid depth and intensity estimation bias under low SBR conditions. However, in strong noise scenarios, dead time in detectors leads to photon loss, causing histograms to deviate from the Poisson model [28], [29], [30], which introduces reconstruction errors. Additionally, many of these algorithms depend on user-defined parameters and hand-crafted priors [31], such as noise rate and thresholds, which need to be manually adjusted according to different detection scenarios to obtain better reconstruction results. This reliance limits the algorithm's adaptability and compatibility with both sparse counting and strong noise counting cases.

To improve the compatibility and accuracy of reconstruction algorithms under sparse count and strong noise count conditions with low signal-to-background ratio (SBR < 1) [22], we propose an adaptive photon count reconstruction algorithm. This algorithm incorporates an automatic parameter setting mechanism, eliminating the need for manual tuning across various detection scenarios and counting conditions. The proposed method directly estimates the echo position and reconstructs depth and intensity images based on the temporal distribution of histogram counts, without relying on prior distribution or the Poisson statistical model, this approach results in higher reconstruction accuracy. The core ideas of the proposed algorithm include: (1) A novel method for separating echo and noise: Since the echo signal is concentrated within a time range corresponding to the pulse width near the true depth, we introduce a count aggregation method to adaptively separate echo and noise regions in the histogram. The number of empty bins surrounding each bin is quantified and fed into a decay exponential function to derive the aggregation parameter. A lower count of empty bins indicates stronger aggregation. Regions where the aggregation parameter exceeds a predefined threshold are classified as potential echo regions, while those below the threshold are designated as noise and excluded. This method enables effective separation of echo and noise regions in both sparse and strong noise counting cases. (2) A new depth optimization method: An objective function is formulated based on the degree of count differences within the time neighborhood. The count differences between each bin and its neighboring bins are modeled using a Gaussian function.
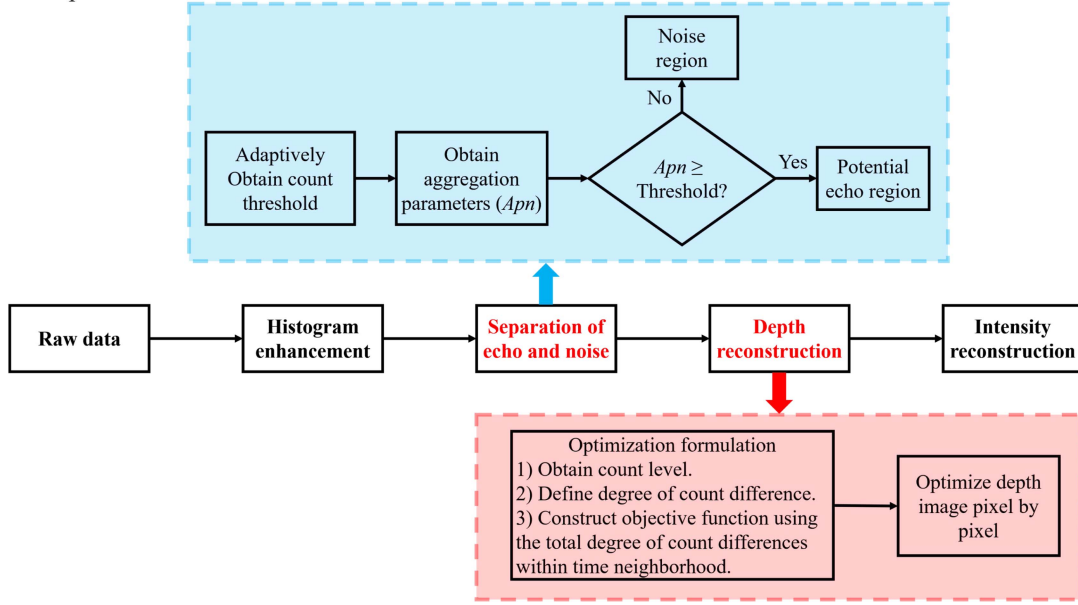
Fig. 1. The overall flow chart of the proposed algorithm.

Depth and intensity are then reconstructed pixel by pixel through optimization.

Simulation and experimental data demonstrate that the proposed algorithm significantly improves reconstruction accuracy in both sparse and strong noise counting scenarios under low SBR conditions. Compared to state-of-the-art methods, the algorithm reduces the absolute reconstruction depth error by nearly 50% in both counting cases, while the reconstruction error at edges and fine details is reduced by an order of magnitude. These improvements are highly advantageous for target recognition [32] and all-day imaging applications.

A. Main Contributions

1) Adaptive Separation of Noise and Echo Regions: This study introduces the aggregation parameter, using a unified framework to separate noise and echo regions under both sparse and strong noise counting conditions. The aggregation parameter is derived from the temporal distribution characteristics of the count data, enabling precise extraction of potential echo regions from the histogram in diverse detection scenarios.

2) Depth Optimization Based on the Degree of Count Differences Within the Time Neighborhood: The proposed echo localization method leverages the inherent concentration of echo counts. By formulating the objective function for depth optimization solely based on count differences within the time neighborhood, this approach eliminates the need for prior models of photon count data. This reduces errors that may arise from discrepancies between the prior model and real-world conditions, thereby enhancing reconstruction accuracy.

B. Outline

Section II explains the overall flow of the reconstruction algorithm and its implementation method. Section III evaluates the performance of the proposed algorithm for both sparse

and strong noise counting cases using simulation and real experimental data. Section IV provides the conclusions, while Section V discusses the limitations of the proposed algorithm and suggests future research directions.

II. ALGORITHM

Fig. 1 illustrates the overall flowchart of the proposed algorithm. The process starts with histogram enhancement, where spatial correlation between pixels is utilized to improve the echo count data. In the second step, the aggregation parameter is defined based on count aggregation in the time dimension, enabling the adaptive separation of echo and noise regions within the histogram. The third step involves formulating an optimization model that quantifies the total degree of count differences within the time neighborhood, facilitating pixel-by-pixel depth reconstruction. Finally, the echo counts are extracted pixel by pixel using the reconstructed depth information, which is then used to reconstruct the intensity image. The visualization of each step in Fig. 1 is provided in Section I of the supplemental material, offering a clear illustration of how each step contributes to the overall reconstruction process.

A. Histogram Enhancement

For sparse count data with low SBR, the echo counts are sparse, and some pixels may not register any detected photons. For strong noise count data with low SBR, the echo counts are obscured by noise. Estimating the echo position based solely on the histogram of a single pixel leads to inaccurate results in both scenarios. Therefore, a histogram enhancement method based on Gaussian weighting, as described in [33] and [34], is proposed to supplement and enhance the echo counts in the histogram during the processing of each pixel. The weight of each pixel is

defined as follows:

$$w_{(i',j')} = \exp\{-[(i-i')^2 + (j-j')^2]/\sigma_w)\}, (i',j') \in Sp_{(i,j)}$$

$$Sp_{(i,j)} = \{(x,y) \in \{1, \ldots, Nr\} \times \{1, \ldots, Nc\} :$$

$$|i-x| \le N/2, |j-y| \le N/2\} \tag{1}$$

Where $Sp_{(i,j)}$ is the set of spatial neighborhood pixels of pixel (i,j), the spatial neighborhood is set to be a $N \times N$ square area. The $\sigma_w > 0$ is the pixel weight adjustment parameter, Nr and Nc represent the number of rows and columns in the scene image, respectively.

Assum that the photon detection period is $Tr = \{1, 2, 3, \ldots, T\}$, and the raw histogram at pixel (i,j) is $y_{(i,j)} = \{y_{(i,j),t} : t \in Tr, (i,j) \in \{1, \ldots, Nr\} \times \{1, \ldots, Nc\}\}$. The enhanced histogram at pixel (i,j), $Y_{(i,j),t}$ is defined as follows:

$$Y_{(i,j),t} = \sum_{(i',j') \in Sp_{(i,j)}} w_{(i',j')} \cdot y_{(i,j),t} \tag{2}$$

Due to the discrete nature of the counts, the accumulated count values are rounded to the nearest integer.

B. Adaptive Separation of Noise and Echo Regions

The echo counts are primarily concentrated within the time range corresponding to the pulse width, indicating a more aggregated distribution. In contrast, noise counts are randomly distributed throughout the entire detection period. Unlike the UA algorithm [22], which simply employs a sliding window to filter out regions with fewer counts and identify echo clusters, our method introduces an aggregation parameter to quantify count aggregation along the time dimension. This allows for the separation of noise and echo regions, with potential echo regions being extracted based on a threshold for the aggregation parameter. Only the counts within these extracted echo regions are processed, thus reducing noise interference and minimizing the data to be handled.

The aggregation parameter quantifies the count aggregation at a specific bin by assessing the number of empty bins (i.e., bins without counts) in its vicinity. The fewer the empty bins, the more aggregated the counts are. However, in scenarios with strong noise counts, the aggregation of counts may not be effectively distinguished by the number of empty bins due to the high noise volume. To address this, the count threshold is adaptively determined based on the distribution characteristics of the count data, enabling the removal of most non-zero noise counts. Subsequently, analyzing the number of empty bins in the vicinity can be effectively applied in both sparse and high noise counting scenarios. This methodology facilitates the adaptive separation of echo and noise regions within a unified framework, efficiently handling both cases.

1) Adaptively Obtaining the Count Threshold: Given the disparate count distributions in sparse and strong noise counting scenarios—where sparse counting results in many empty bins and strong noise counting exhibits numerous non-zero count—the count threshold is adaptively obtained. This is achieved by analyzing the most frequently occurring count value within a sliding window, with counts at empty bins considered as zero.
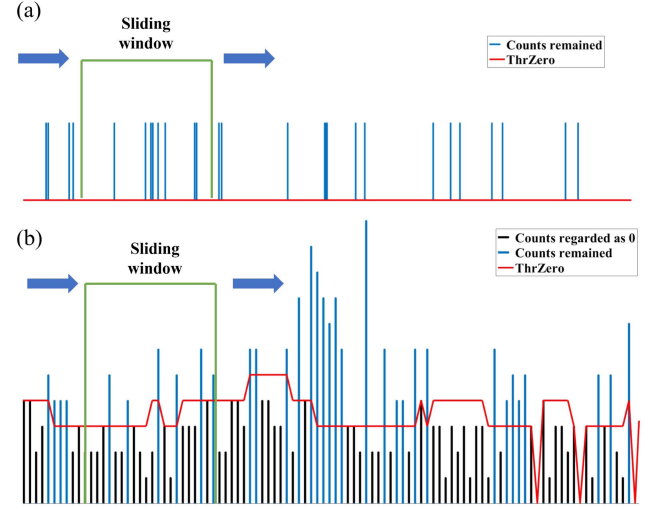

Fig. 2.   Adaptive obtaining of the count threshold using a sliding window. (a) and (b) Illustrate the count thresholds for sparse and strong noise counting cases, respectively. The red line represents the count threshold for each bin, with counts less than or equal to this threshold treated as zero (shown as black lines). The blue lines correspond to counts exceeding the threshold.

