Optics Letters

Fast non-line-of-sight imaging based on first photon event stamping

Zhupeng Li,[1,2] Xintong Liu,[3] Jianyu Wang,[3] Zuoqiang Shi,[4,5,6] Lingyun Qiu,[3,5,7] and Xing Fu[1,2,8]

[1]State Key Laboratory of Precision Measurement Technology and Instruments, Department of Precision Instruments, Tsinghua University, Beijing 100084, China
[2]Key Laboratory of Photonic Control Technology (Tsinghua University), Ministry of Education, Beijing 100084, China
[3]Yau Mathematical Sciences Center, Tsinghua University, Beijing 100084, China
[4]Department of Mathematical Sciences, Tsinghua University, Beijing 100084, China
[5]Yanqi Lake Beijing Institute of Mathematical Sciences and Applications, Beijing 101408, China
[6]e-mail: zqshi@tsinghua.edu.cn
[7]e-mail: lyqiu@tsinghua.edu.cn
[8]e-mail: fuxing@mail.tsinghua.edu.cn

Received 19 October 2021; revised 20 January 2022; accepted 10 March 2022; posted 10 March 2022; published 4 April 2022

Non-line-of-sight (NLOS) imaging enables people to see a hidden scene based on multiple interaction information between the object and the carrier. There have been numerous studies focusing on the physical modeling of photon scattering, but few have explored the detection process, which also plays a vital role. In this paper, we put forward a novel, to the best of our knowledge, detection methodology for NLOS imaging based on time-sequential first photon (TSFP) data. We verify the method with both synthetic and experimental data, showing a dramatic reduction in acquisition time cost compared with traditional methods for the same reconstruction quality. This work may contribute to real-time and photon-starved NLOS imaging for practical applications.

https://doi.org/10.1364/OL.446079

Many solutions for non-line-of-sight (NLOS) imaging have emerged and have attracted lots of attention for potential applications, such as automatic driving, object tracking, and surveillance. In the optics domain, NLOS imaging can be interpreted as a photon-scattering process based on geometric optics [1–5] or wave propagation approaches [6,7]. In geometric perspective, the hidden curved surface can be reconstructed based on solving normal information [8–10] and occlusion consideration [10]. In mathematics, the NLOS problem can be viewed as an inverse problem, and there are many algorithm innovations that aim to lower the time and space complexity for possible real-time applications, such as the light-cone transform (LCT) [4], f-k migration [6], fast backprojection [11], phasor field [7,12] and deep-learning-based approaches [13]. For precise reconstructions, signal-object collaborative regularization [14] was recently reported to show high robustness against noise.

The detection process in NLOS imaging is also crucial for signal acquisition in practice. A classical approach is to scan a laser, illuminating positions on the visible wall, and capture temporal signals that contain information on the hidden scene. To record the temporal signals, time-of-flight (TOF) measurements together with time-correlated single-photon-counting (TCSPC) technology is widely adopted. This involves registering every photon event with a specific arrival time stamp based on a common start time at picosecond resolution and with high sensitivity for weak light detection, relying on a Geiger-mode single-photon avalanche diode (SPAD). Until now, most methods have employed the accumulated data from the photon event histogram (PEH) method for reconstruction in NLOS imaging, which requires a long and fixed measurement time at each scanning position to achieve an acceptable signal-to-noise ratio (SNR).

In contrast, to realize fast measurement, one can operate the TCSPC device in time-tagged time-resolved (TTTR) mode to record every photon event arrival time sequentially in time, and can adaptively stop the measurement whenever enough information has been acquired for reconstruction, leading to a much more efficient detection process. In this way, first-photon imaging [15] was first demonstrated in a line-of-sight (LOS) scenario in which the measured first photon event is contributed by only one voxel from the scene, thus allowing the albedo of each illuminated point to be solved using only one observation of the first photon event.