This method ensures that the count threshold approximately reflects the local noise level while maintaining a sufficiently low value in the echo regions to preserve echo counts.

The sliding window traverses the histogram to determine the count threshold for each bin. Bins with count values less than or equal to their respective thresholds are subsequently considered empty bins in the calculation of aggregation parameters. The width of the sliding window, Tw, must be greater than the pulse width, Tp, to prevent the count threshold from being set too high in the echo region, which could lead to echo counts being erroneously identified as zero. However, Tw should not be excessively large, as this would hinder the accurate reflection of noise levels across different regions of the histogram. Therefore, Tw is set in accordance with the pulse width Tp, with the fixed relationship $Tw = 2 \cdot Tp$. As a result, there is no need for manual adjustment of Tw. The impact of this fixed relationship on reconstruction accuracy is discussed in Section V of the supplemental material.

Based on the above, the adaptive obtaining of the count threshold is defined as follows:

$$ThrZero_{(i,j),t}$$
$$= \arg\max histogram(\{Y_{(i,j),k}|k \in Wn_t\})$$
$$Wn_t = \left\{t - \frac{Tw}{2}, \ldots, t, t+1, \ldots, t + \frac{Tw}{2}\right\} \tag{3}$$

Where $ThrZero_{(i,j),t}$ represents the count threshold at the t-th bin, Wn_t represents the bins contained within the sliding window when its center is at the t-th bin, $histogram(.)$ represents the statistical count of the occurrences of count values, $\arg\max histogram(.)$ represents obtaining the most frequently occurring count value within the window.

As illustrated in Fig. 2, for sparse counting cases, most of the bins are empty, resulting in the count threshold remaining zero,

as indicated by the red line in Fig. 2(a). In contrast, in strong noise counting cases, the count threshold is influenced by the noise count data and is predominantly non-zero, as shown by the red line in Fig. 2(b). Consequently, the *ThrZero* method has no effect on sparse counting scenarios but can provide some indication of noise levels in situations with strong noise. In this study, it is assumed that pile-up effects have been mitigated during data processing. However, the proposed count threshold method demonstrates potential for mitigating pile-up effects in count data under strong noise conditions, as discussed in Section X of the supplemental material.

2) Obtaining the Aggregation Parameters of the Counts:
After applying the count threshold *ThrZero*, sparse count data remain largely unaffected, while strong noise count data becomes sparser. As a result, the count aggregation characteristics are effectively reflected by the number of empty bins in both cases. The histogram is subsequently traversed again using the sliding window (with the same width, *Tw*) to quantify the number of empty bins in the vicinity of each bin.

To establish a unified analysis framework for both sparse and strong noise counting scenarios, the number of empty bins is incorporated into a decaying exponential function, which quantifies and normalizes the count aggregation characteristics. The aggregation parameter (normalized)$Apn_{(i,j),t}$ is defined as follows:

$$Ap_{(i,j),t} = \frac{\sum\limits_{k \in Wn_t} Y_{(i,j),k}}{Tw} \cdot \exp(-\alpha_0 \cdot Nzero_{(i,j),t})$$

$$Apn_{(i,j),t} = Ap_{(i,j),t}/\max(\{Ap_{(i,j),k}|k=1,2,\ldots,T\}) \quad (4)$$

Where $Nzero_{(i,j),t}$ represents the number of empty bins contained in the sliding window when its center is at the *t*-th bin, and $\alpha_0 > 0$ denotes the sensitivity adjustment parameter for the number of empty bins. Larger values of α_0 can highlight differences in the aggregation characteristics of counts across different time regions, a more detailed analysis of α_0 is provided in Section V of the supplemental material. Once an appropriate α_0 is set, manual adjustment is unnecessary across different scenes and SBR conditions.

The aggregation parameters incorporate both the count level and the time aggregation characteristics, making them more robust under varying noise conditions. The count level term, $(\sum_{k \in Wn_t} Y_{(i,j),k})/Tw$, can highlight the echo region when the echo count level is high, while the time aggregation characteristic term,$\exp(-\alpha_0 \cdot Nzero_{(i,j),t})$, becomes more significant when the echo count level is weak. As observed from (4), the aggregation parameter is large when there are fewer empty bins or when the count level is high, indicating stronger count aggregation characteristics.

3) Separating the Echo and Noise Regions: Since the count aggregation in the echo region tends to be stronger than that in the noise region, the separation of echo and noise regions is achieved by setting a threshold for the aggregation parameters. Regions where the aggregation parameters exceed this threshold are identified as potential echo regions, while those below the threshold are considered noise regions. Given that the pulse width is much shorter than the detection period, most counts in the histogram
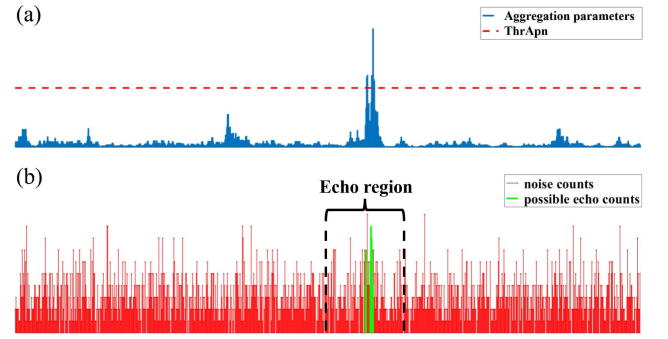

Fig. 3. The echo and noise regions are separated based on the aggregation parameters. (a) illustrates the aggregation parameters *Apn* and its threshold *ThrApn*, and (b) shows the separation of potential echo counts (green lines) and noise counts (red lines) according to the aggregation parameters. The time interval between the dotted lines represents the region that potentially contains the echo counts.

originate from noise photons, resulting in generally lower aggregation levels for noise bins and higher levels for echo bins. As the aggregation parameters are normalized (with a maximum value of 1), the threshold is set to 1/2. This ensures that regions with relatively strong aggregation are included, minimizing the risk of missing true echo regions under low SBR conditions. The impact of this threshold setting on reconstruction results is discussed in Section V of the supplemental material.

The threshold $ThrApn_{(i,j)}$is:

$$ThrApn_{(i,j)} = \max(Apn_{(i,j)})/2 = 1/2 \quad (5)$$

To achieve regions separation, the detection period *Tr* is divided evenly into several non-overlapping time intervals, $Tr = \bigcup_{l=1}^{s} T^l$. The length of each time interval, L_{sub}, is determined by considering both the pulse width *Tp* and the variations in the noise count level. Specifically, L_{sub} must be greater than *Tp* to ensure it encompasses all echo counts. However, L_{sub} should not be too large, as this would include an excessive amount of noise counts, potentially affecting the accuracy of echo position estimation and increasing the computational burden. To mitigate this, the length of the time interval, L_{sub}, is set according to a fixed relationship with the pulse width *Tp*: $L_{sub} = 15 \cdot Tp$. This eliminates the need for manual adjustment of L_{sub} across different scenarios. A more detailed analysis of the impact of this fixed relationship on the reconstruction results is provided in Section V of the supplemental material.

After dividing the detection period, the histogram $Y_{(i,j)}$ and the normalized aggregation parameters $Apn_{(i,j)}$ are also divided into several parts accordingly:

$$Y_{(i,j)} = \bigcup_{l=1}^{s} Y_{(i,j)}^l, Y_{(i,j)}^l = \{Y_{(i,j),k}|k \in T^l\}$$

$$Apn_{(i,j)} = \bigcup_{l=1}^{s} Apn_{(i,j)}^l, Apn_{(i,j)}^l = \{Apn_{(i,j),k}|k \in T^l\} \quad (6)$$

As illustrated in Fig. 3, the separation of noise and echo regions is performed adaptively based on the aggregation parameter. The time intervals that are likely to contain echo counts

are identified and extracted as potential echo regions. The set of echo regions for a given pixel (i,j), denoted as $V_{(i,j)}$, can be derived as follows:

$$V_{(i,j)} = \left\{ \begin{matrix} l_n | \exists t_a \in T^{l_n}, \\ Apn^{l_n}_{(i,j),t_a} \geq ThrApn_{(i,j)}, n = 1, 2, \dots \end{matrix} \right\} \quad (7)$$

In the subsequent processing steps, only the counts within the extracted potential echo regions, $Y_{(i,j),t}, t \in T^{l_n}, l_n \in V_{(i,j)}$, are considered. This method effectively removes the majority of the noise counts, thereby reducing the data that needs to be processed. Furthermore, it enables sparse reconstruction algorithms to better handle cases with strong noise counting, as further discussed in Section VI of the supplemental material.

C. Reconstruction Method

To accommodate both sparse and strong noise counting scenarios, the echo position is directly estimated based on the distribution differences between echo and noise counts in the time dimension. So based on the degree of count differences within the time neighborhood, an objective function is constructed to optimize the depth pixel by pixel. Concurrently, the intensity image is reconstructed by extracting the echo counts for each pixel.

1) Depth Reconstruction: Echo counts are predominantly distributed within the time range of the pulse width, while noise count levels generally exhibit minimal variation or remain constant over short time intervals. Therefore, abrupt changes in count levels within a short time neighborhood can be attributed to echo counts, in other words, the echo counts have a stronger degree of difference relative to other counts in the time interval. By analyzing the total degree of count differences within the time interval, the echo position can be effectively estimated, thereby facilitating accurate reconstruction without the need for external prior information.

However, due to the inherent randomness of single-photon detection, the count values in the histogram are subject to random fluctuations. These fluctuations can obscure the measurement of count differences, especially under low SBR conditions, making it insufficient to rely solely on the difference between two count values. To address this, the count level is initially estimated by smoothing the histogram, which reduces random fluctuations and captures the overall trend in the count data. Both the count value and count level are subsequently integrated into a Gaussian function to quantify and normalize the degree of count difference. This quantification serves as a standard for determining the echo position within the time interval. The total degree of count differences within the time interval is then used as the objective function for pixel-by-pixel depth optimization.

a) Count level obtaining: Given that multiple potential echo regions may be identified (due to *ThrApn* < 1) and noise counts may still persist within these regions, the aggregation parameter is employed to further enhance the counts. This enhancement serves to make the potential echo counts more

distinguishable, thereby highlighting them against the background noise. Counts with aggregation parameters exceeding the threshold are enhanced, while others remain unchanged.

The enhanced histogram $Yen_{(i,j)}$ is defined as follows:

$$Yen_{(i,j),t} = \begin{cases} Y_{(i,j),t} + Apn_{(i,j),t} \cdot Y_{(i,j),t}, Apn_{(i,j),t} \\ \geq ThrApn_{(i,j)} \\ Y_{(i,j),t}, Apn_{(i,j),t} < ThrApn_{(i,j)} \end{cases} \quad (8)$$

The count level is defined as the average of the counts within a short, adjacent time range, akin to downsampling or smoothing the data. This process mitigates the effects of random fluctuations in the count data and emphasizes the underlying trend in the count variations. The count level at t-th bin at pixel (i,j), denoted as $\mu_{(i,j),t}$, is defined as follows:

$$\mu_{(i,j),t} = \frac{1}{Tp} \sum_{k \in Wp_t} Yen_{(i,j),k} Wp_t =$$

$$\left\{ t - \frac{Tp}{2}, \dots, t, t+1, \dots, t + \frac{Tp}{2} \right\} \quad (9)$$

Where Wp_t represents the bins contained in the time range adjacent to the t-th bin.

As with the division of the detection period, the count level is divided accordingly:

$$\mu_{(i,j)} = \bigcup_{l=1}^{s} \mu^l_{(i,j)}, \mu^l_{(i,j)} = \{\mu_{(i,j),k} | k \in T^l\} \quad (10)$$

b) The degree of count difference: To fully utilize both the relative magnitude and the trend of changes between counts, the count values and count levels within the time interval are considered simultaneously. The deviation between these two metrics is then used to define the degree of count difference. To establish a standardized measurement, this deviation is incorporated into a Gaussian function. The degree of count difference between t_c-th count value and t_o-th count level in the l-th time interval at pixel (i,j), denoted as $Dp^l_{(i,j),t_c,t_o}$, is defined as:

$$Dp^l_{(i,j),t_c,t_o} = 1 - (1 - Apn_{(i,j),t_o})$$

$$\cdot \exp\left(-\frac{(Yen^l_{(i,j),t_c} - \mu^l_{(i,j),t_o})^2}{\sigma_{st}} \right) \quad t_c, t_o = 1, 2, \dots, L_{sub}$$

$$l = 1, 2, \dots, s \quad (11)$$

Where $\sigma_{st} > 0$ is the deviation-sensitive adjustment parameter. The impact of this parameter on the reconstruction results is discussed in Section V of the supplemental material, which shows that σ_{st} is a non-sensitive parameter. Furthermore, the $(1 - Apn_{(i,j),t_o})$ term in the formula serves an important function in addressing scenarios where multiple bins within the time interval exhibit high aggregation parameters. In such cases, these bins may all correspond to potential echo locations, leading to similar count levels. Consequently, the degree of count differences between these bins may be low, potentially affecting the accuracy of measuring the total degree of count differences for the true echo counts within the time interval.
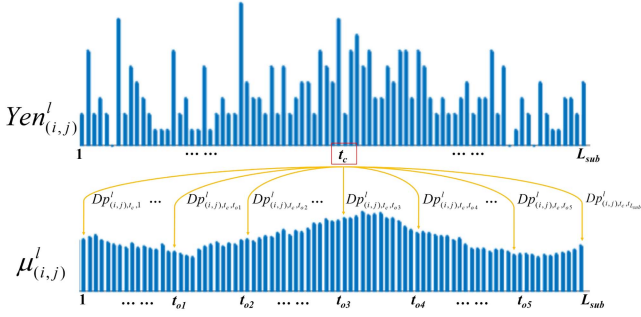
Fig. 4. The degree of count differences between the count value *Yen* and the count level μ in the l-th time interval.

When the aggregation at the t_o-th bin of the time interval is high, the $(1 - Apn_{(i,j),t_o})$ term in the formula diminishes, leading to an increase in the degree of count difference $Dp^l_{(i,j),t_c,t_o}$. Therefore, the $(1 - Apn_{(i,j),t_o})$ term is crucial in mitigating the mutual influence of counts with high aggregation parameters, thereby ensuring more accurate echo position determination.

As shown in Fig. 4, for each bin within the time interval, the degree of count differences between that bin and all other bins can be determined.

c) Objective function: The total degree of count differences is defined as the product of the degree of count differences within the given time interval (i.e., the potential echo region). It serves as the objective function for optimizing depth estimation.

As discussed previously, the length of the time interval L_{sub} is set to $L_{sub} = 15 \cdot Tp$. This length is sufficient to encompass all the echo counts and also includes enough noise count bins to serve as bins with low degrees of count differences, thereby enhancing the prominence of the echo counts in the output of the objective function. Additionally, the degree of count difference relative to a count's own position is not considered and is set to 1 in the objective function formulation. The objective function in the l-th time interval, denoted as $TDp^l_{(i,j),t_c}$, is defined as follows:

$$TDp^l_{(i,j),t_c} = \prod_{\substack{t_o=1 \\ t_o \neq t_c}}^{L_{sub}} Dp^l_{(i,j),t_c,t_o} = \prod_{\substack{t_o=1 \\ t_o \neq t_c}}^{L_{sub}}$$

$$\left[1 - (1 - Apn_{(i,j),t_o}) \cdot \exp\left(-\frac{\left(Yen^l_{(i,j),t_c} - \mu^l_{(i,j),t_o}\right)^2}{\sigma_{st}} \right) \right]$$

$$t_c = 1, 2, \ldots, L_{sub}; l = 1, 2, \ldots, s \tag{12}$$

Where the objective function $TDp^l_{(i,j),t_c} \in (0,1)$ represents the total degree of count differences of the t_c-th count in the l-th time interval of the histogram at pixel (i,j), and the bin position with highest *TDp* is used as the estimation of the echo position.

d) Depth optimization: To ensure smooth transitions between depth values in the reconstructed image, a variation term (TV) is introduced. This term helps guide the depth range of the current pixel based on the depths of adjacent pixels that have already been reconstructed (i.e., the left and top neighboring

pixels when reconstruction is performed column-wise). The TV term used in the subsequent optimization is defined as follows:

$$TV^l_{(i,j),t_c} = \sqrt{\left(z^l_{(i,j),t_c} - \hat{z}_{(i,j-1)}\right)^2 + \left(z^l_{(i,j),t_c} - \hat{z}_{(i-1,j)}\right)^2}$$

$$z^l_{(i,j),t_c} = t^l_{(i,j),t_c} \cdot \tau \cdot c/2 \tag{13}$$

Where $\hat{z}_{(i,j-1)}$ and $\hat{z}_{(i-1,j)}$ represent the depth estimates of the left and top pixels, respectively, τ is the time length of the bin, and c is the speed of light.

The echo position is estimated only within the extracted potential echo regions $V_{(i,j)}$. The optimization formulation for the echo position is as follows: equation (14) shown at the bottom of the next page

Where $l^{echo}_{(i,j)}$ denotes the time interval which contains the echo and $tc^{echo}_{(i,j)}$ denotes the bin position of the echo peak in this time interval. Therefore, the absolute bin position of the echo peak is $\hat{t}^{echo}_{(i,j)} = (l^{echo}_{(i,j)} - 1) \cdot L_{sub} + tc^{echo}_{(i,j)}$, and depth estimation at the pixel (i,j) is as follows:

$$\hat{z}_{(i,j)} = \hat{t}^{echo}_{(i,j)} \cdot \tau \cdot c/2 \tag{15}$$

Under low SBR conditions, the proposed depth optimization method outperforms ML-based optimization in terms of reconstruction accuracy. For further details, please refer to Section VII of the supplemental material.

2) Intensity Reconstruction: The echo intensity at each pixel is estimated based on the reconstructed depth information. The counts in a time range with pulse width length adjacent to the echo peak position are extracted and summed, and the other counts in the time interval are used to estimate the noise level. The average single-pulse echo counts are then utilized for intensity estimation. Specifically, incorporating the Gaussian weights from the spatial neighborhood pixels introduced earlier, the intensity is estimated as follows:

$$\hat{I}_{(i,j)} =$$

$$\frac{\sum_{t_c \in W_{p_{tc}^{echo}_{(i,j)}}} Y^{l^{echo}_{(i,j)}}_{(i,j),t_c} - \frac{Tp}{L_{sub}-Tp} \cdot \sum_{t_c \notin W_{p_{tc}^{echo}_{(i,j)}}} Y^{l^{echo}_{(i,j)}}_{(i,j),t_c}}{M \cdot \sum_{i'=1}^{N} \sum_{j'=1}^{N} w_{(i',j')}} \tag{16}$$

Where $\hat{I}_{(i,j)}$ denotes the reconstructed intensity at pixel (i,j) and M denotes the number of pulses accumulated at pixel (i,j) during the detection.

III. SIMULATION RESULTS

Simulation data for both sparse and strong noise counting cases, as well as real experimental datasets, are utilized to evaluate the performance of the proposed algorithm. Additionally, the algorithm is also compared with several state-of-the-art reconstruction algorithms to assess the improvements in reconstruction accuracy.

A. Evaluation Metrics

For the reconstructed depth image, the overall reconstruction error is evaluated using Depth Absolute Error (DAE), Root Mean

Square Error (RMSE), and True Percentage (True pct) (i.e., the percentage of pixels satisfying a given depth absolute error) [24]. Additionally, Soft Edge Error (SEE) [35], [36] is employed to assess reconstruction accuracy of edges and details within the scene. The Multi-Scale Structural Similarity Index Measure (MS-SSIM) [37] is utilized to measure the structural similarity between the reconstructed depth image and the reference depth image. For the reconstructed intensity image, the Log Mean Squared Error (log MSE, abbreviated as MSE) [22] is used to evaluate the reconstruction performance. The definitions of these quantitative metrics are as follows:

$$DAE = \frac{1}{Nr \cdot Nc} \sum_{i=1}^{Nr} \sum_{j=1}^{Nc} |z_{(i,j)} - \hat{z}_{(i,j)}| \tag{17}$$

$$RMSE = \sqrt{\frac{1}{Nr \cdot Nc} \sum_{i=1}^{Nr} \sum_{j=1}^{Nc} (z_{(i,j)} - \hat{z}_{(i,j)})^2} \tag{18}$$

$$SEE = \frac{1}{|Edge(z)|} \cdot \sum_{(i,j) \in Edge(z)} \min_{(i',j') \in v_{(i,j)}} |z_{(i,j)} - \hat{z}_{(i,j)}| \tag{19}$$

$$MSE = 10 \cdot \log_{10} \left(\frac{1}{Nr \cdot Nc} \sum_{i=1}^{Nr} \sum_{j=1}^{Nc} (I_{(i,j)} - \hat{I}_{(i,j)})^2 \right) \tag{20}$$

Where Nr and Nc is the size of image, $z_{(i,j)}$ and $\hat{z}_{(i,j)}$ represent the reference depth and reconstructed depth at the pixel (i,j), respectively. $Edge(z)$ is the set of edge pixels in reference depth image, $|Edge(z)|$ is the total number of pixels in $Edge(z)$, and $v_{(i,j)}$ represents the spatial neighborhood of pixel (i,j) in $Edge(z)$. The $I_{(i,j)}$ and $\hat{I}_{(i,j)}$ represent the reference intensity and reconstructed intensity at the pixel (i,j), respectively.