However, the application of first-photon imaging in a NLOS setup is extremely challenging. First, different voxels in the hidden scene with equal propagation distances, and even different propagation distances (due to the temporal broadening that occurs in practical measurements), may contribute simultaneously to the measured signal in the same time bin. These contributions cannot be distinguished by a single observation of the first photon event. Second, the signal received using the first-photon method is noisier than that obtained using the traditional accumulation method because the parameter estimation for each time bin is based on only one time measurement. Furthermore, the NLOS problem is naturally ill posed, as the same measured signals may correspond to multiple reconstructions. To achieve fast NLOS measurement based on the first-photon

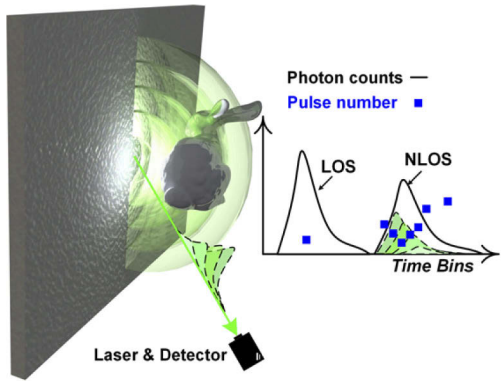

Fig. 1. Comparison of first-photon detection methods for the LOS and NLOS cases. The solid-line waveforms are the ideal photon intensity signals, and the dashed waveforms are contributions from different spherical shell regions. The squares are the first photon event measurements.

method and mitigate the reconstruction difficulties from noisy data and an ill-posed operator, in this paper we put forward a time-sequential first photon (TSFP) method that employs first-photon data from several time bins for reconstruction in the time stream of each scanning position measurement. For the proposed detection process, the acquisition time is so short that the variance of the histogram signal is outside the tolerance of conventional PEH methods.

Figure 1 compares TOF signal measurement in the confocal case between the LOS and the NLOS scenarios at the same scanning position. The temporal waveform of the echo in NLOS imaging is sampled by several first-photon events distributed over different time bins, which are more likely to carry the most significant information for reconstructing the hidden scene than the subsequent first-photon events (received by other time bins).

In [16], the authors take dead time into account and reconstruct the depth and reflectance of the LOS scene by solving a Bayesian model with PEH data. In the case of NLOS imaging, the dead time can be neglected due to the low photon flux detected after multiple diffuse reflections. In this work, we apply the maximum likelihood estimation (MLE) method for NLOS reconstruction under extremely low-dose acquisition conditions by making use of the TSFP detection process.

We adopt the physical model introduced in the LCT method [4]:

$$\tau(z,t,\alpha) = b + \int_{\Omega} \frac{\alpha(\boldsymbol{x})}{||\boldsymbol{x}-z||^4}\delta(ct-2||\boldsymbol{x}-z||)\mathrm{d}\boldsymbol{x}, \tag{1}$$

where $\tau(z,t,\alpha)$ is the photon intensity detected at position z on the visible wall at time t, $b \neq 0$ is a constant for the intensity value of slow-varying background noise, $\alpha(\boldsymbol{x})$ is the albedo value at point $\boldsymbol{x}$ in the hidden scene, δ is the Dirac delta function, c is the speed of light, and Ω is the reconstruction domain.

To detect the photon intensity, every focal point z is illuminated by a series of laser pulses. For a specified scanning position and time bin, the detection process for photon events is considered as a sequence of Bernoulli trials. The probability of a failure outcome is modeled as

$$p(\tau) = \exp(-\eta\tau), \tag{2}$$

where η is the detection efficiency. The probability of detecting the first photon at the kth pulse is

$$P(\tau) = p(\tau)^{k-1}[1-p(\tau)]. \tag{3}$$

The set of scanning positions is denoted $\{z_i\}_{i=1}^{S}$ and the time is discretized with $\{t_j = j\Delta t\}_{j=1}^{T}$, in which $i = 1, \ldots, S$ is the index of scanning positions, $j = 1, \ldots, T$ is the index of time bins, and Δt is the time resolution. The probability of failing to detect a photon event at z_i and t_j is

$$p_{ij}(\alpha) = \exp(-\eta\tau(z_i,t_j,\alpha)). \tag{4}$$

By assuming that the detection process of photon events is independent among all scanning positions and time bins, the joint probability of detecting the first photon at point z_i and time t_j with a pulse number of k_{ij} is written as

$$P(\alpha) = \prod\nolimits_{i,j} p_{ij}^{k_{ij}-1}(\alpha)[1-p_{ij}(\alpha)]. \tag{5}$$

In the case of low light flux, we have $\eta\tau \approx 0$, and $1 - p_{ij}(\alpha) \approx \eta\tau(z_i,t_j,\alpha)$. Using the maximum likelihood principle, we obtain the following negative log-likelihood function:

$$L(\alpha|\boldsymbol{k}) = \sum\nolimits_{i,j}(k_{ij}-1)[\eta\tau(z_i,t_j,\alpha)] - \ln[\eta\tau(z_i,t_j,\alpha)]. \tag{6}$$