B. Simulated Data

The scenes Art, Bowling, and Reindeer from the Middlebury dataset [38], [39] are utilized to generate the simulated data. Among these, the Art scene is highly detailed, representing a complex scene; the Bowling scene exhibits smooth, planar surface characteristics, representing a simpler scene; while the Reindeer scene offers a moderate level of complexity. The depth and intensity reference images for these scenes are shown in Fig. 5. Each simulation scene has an image resolution of 278 × 348 pixels. The histogram for each pixel contains 1024 bins, with a detection period of 100 ns and a laser pulse width of 0.27 ns.
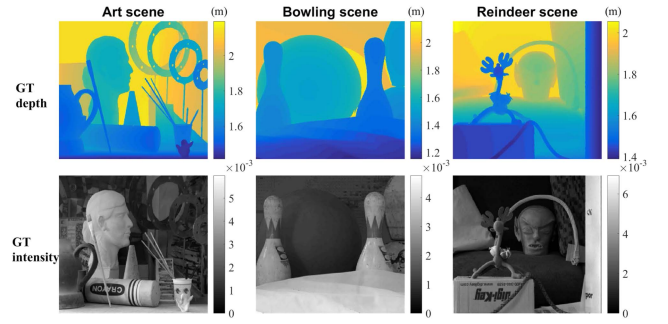

Fig. 5. Reference depth and intensity images for the art, bowling, and reindeer scenes. The top row displays the reference depth images, while the bottom row shows the reference intensity images.

The simulated data are generated by combining the reference image and the average photon number. The grayscale image provided by the dataset is used to represent the scene's reflectivity image. The echo count at each pixel is set to be proportional to the corresponding reflectivity, thereby generating the reference intensity, as previously used in [22]. A Gaussian pulse shape with a specified mean of $2 \cdot z_{(i,j)}/(c \cdot \tau)$ and variance $\sigma = Tp/2$ is used to simulate the impulse response of the system. The amplitude of the pulse is set equal to the intensity, and the average counting rate is obtained by summing the system's impulse response and the noise counting rate. Subsequently, the count data are generated using the Poisson statistical model of arriving photon [40].

Additional simulation reconstruction results for other scenes will be provided in Section III of the supplemental material, which includes four additional scenes from the Middlebury dataset [38], [39], as well as three scenes from the NYU RGB-D v2 dataset [42].

C. Results on Simulated Data

In the following, the simulated data for sparse counting and strong noise counting cases are reconstructed separately to analyze the adaptability of the proposed algorithm to both scenarios. The proposed algorithm is quantitatively compared with the classical ML [17], Shin [19], UA [22], ManiPoP [6], [21], and CASPI [48] algorithms under varying SBR conditions to evaluate reconstruction performance. For comparisons with other recent related works, please refer to Section IV of the supplemental material. The proposed algorithm uses the following parameters in all reconstruction tests: $\alpha_0 = 0.8$, $N = 3$, $\sigma_{st} = 1$, and Tw and L_{sub} have fixed relationships with the pulse width Tp:

$$
\begin{aligned}
(l_{(i,j)}^{echo}, tc_{(i,j)}^{echo}) &= \underset{(l_n, t_c)}{\arg\min} (-\log TDp_{(i,j),t_c}^{l_n} + \beta \cdot TV_{(i,j),t_c}^{l_n}) \\
&= \underset{(l_n, t_c)}{\arg\min} \left\{ -\sum_{\substack{t_o=1 \\ t_o \neq t_c}}^{L_{sub}} \log \left[1 - (1 - Apn_{(i,j),t_o}) \cdot \exp\left(-\frac{(Yen_{(i,j),t_c}^{l_n} - \mu_{(i,j),t_o}^{l_n})^2}{\sigma_{st}} \right) \right] + \beta \cdot TV_{(i,j),t_c}^{l} \right\} \\
l_n &\in V_{(i,j)}; t_c = 1, 2, \ldots, L_{sub}
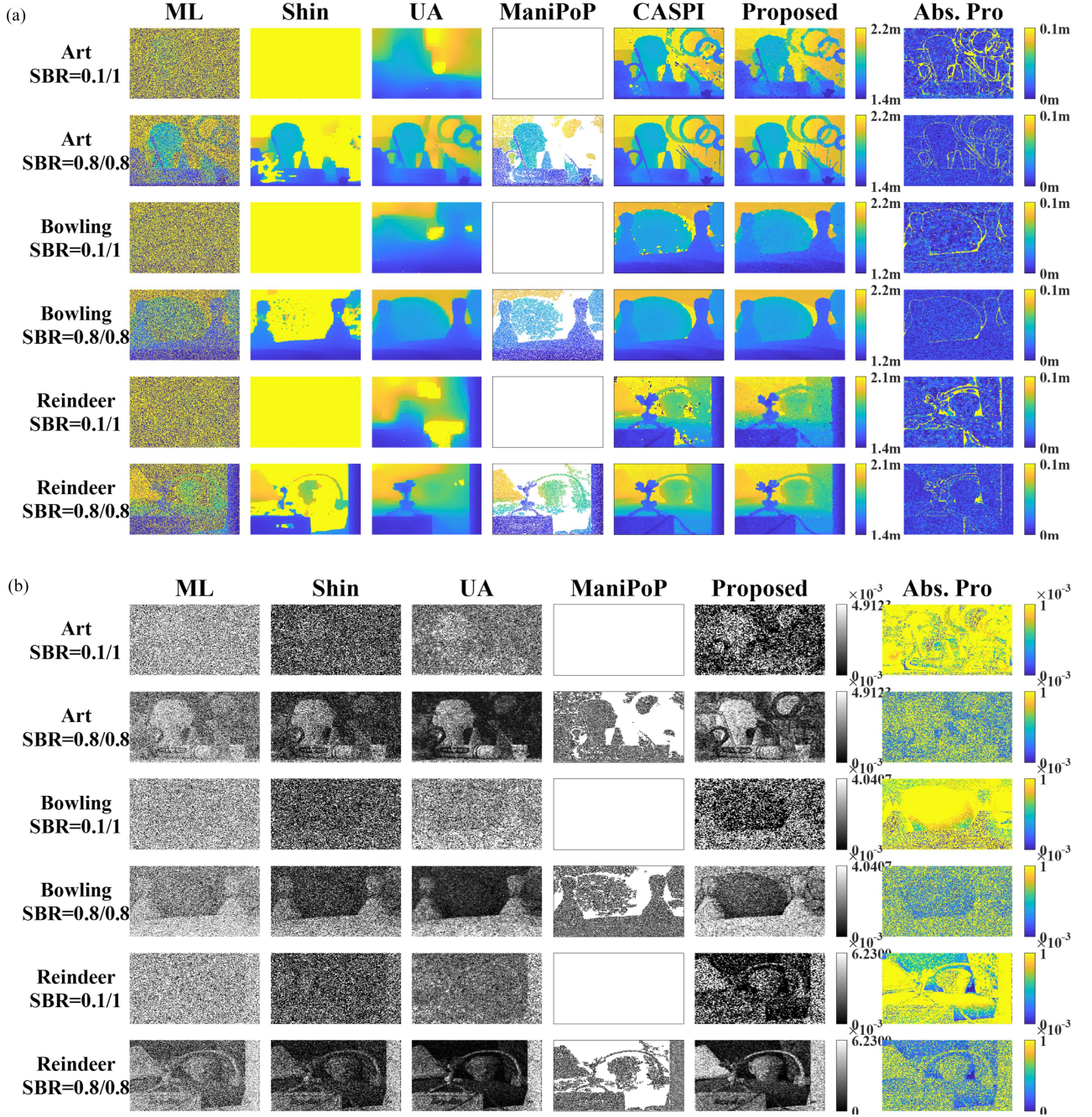\end{aligned}
\tag{14}
$$

Fig. 6. Reconstruction images and absolute error images of depth and intensity for the art, bowling, and reindeer scenes under sparse counting cases, with Sppp set to 0.1 and 0.8, and corresponding SBR set to 0.1 and 1. (a) shows the reconstructed depth images, and the absolute error images of the proposed algorithm, while (b) shows the reconstructed intensity images, and the absolute error images of the proposed algorithm.

$Tw = 2·Tp$, $L_{sub} = 15·Tp$. Furthermore, no manual adjustment of these parameters is required across different scenes and SBR conditions.

1) Reconstruction Results for Sparse Counting Case: To analyze the reconstruction performance in the sparse counting case, where more than 95% of the bins in the histogram are empty and some pixels lack echo count data [6], the average signal photons per pixel (Sppp) is set to 0.1 (SBR = 0.1). Additionally, to further evaluate reconstruction performance when the echo

count is slightly higher, simulation data with Sppp = 0.8 (SBR = 1) are also used for reconstruction.

The reconstruction results of the scene depth image are shown in Fig. 6(a). The traditional ML algorithm has a high requirement for SBR and needs more echo photons to achieve better results, leading to reconstructed images that are noisy and lack all detail information in the scene. The Shin algorithm is significantly affected by noise, rendering depth reconstruction completely invalid when Sppp is low. The UA algorithm effectively reduces

TABLE I
QUANTITATIVE COMPARISON ON ART, BOWLING, AND REINDEER SCENES IN THE SPARSE COUNTING CASE WITH DIFFERENT LEVELS OF SPPP AND SBR

	SBR = 0.1/1 = 0.1						SBR = 0.8/0.8 = 1					
	DAE (m)	RMSE (m)	SEE (m)	MS-SSIM	True pct	MSE (dB)	DAE (m)	RMSE (m)	SEE (m)	MS-SSIM	True pct	MSE (dB)
Art												
ML	4.2355	5.7142	0.6126	0.0007	0.0611	-30.77	2.6360	4.4487	0.0594	0.0088	0.3852	-48.59
Shin	4.9175	4.9311	4.8140	0.1070	0	-39.30	0.7887	1.2662	0.6733	0.0769	0.3990	-54.67
UA	0.1580	0.4079	0.1233	0.1503	0.3952	-50.59	0.0424	0.0779	0.0210	0.8359	0.7257	-59.79
ManiPoP	1.8279	1.8427	1.6815	0	0	X	1.1228	1.4643	1.0572	0	0.4058	X
CASPI	0.4081	1.8231	0.1133	0.0665	**0.8661**	X	**0.0173**	**0.0495**	**0.0037**	**0.9658**	**0.9552**	X
Proposed	**0.0488**	**0.1871**	**0.0095**	**0.7835**	0.8496	**-51.65**	0.0211	0.0519	0.0045	0.9603	0.9456	**-60.81**
Bowling												
ML	4.2606	5.8527	0.5030	0.0002	0.0616	-30.77	2.6622	4.5919	0.0520	0.0029	0.3961	-48.51
Shin	5.2176	5.2297	4.9786	0.1037	0	-39.24	0.9532	1.4920	0.5538	0.0214	0.4164	-54.60
UA	0.1400	0.3632	0.0583	0.1423	0.4974	-50.86	0.0199	0.0604	0.0157	0.9215	0.9267	-60.32
ManiPoP	1.5112	1.5307	1.4217	0	0	X	0.7776	1.1278	0.6010	0	0.5159	X
CASPI	0.1908	1.2677	0.1236	0.1100	**0.9483**	X	**0.0124**	**0.0425**	**0.0034**	**0.9816**	**0.9843**	X
Proposed	**0.0275**	**0.0778**	**0.0101**	**0.8576**	0.9338	**-51.57**	0.0163	0.0495	0.0044	0.9464	0.9799	**-60.72**
Reindeer												
ML	4.2368	5.7498	0.6147	0	0.0626	-30.75	2.7312	4.5670	0.0984	0.0042	0.3733	-48.50
Shin	5.0365	5.0607	5.0653	0.0260	0	-39.20	1.4035	2.1641	0.9319	0.0231	0.3987	-54.58
UA	0.2177	0.5749	0.2346	0.0727	0.5085	-50.06	0.0474	0.3124	0.0164	0.6891	0.8429	-57.98
ManiPoP	1.7145	1.7277	1.6420	0	0	X	1.0462	1.3712	0.6728	0	0.4134	X
CASPI	0.7786	2.6144	0.1929	0.0272	0.8500	X	0.0287	0.3979	0.0039	0.6213	**0.9775**	X
Proposed	**0.0404**	**0.2612**	**0.0087**	**0.6701**	**0.8886**	**-51.42**	**0.0175**	**0.0399**	**0.0040**	**0.9456**	0.9616	**-60.95**

noise and limits error through smoothing operations. However, excessive smoothing occurs when echo count data is insufficient, resulting in large reconstruction errors at the edges and details. The ManiPoP algorithm requires a high number of echo photons and fails to reconstruct images when the SBR is low, resulting in many empty pixels (white areas) in the reconstructed images. The CASPI algorithm demonstrates the capability to achieve overall complete reconstruction, but some areas in the scene exhibit reconstruction failures, particularly around the edges of objects. The proposed algorithm achieves effective reconstruction even when Sppp is low, retaining most target details and edges, and the reconstruction of details improves further as Sppp increases. As shown in the absolute error images, the overall error of the proposed algorithm is low and the reconstruction errors are primarily concentrated in very narrow areas along the edges of the target, demonstrating that the proposed algorithm reduces reconstruction error caused by abrupt depth changes at edges.

The reconstruction results of the scene intensity image are shown in Fig. 6(b). The traditional ML method estimates echo intensity by summing the counts near the echo position, causing the reconstructed intensity to fluctuate with noise counts. The Shin and ManiPoP algorithms fail to reconstruct the intensity image when echo count data are insufficient. While the UA algorithm can effectively reconstruct the intensity image under low SBR, but the images are blurred due to the smoothing operation. The intensity image reconstruction effect of the proposed algorithm is better than that of the Shin and ManiPoP algorithms

and comparable to that of the UA algorithm, with relatively better performance in capturing details and edges. Intensity reconstruction comparison with CASPI was not conducted because its intensity does not represent the photon counts from a single pulse.

Quantitative comparison: Table I presents the evaluation metrics corresponding to the reconstructed images in Fig. 6. The MSE is calculated based on the reconstructed intensity images, while the other evaluation metrics are calculated based on the reconstructed depth images (this applies to subsequent tables as well).

For the Art scene, the proposed algorithm's reconstruction metrics outperform CASPI in low SBR conditions. The SEE values are an order of magnitude lower compared to other algorithms, indicating a significantly better preservation of scene detail. Furthermore, the True pct (with an absolute error range set to 0.05 m) for the proposed algorithm is also significantly higher compared to other methods, demonstrating that the reconstructed image has fewer anomalous pixels. The UA algorithm reduces errors through smoothing operations, but this smoothing also causes blurring of edges in the reconstructed images, leading to a higher SEE. The MS-SSIM value for the proposed algorithm remains high even under SBR = 0.1 (0.1/1), substantially surpassing other algorithms, which highlights its robust performance under photon-scarce conditions. As SBR increases to 1 (0.8/0.8), the MS-SSIM value of the proposed algorithm approaches 1, indicating improved preservation of scene structure, and to some extent, a better subjective similarity

to the reference image. Regarding the MSE of the reconstructed intensity images, the proposed algorithm performs comparably to the UA algorithm and outperforms both the Shin and ManiPoP algorithms.

For the Bowling scene, which features few details and edges with most areas consisting of slowly changing planes, the UA algorithm addresses noise pixels with abrupt depth changes through smoothing. However, the proposed algorithm achieves even lower reconstruction errors, particularly when SBR = 0.1 (0.1/1). The DAE and RMSE values are significantly lower than those of the UA algorithm, indicating that the proposed algorithm not only provides superior detail reconstruction but also maintains smooth reconstruction for planar areas. Additionally, when SBR = 0.1 (0.1/0.1), the TruePct of the proposed algorithm is already close to 1, indicating that it accurately reconstructs the majority of pixels even under low echo counts. This highlights the proposed algorithm's capability to handle photon scarcity and noise interference effectively.

For the Reindeer scene, where some details exhibit minimal contrast with the background, there is a higher risk of detail loss. The smoothing operation employed by the UA algorithm results in the complete removal of the head model on the right side of the scene, leading to a large SEE. In contrast, the proposed algorithm effectively reconstructs the scene's detail information, achieving lower reconstruction errors and outperforming all other methods across the evaluation metrics.

2) Reconstruction Results for Strong Noise Counting Case: To analyze the reconstruction effect in strong noise counting cases [11], [22], the Sppp is set to 2 (SBR = 0.01). Additionally, simulation data with Sppp = 2 (SBR = 0.04) are used to further evaluate reconstruction performance under slightly reduced noise conditions.

The reconstruction results of the scene depth image are shown in Fig. 7(a). In the case of strong noise count, the ML algorithm produces results heavily corrupted by noise, while the ManiPoP algorithm fails to reconstruct most pixels. The reconstructed depth of the Shin algorithm is entirely outside the true depth range of the scene. The UA algorithm shows improved reconstruction due to the increased number of echo photons; however, it still exhibits substantial reconstruction errors at the edges. For instance, in the Reindeer scene, the UA algorithm only produces a blurred representation of the reindeer toy on the left. Additionally, the UA algorithm struggles to distinguish between the head model and the backdrop on the right side, leading to a loss of the target. In contrast, the CASPI and proposed algorithm successfully reconstructs both areas with high accuracy, demonstrating lower edge error and enhanced capability to differentiate targets with similar depth. However, the CASPI algorithm still exhibits some reconstruction failures at the edges of objects in the reconstructed images

The intensity image reconstruction results of the scene are shown in Fig. 7(b). The traditional ML method fails completely due to the excessive noise counts in the histogram., This results in reconstructed intensity values that are disproportionately influenced by the noise level and exceeds the reference intensity range. The reconstructed intensity images produced by the Shin algorithm are highly blurred and overlaid with noise.

The ManiPoP algorithm yields intensity images with significant areas left empty. The proposed algorithm performs comparably to the UA algorithm, with the reconstruction of details and edges showing relatively better results

Quantitative comparison: Table II presents the evaluation metrics corresponding to the reconstructed images in Fig. 7. In the strong noise counting case, the proposed algorithm exhibits superior performance in almost all evaluation metrics except for MSE when SBR is low, indicating a significant overall improvement in performance. While the CASPI algorithm shows slight improvements in some metrics compared to the proposed algorithm at SBR = 2/50, it has a significantly longer runtime, taking several hours to complete the reconstruction of a single image. For a more detailed analysis of the CASPI algorithm in comparison with the proposed algorithm, please refer to Sections IV and IX of the supplemental material. This suggests a comprehensive improvement in performance by the proposed algorithm, demonstrating the proposed algorithm's effectiveness in managing strong noise conditions. Additionally, the increase in echo count data has led to improved evaluation metrics for the reconstruction results compared to the sparse counting case.

3) Performance Evaluation: This section evaluates the reconstruction performance of the proposed algorithm using simulated data across various SBR conditions. The scenes including Art, Bowling, and Reindeer (size $278 \times 348 \times 1024$) are used for analysis. The SBR is set to {0.1, 1, 1.5, 2}, and the corresponding Sppp under each SBR condition is set to {0.1, 0.2, 0.4, 0.8, 1.6, 3.2, 6.4, 10} [22]. The evaluation metrics curves, as shown in Fig. 8, illustrate the trend in the reconstruction accuracy of the proposed algorithm across different SBR conditions over a wide range of Sppp. This allows for a comprehensive analysis of the algorithm's performance across various Sppp intervals and SBR conditions. To better visualize the trend in reconstruction metrics, Fig. 8 displays averaged metrics across the three scenes. More granular scene comparisons are provided in Section VIII of supplemental material.

The DAE curves of the reconstructed depth image, as shown in Fig. 8(a), show that as Sppp increases, the DAE of all algorithms decreases and eventually stabilizes. However, the DAE of the proposed algorithm is reduced by nearly 50% than the other algorithms when Sppp is lower, it shows the proposed algorithm's superior performance in maintaining low overall reconstruction error even with insufficient echo count data. The trend of the RMSE curves, as shown in Fig. 8(b), exhibit a similar trend.

The SEE curves of the reconstructed depth image, as shown in Fig. 8(c), show that the SEEs of these algorithms decrease as Sppp increases, indicating that the reconstruction effect for the edges is improved with the increase of echo counts. Notably, the proposed algorithm shows a marked improvement, with SEE values being an order of magnitude lower compared to other algorithms at lower Sppp values. This highlights the proposed algorithm's exceptional capability in preserving structural details and edge integrity, even under conditions of insufficient echo counts.

The MS-SSIM curves of the reconstructed depth image, as shown in Fig. 8(d), show that the MS-SSIMs of both the proposed algorithm and the UA algorithm increase rapidly with the
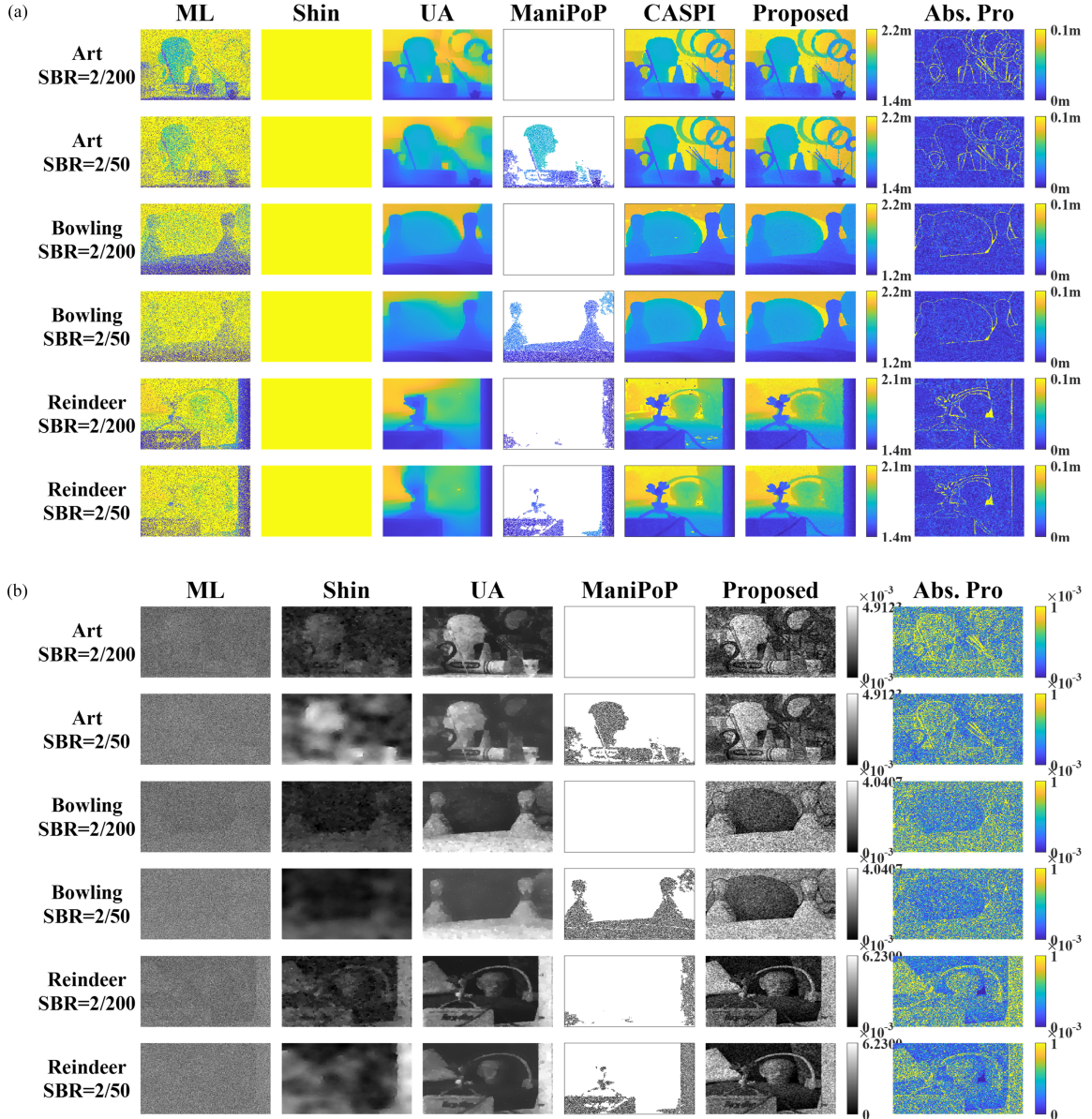
Fig. 7.   Reconstructed images and absolute error images of depth and intensity for art, bowling and reindeer scenes in the strong noise counting cases, with Sppp set to 2 and the corresponding SBR set to 0.01 and 0.04. (a) shows the reconstructed depth images and the absolute error images, while (b) shows the reconstructed intensity images and the absolute error images.

increase of Sppp and eventually converge to 1. However, the proposed algorithm consistently outperforms the UA algorithm, especially when the Sppp and SBR are low, while the MS-SSIMs of other algorithms are always at a lower level. This indicates superior structural preservation by the proposed algorithm under conditions of low SBR and insufficient echo counts.

The True pct curves of the reconstructed depth image, as shown in Fig. 8(e), show that the proposed algorithm already achieves a True pct of 90% even when echo counts are extremely low (Sppp = 0.1), which is significantly higher compared to other algorithms.

The MSE curves of the reconstructed intensity image, as shown in Fig. 8(f)), show that the reconstruction effect for intensity image of proposed algorithm is comparable to that

of the UA algorithm and superior to the Shin and ManiPoP algorithms.

All evaluation metrics improve and converge as the Sppp increases. Notably, the proposed algorithm significantly outperforms the other algorithms, particularly under low Sppp and low SBR conditions. This indicates the proposed algorithm's exceptional reconstruction performance under conditions of limited echo photons and low SBR.

D. Results on Real Experimental Dataset

In order to compare and verify the reconstruction effect of the proposed algorithm on the real-world scene, we further evaluate the proposed algorithm on the real experimental dataset provided

TABLE II

QUANTITATIVE COMPARISON ON ART, BOWLING, AND REINDEER SCENES IN THE STRONG NOISE COUNT CASE WITH DIFFERENT LEVELS OF SPPP AND SBR

|	SBR = 2/200 = 0.01						SBR = 2/50 = 0.04					
	DAE (m)	RMSE (m)	SEE (m)	MS-SSIM	True pct	MSE (dB)	DAE (m)	RMSE (m)	SEE (m)	MS-SSIM	True pct	MSE (dB)
Art												
ML	4.8614	6.4714	0.3184	0.0036	0.1731	-13.97	3.8858	5.7789	0.1285	0.0033	0.3328	-25.93
Shin	5.6016	5.6058	5.4549	0.1220	0	-32.55	5.3760	5.3803	5.2459	0.1612	0	-55.79
UA	0.0802	0.1257	0.0307	0.4123	0.5665	**-65.31**	0.0477	0.0932	0.0217	0.7261	0.7247	**-67.93**
ManiPoP	1.8279	1.8427	1.6815	0	0	X	1.4861	1.6842	1.4391	0	0.2112	X
CASPI	0.2842	1.7397	0.0775	0.0909	**0.9369**	X	**0.0169**	0.0737	**0.0037**	**0.9677**	**0.9595**	X
Proposed	**0.0256**	**0.0701**	**0.0044**	0.9186	0.9274	-62.67	0.0195	0.0530	0.0038	0.9555	0.9509	-64.07
Bowling												
ML	5.0813	6.7091	0.2996	0.0017	0.1736	-13.96	4.0049	5.9486	0.1583	0.0017	0.3409	-25.93
Shin	5.9197	5.9237	5.7424	0.1742	0	-32.48	5.6918	5.6949	5.4941	0.1460	0	-56.07
UA	0.0508	0.1104	0.0331	0.6995	0.7030	**-67.80**	0.0203	0.0814	0.0151	0.8768	0.9152	**-71.44**
ManiPoP	1.5112	1.5307	1.4217	0	0	X	1.1245	1.3543	0.9233	0	0.2969	X
CASPI	0.2715	1.7204	0.0394	0.3029	0.9653	X	**0.0129**	0.0811	**0.0035**	0.9643	0.9855	X
Proposed	**0.0174**	**0.0577**	**0.0041**	0.9265	0.9698	-63.52	0.0145	0.0491	0.0039	0.9506	0.9816	-64.82
Reindeer												
ML	4.8382	6.4860	0.3463	0.0014	0.1904	-13.94	3.9673	5.8627	0.1667	0.0018	0.3324	-25.95
Shin	5.7154	5.7185	5.6359	0.1331	0	-32.48	5.4885	5.4927	5.4358	0.1272	0	-56.06
UA	0.0659	0.1534	0.0276	0.5300	0.6870	**-63.31**	0.0357	0.1045	0.0157	0.6505	0.8212	**-67.36**
ManiPoP	1.6191	1.6877	1.6371	0	0.0670	X	1.4286	1.6030	1.3496	0	0.1979	-53.88
CASPI	0.4951	2.3115	0.3374	0.0509	0.9272	X	0.0408	0.4801	0.0047	0.5392	**0.9748**	X
Proposed	**0.0195**	**0.0527**	**0.0042**	0.9117	0.9466	-63.03	**0.0166**	**0.0422**	**0.0037**	0.9423	0.9621	-64.45
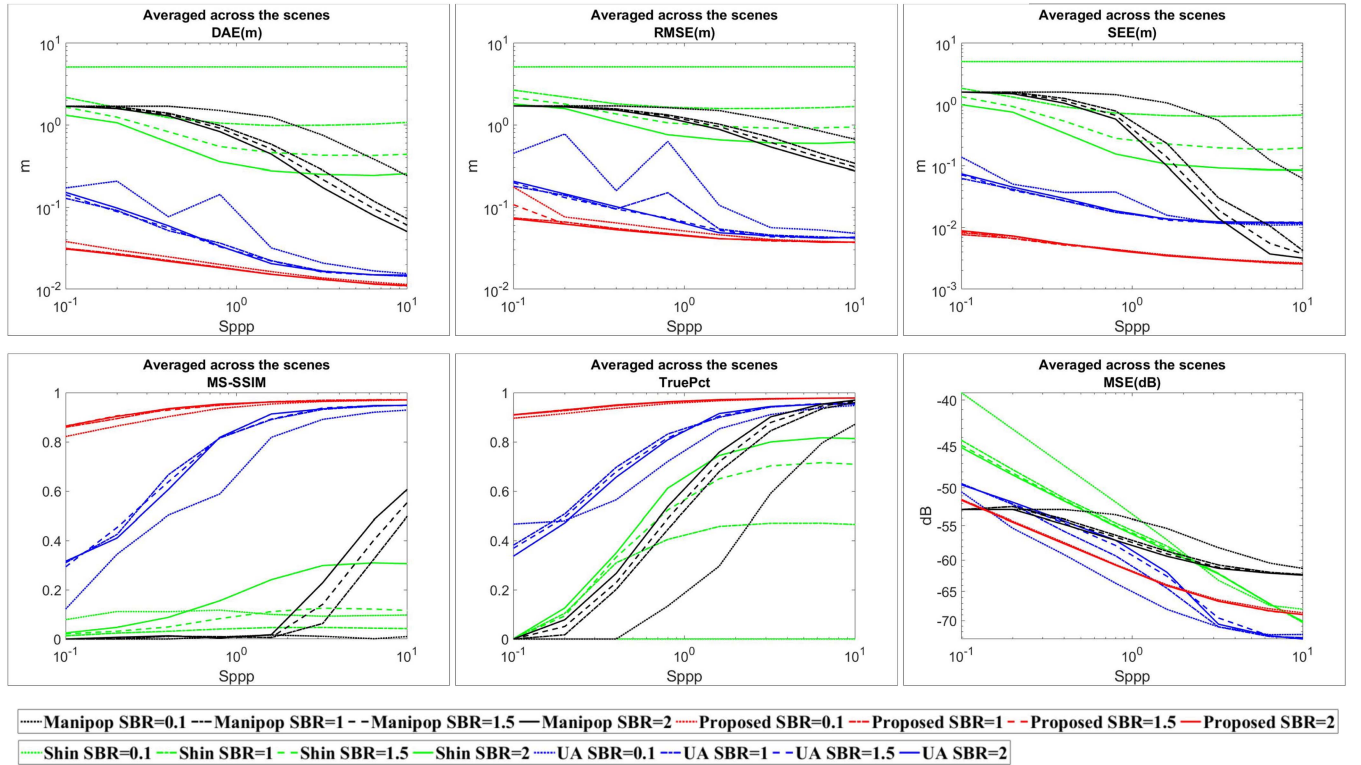
Fig. 8. Performance evaluation for simulated data with SBR values of {0.1, 1, 1.5, 2} across varying Sppp. The metric curves represent averaged results across the Art, Bowling, and Reindeer scenes.
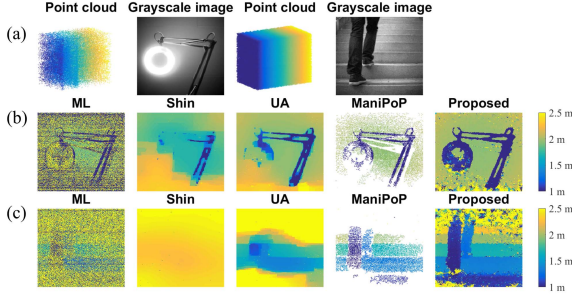
Fig. 9. Reconstructed images of the real experimental scene from [16]. (a) shows the original point clouds and visible grayscale images of the scenes. (b) shows the reconstructed depth images of the indoor lamp scene (sparse counting case). (c) shows the reconstructed depth images of the outdoor stairs scene under strong ambient light (strong noise counting case).
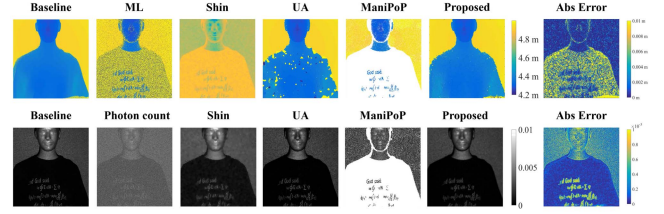

Fig. 10. Reconstructed images and absolute error images of the mannequin scene from [41]. The top row displays the depth baseline and reconstructed depth images, while the bottom row shows the intensity baseline and reconstructed intensity images.

TABLE III
QUANTITATIVE COMPARISON ON MANNEQUIN SCENE WITH

	DAE (m)	RMSE (m)	SEE (m)	MS-SSIM	True pct	MSE (dB)
Mannequin						
ML	2.5904	4.0710	0.5962	0.0250	0.4027	-29.49
Shin	2.6242	2.6715	2.3807	0.5669	0	-57.38
UA	0.1692	1.0239	0.9862	0.4583	0.9468	-54.80
Mani PoP	2.7462	3.4935	4.1728	0.0369	0.3828	-42.45
Propo sed	**0.0281**	**0.1834**	**0.2088**	**0.9770**	**0.9810**	**-58.38**

in [16] and [41]. When reconstructing real scenes, the proposed algorithm utilizes the same parameters as those applied in the simulation data tests described above.

In the dataset provided by [16], the indoor lamp scene is selected to represent the sparse count case (satisfies the sparse counting condition [6]), and the outdoor stairs scene with strong ambient light is used to represent strong noise counting case (strong noise environment [11], [16]). The count data cubes have the resolution of $256 \times 256 \times 1536$, Fig. 9 shows the raw point clouds, visible greyscale images and the reconstructed depth images of the lamp and stairs scenes. Since the grayscale images in the dataset do not serve as intensity references, we did not perform intensity image reconstruction. Therefore, the reconstruction results are not quantitatively analyzed due to the lack of reference images in the dataset.

As shown in the lamp scene, there is lots of noise in the reconstruction results of ML algorithm, and the results of Shin algorithm are blurred and lose most of the detail information. The UA algorithm successfully reconstructs most of the lamp's structure, but it fails to fully capture the profile of the ring area. The reconstruction result of the ManiPoP algorithm contains many pixels that failed to be reconstructed. In contrast, the proposed algorithm reconstructs the lamp and its bracket more completely, preserving clear edge features.

In the stairs scene, the detection data includes a significant amount of strong ambient light noise counts. While the ManiPoP algorithm manages to reveal some scene features, many pixels remain unreconstructed. In contrast, the proposed algorithm effectively reconstructs the scene, preserving the structural integrity and details, whereas the other algorithms fail to produce effective reconstructions, highlighting the robustness and efficacy of the proposed method in handling real-world noisy conditions.

The proposed algorithm achieves accurate depth reconstruction in both real scenes, demonstrating its adaptability to sparse count and strong noise count data, as well as its higher accuracy compared to other methods.

In addition, the experimental data of the Mannequin scene provided by [41], which includes high SBR baseline data, is used for quantitative analysis of reconstruction effects. The count data cubes have a resolution of $1000 \times 1000 \times 12500$, and the reference images (baseline images) are generated from the high

SBR detection data of the scene. The reconstruction results are presented in Fig. 10, and the corresponding evaluation metrics are presented in Table III. The reconstruction results of the traditional ML algorithm exhibit significant noise, particularly in areas of the mannequin's clothing where reflectivity is low, leading to nearly complete failure in depth reconstruction and resulting in large DAE and RMSE values. The Shin algorithm reconstructs depth beyond the mannequin's range, resulting in high overall error. The ManiPoP algorithm produces reconstruction results with large blank areas (white areas), particularly failing in low reflectivity clothing areas. The UA algorithm achieves better reconstruction accuracy than the aforementioned algorithms; however, there are noticeable false plaques in low reflectivity and edge areas within the depth image, contributing to larger overall and edge errors. In contrast, the proposed algorithm effectively reconstructs the depth image, achieving complete reconstruction in low reflectivity areas, resulting in lower DAE and RMSE values, and achieves clear and continuous edge reconstruction, as indicated by lower SEE values. This demonstrates the algorithm's improved robustness and precision in challenging conditions.

For intensity image reconstruction, traditional methods relying on photon counts are severely affected by noise. While the other algorithms show similar results, and the proposed algorithm achieves slightly better MSE.

IV. CONCLUSION

In this paper, an adaptive photon count reconstruction algorithm is proposed for sparse and strong noise count data with low SBR. The algorithm introduces a histogram enhancement method based on Gaussian weighting, which complements the

echo count data and improves reconstruction accuracy, particularly at scene edges. Given that echo counts are primarily aggregated within a time range corresponding to the pulse width, an aggregation parameter is proposed to enable adaptive separation of echo and noise regions. To ensure compatibility and accuracy in both sparse and strong noise counting scenarios, the objective function is constructed based on the degree of count differences, enabling pixel-wise optimization of both depth and intensity without requiring external prior information. The proposed algorithm demonstrates significant performance improvements across simulated and real experimental datasets. Specifically, when both Sppp and SBR are low, the algorithm reduces overall reconstruction errors by approximately 50%. It achieves a high proportion of correctly reconstructed pixels—up to 90% when Sppp is as low as 0.1 (SBR = 0.1)—and reduces edge reconstruction errors by an order of magnitude compared with existing methods. This demonstrates the superior performance of the proposed algorithm in preserving structural integrity and reducing reconstruction errors for count data under low SBR conditions. Furthermore, robust experimental results validate the algorithm's efficacy in real-world scenarios, underscoring its practical applicability and reliability in diverse and challenging environments. The algorithm shows great potential for improving long-range detection, low-reflectivity detection, foliage target detection, all-day detection, and underwater target detection.

V. FUTURE WORK

The proposed algorithm yields effective reconstruction results; however, its runtime performance requires further optimization. As indicated by the runtime comparisons in Section IX of the supplemental material, the proposed algorithm's runtime is approximately 1.8 times that of the UA algorithm and significantly lags behind neural network-based approaches in terms of reconstruction time. This discrepancy arises from the fact that the algorithm has not yet been optimized for parallel computation. The use of a sliding window to traverse the histogram twice at each pixel is computationally intensive. To address this, future work will explore parallel computing techniques, as the processing of different sections of the histogram data is independent and could be significantly accelerated through parallelization. Moreover, our current method focuses on localizing a single echo by identifying the depth value that maximizes the objective function during optimization. However, since the optimization formula is designed based on the degree of count differences within the time neighborhood, echoes within their neighborhoods tend to exhibit higher count levels or temporal clustering compared to noise. This characteristic suggests that the optimization formula could potentially identify multiple echoes simultaneously. Therefore, future research will focus on optimizing depth reconstruction for scenarios involving multiple echo returns, building on the current objective function formulation.

REFERENCES

[1] Y. J. Zhao et al., "Forest species diversity mapping using airborne LiDAR and hyperspectral data in a subtropical forest in China," *Remote Sens. Environ.*, vol. 213, pp. 104–114, Aug. 2018, doi: 10.1016/j.rse.2018.05.014.

[2] T. Hakala, J. Suomalainen, S. Kaasalainen, and Y. Chen, "Full waveform hyperspectral LiDAR for terrestrial laser scanning," *Opt. Exp.*, vol. 20, no. 7, pp. 7119–7127, Mar. 2012, doi: 10.1364/OE.20.007119.

[3] S. Kaasalainen, T. Malkamäki, J. Ilinea, and L. Ruotsalainen, "Multispectral terrestrial laser scanning: New developments and applications," in *Proc. 2018 IEEE Int. Geosci. Remote Sens. Symp.*, Valencia, Spain, 2018, pp. 8737–8740, doi: 10.1109/IGARSS.2018.8517590.

[4] D. Kukolj, I. Marinović, and S. Nemet, "Road edge detection based on combined deep learning and spatial statistics of LiDAR data," *J. Spatial Sci.*, vol. 68, no. 2, pp. 245–259, Aug. 2023, doi: 10.1080/14498596.2021.1960912.

[5] K. L. Lim, T. Drage, and T. Bräunl, "Implementation of semantic segmentation for road and lane detection on an autonomous ground vehicle with LIDAR," in *Proc. 2017 IEEE Int. Conf. Multisensor Fusion Integration Intell. Syst.*, Daegu, South Korea, 2017, pp. 429–434, doi: 10.1109/MFI.2017.8170358.

[6] 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, Mar. 2019, doi: 10.1137/18M1183972.

[7] A. Halimi, P. Ciuciu, A. McCarthy, S. McLaughlin, and G. S. Buller, "Fast adaptive scene sampling for single-photon 3D lidar images," in *Proc. 2019 IEEE 8th Int. Workshop Comput. Adv. Multi-Sensor Adaptive Process.*, Le gosier, Guadeloupe, 2019, pp. 196–200, doi: 10.1109/CAM-SAP45676.2019.9022519.

[8] Z. P. Li et al., "Single-photon computational 3D imaging at 45 km," *Photon. Res.*, vol. 8, no. 9, pp. 1532–1540, Sep. 2020, doi: 10.1364/PRJ.390091.

[9] Z. P. Li et al., "Single-photon imaging over 200 km," *Optica*, vol. 8, no. 3, pp. 344–349, Mar. 2021, doi: 10.1364/OPTICA.408657.

[10] 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, May 2017, doi: 10.1364/OE.25.011919.

[11] K. Drummond, S. McLaughlin, Y. Altmann, A. Pawlikowska, and R. Lamb, "Joint surface detection and depth estimation from single-photon lidar data using ensemble estimators," in *Proc. 2021 Sensor Signal Process. Defence Conf.*, Edinburgh, U.K., 2021, pp. 1–5, doi: 10.1109/SSPD51364.2021.9541522.

[12] R. A. Barton-Grimley, R. A. Stillwell, and J. P. Thayer, "High resolution photon time-tagging lidar for atmospheric point cloud generation," *Opt. Exp.*, vol. 26, no. 20, pp. 26030–26044, Oct. 2018, doi: 10.1364/OE.26.026030.

[13] Y. Jiang et al., "Photon counting lidar working in daylight," *Opt. Laser Technol.*, vol. 163, Aug. 2023, Art. no. 109374, doi: 10.1016/j.optlastec.2023.109374.

[14] G. Chen, C. Wiede, and R. Kokozinski, "Data processing approaches on SPAD-based d-TOF LiDAR systems: A review," *IEEE Sensors J.*, vol. 21, no. 5, pp. 5656–5667, Mar. 2021, doi: 10.1109/JSEN.2020.3038487.

[15] R. Tobin et al., "Long-range depth profiling of camouflaged targets using single-photon detection," *Opt. Eng.*, vol. 57, no. 3, Dec. 2017, Art. no. 031303, doi: 10.1117/1.OE.57.3.031303.

[16] D. B. Lindell, M. O'Toole, and G. Wetzstein, "Single-photon 3D imaging with deep sensor fusion," *ACM Trans. Graph.*, vol. 37, no. 4, Jul. 2018, Art. no. 113, doi: 10.1145/3197517.3201316.

[17] A. Kirmani et al., "First-photon imaging," *Science*, vol. 343, no. 6166, pp. 58–61, 2014, doi: 10.1126/science.1246775.

[18] A. Halimi, R. Tobin, A. McCarthy, J. Bioucas-Dias, S. McLaughlin, and G. S. Buller, "Robust restoration of sparse multidimensional single-photon LiDAR images," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 138–152, 2020, doi: 10.1109/TCI.2019.2929918.

[19] 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. Imag.*, vol. 1, no. 2, pp. 112–125, Jun. 2015, doi: 10.1109/TCI.2015.2453093.

[20] 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, doi: 10.1109/TIP.2016.2526784.

[21] J. Tachella, Y. Altmann, S. McLaughlin, and J. Y. Tourneret, "Fast surface detection in single-photon lidar waveforms," in *Proc. 27th Eur. Signal Process. Conf.*, A Coruna, Spain, 2019, pp. 1–5, doi: 10.23919/EU-SIPCO.2019.8903062.

[22] J. Rapp and V. K. Goyal, "A few photons among many: Unmixing signal and noise for photon-efficient active imaging," *IEEE Trans. Comput. Imag.*, vol. 3, no. 3, pp. 445–459, Sep. 2017, doi: 10.1109/TCI.2017.2706028.

[23] R. Tobin et al., "Three-dimensional single-photon imaging through obscurants," *Opt. Exp.*, vol. 27, no. 4, pp. 4590–4611, Feb. 2019, doi: 10.1364/OE.27.004590.

[24] 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, May 2021, Art. no. 11236, doi: 10.1038/s41598-021-90587-8.

[25] X. F. Wang, T. Y. Zhang, Y. Kang, W. W. Li, and J. T. Liang, "High-flux fast photon-counting 3D imaging based on empirical depth error correction," *Photonics*, vol. 10, no. 12, Nov. 2023, Art. no. 1034, doi: 10.3390/photonics10121304.

[26] S. Plosz, A. Maccarone, S. McLaughlin, G. S. Buller, and A. Halimi, "Real-time reconstruction of 3D videos from single-photon LiDaR data in the presence of obscurants," *IEEE Trans. Comput. Imag.*, vol. 9, pp. 106–119, 2023, doi: 10.1109/TCI.2023.3241547.

[27] L. Ma, J. F. Sun, D. Liu, and X. Zhou, "Noise-suppression algorithm of GM-APD lidar based on Markov random field," *Opt. Laser Technol.*, vol. 169, Feb. 2024, Art. no. 110026, doi: 10.1016/j.optlastec.2023.110026.

[28] M. Patting, M. Wahl, P. Kapusta, and R. Erdmann, "Dead-time effects in TCSPC data analysis," in *Proc. SPIE*, vol. 6583, 2007, Art. no. 658307, doi: 10.1117/12.722804.

[29] J. Rapp, Y. T. Ma, R. M. A. Dawson, and V. K. Goyal, "High-flux single-photon lidar," *Optica*, vol. 8, no. 1, pp. 30–39, Jan. 2021, doi: 10.1364/OPTICA.403190.

[30] J. Rapp, Y. Ma, R. M. A. Dawson, and V. K. Goyal, "Predicting dead time distortion for high-flux single-photon lidar," in *Proc. 2020 Conf. Lasers Electro-Opt.*, San Jose, CA, USA, 2020, pp. 1–2.

[31] J. Koo, A. Halimi, A. Maccarone, G. S. Buller, and S. McLaughlin, "A Bayesian based unrolling approach to single-photon lidar imaging through obscurants," in *Proc. 30th Eur. Signal Process. Conf.*, Belgrade, Serbia, 2022, pp. 872–876, doi: 10.23919/EUSIPCO55093.2022.9909749.

[32] X. C. Zhao et al., "Photon-efficient 3D reconstruction employing a edge enhancement method," *Opt. Exp.*, vol. 30, no. 2, pp. 1555–1569, Jan. 2022, doi: 10.1364/OE.446369.

[33] W. Liu, X. Chen, J. Yang, and Q. Wu, "Robust color guided depth map restoration," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 315–327, Jan. 2017, doi: 10.1109/TIP.2016.2612826.

[34] X. Liu, D. Zhai, R. Chen, X. Ji, D. Zhao, and W. Gao, "Depth super-resolution via joint color-guided internal and external regularizations," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1636–1645, Apr. 2019, doi: 10.1109/TIP.2018.2875506.

[35] C. Chen, X. Chen, and H. Cheng, "On the over-smoothing problem of CNN based disparity estimation," in *Proc. 2019 IEEE/CVF Int. Conf. Comput. Vis.*, Seoul, South Korea, 2019, pp. 8996–9004, doi: 10.1109/ICCV.2019.00909.

[36] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986, doi: 10.1109/TPAMI.1986.4767851.

[37] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, 2003, vol. 2, pp. 1398–1402, doi: 10.1109/ACSSC.2003.1292216.

[38] D. Scharstein and C. Pal, "Learning conditional random fields for stereo," in *Proc. 2007 IEEE Conf. Comput. Vis. Pattern Recognit.*, Minneapolis, MN, USA, 2007, pp. 1–8, doi: 10.1109/CVPR.2007.383191.

[39] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *Proc. 2007 IEEE Conf. Comput. Vis. Pattern Recognit.*, Minneapolis, MN, USA, 2007, pp. 1–8, doi: 10.1109/CVPR.2007.383248.

[40] S. Hernandez-Marin, A. M. Wallace, and G. J. Gibson, "Bayesian analysis of lidar signals with multiple returns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 12, pp. 2170–2180, Dec. 2007, doi: 10.1109/TPAMI.2007.1122.

[41] D. Venkatraman, "Data repository for photon-efficient pseudo-array imaging," 2015. [Online]. Available: https://github.com/photon-efficient-imaging/sample-data

[42] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from RGBD images," in *Proc. 12th Eur. Conf. Comput. Vis.*, Florence, Italy, 2012, pp. 746–760.

[43] A. Gupta, A. Ingle, and M. Gupta, "Asynchronous single-photon 3D imaging," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Seoul, South Korea, Oct. 2019, pp. 7908–7917, doi: 10.1109/ICCV.2019.00800.

[44] P. Morris, R. Aspden, J. Bell, R. Boyd, and M. Padgett, "Imaging with a small number of photons," *Nat. Commun.*, vol. 6, Jan. 2015, Art. no. 5913, doi: 10.1038/ncomms6913.

[45] 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, p. 4984, Nov. 2019, doi: 10.1038/s41467-019-12943-7.

[46] J. Peng, Z. Xiong, H. Tan, X. Huang, Z.-P. Li, and F. Xu, "Photon-efficient 3D imaging with a non-local neural network," in *Proc. Comput. Vis. – ECCV 2020: 16th Eur. Conf.*, Nov. 2020, pp. 225–241.

[47] J. Peng, Z. Xiong, H. Tan, X. Huang, Z.-P. Li, and F. Xu, "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, doi: 10.1109/TPAMI.2022.3200745.

[48] J. Lee et al., "CASPI: Collaborative photon processing for active single-photon imaging," *Nat. Commun.*, vol. 14, no. 1, May 2023, Art. no. 3158, doi: 10.1038/s41467-023-38893-9.

Meijun Chen received the B.Eng. degree in electronic information engineering from the Beijing University of Chemical Technology, Beijing, China, in 2022. He is currently working toward the M.S. degree with the Southwest Institute of Technical Physics, Chengdu, China. His research interests include single-photon LiDAR imaging and intelligent signal and information processing.
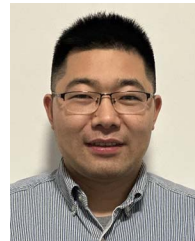
Zhendong Shi received the Ph.D. degree in optical engineering from the Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu, China, in 2015. He is currently a Senior Research Engineer with the Southwest Institute of Technical Physics, Chengdu. His research interests include airborne LiDAR, multispectral LiDAR, oceanographic LiDAR, and single-photon imaging.

Wei Chen received the M.S. degree in optical engineering from Sichuan University, Chengdu, China, in 2008. He is currently a Research Professor and the Assistant President with the Southwest Institute of Technical Physics, Chengdu. He is also the Director of the LiDAR and Device Laboratory in Sichuan Province, China. His research interests include laser technology, photoelectric countermeasure techniques, and multispectral LiDAR.

Fangjie Xu received the B.Eng. degree from the Wuhan University of Technology, Wuhan, China, in 2023. He is currently working toward the M.S. degree with the Southwest Institute of Technical Physics, Chengdu, China. His research interests include single-photon LiDAR denoising, Linear-mode LiDAR denoising, and Airborne three-wavelength LiDAR system.

Yong Jiang, photograph and biography not available at the time of publication.

Yijiang Mao, photograph and biography not available at the time of publication.

Shiyue Xu, photograph and biography not available at the time of publication.

Bowen Chen, photograph and biography not available at the time of publication.

Yalan Wang, photograph and biography not available at the time of publication.

Zecheng Wang, photograph and biography not available at the time of publication.

Jie Leng, photograph and biography not available at the time of publication.