Regularization can be used to improve the reconstruction quality. We choose the L_1 and total variation (TV) regularizations to enhance the sparseness and smoothness of the reconstructions. The reconstruction can be obtained by solving

$$\alpha^* = \arg\min_{\alpha} L(\alpha|\boldsymbol{k}) + \lambda_1||\alpha||_1 + \lambda_2 TV(\alpha). \tag{7}$$

This optimization problem is convex when $\lambda_1 = \lambda_2 = 0$, and is solved by the gradient descent method, as Fig. 2 shows. For general conditions, the optimization problem in Eq. (7) can be solved using the split Bregman method [17] and the FISTA algorithm [18]. The goal of our fast NLOS reconstruction is to find the function $\alpha(\boldsymbol{x})$ that matches measurement $\boldsymbol{k}$. In practice, $\alpha(\boldsymbol{x})$ is discretized with voxels in the domain Ω.

We verify our method with synthetic and experimental data. To generate synthetic data, we employ the signals of Stanford Bunny and USAF from the Zaragoza dataset [19]. The original data are scaled and a constant dark count is added. Random samples of the photon event data are generated using Eq. (2). For each focal point, when a sufficient number of time bins (we choose 10, 50, 75, 100, and 200 in the simulations) receive the first photon, we stop the simulation and record the pulse number, the TSFP data, and the PEH data. By assuming the repetition rate of the laser to be 10 MHz, the acquisition time is computed as the product of the pulse number and pulse period.

In our simulation, the visible wall is sampled at 64×64 focal points and the average photon events per pulse are 0.0046 and 0.0051, respectively, for the instances of the Stanford Bunny

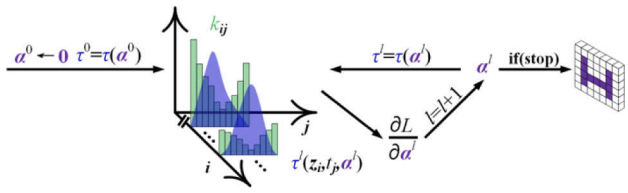

Fig. 2. Sketch of the optimization process. l is the index of iterations.

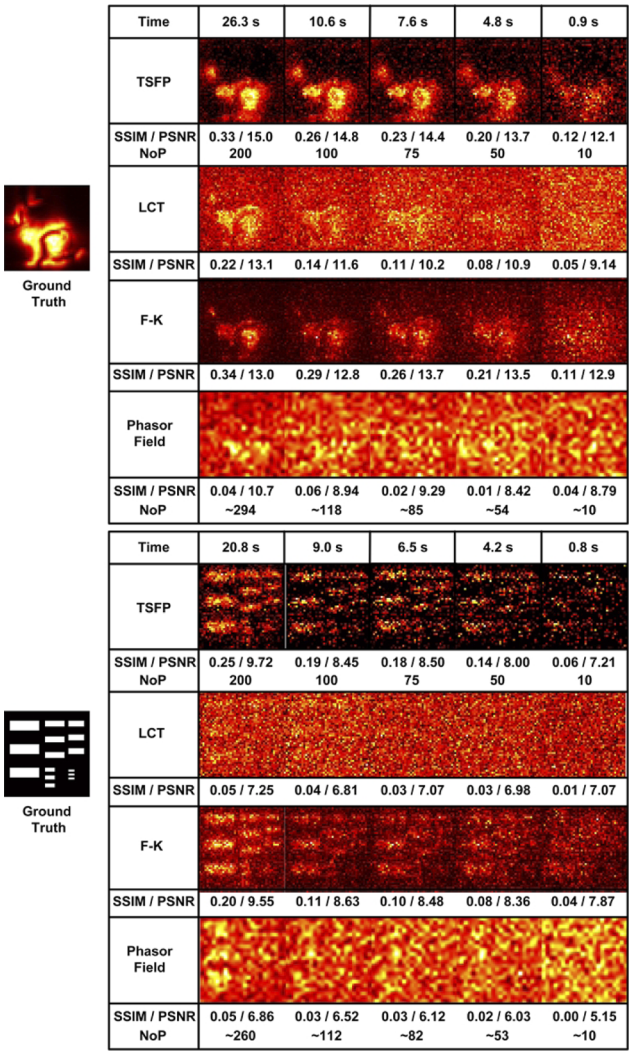

Fig. 3. Reconstruction results for synthetic data from Stanford Bunny (top) and USAF (bottom).

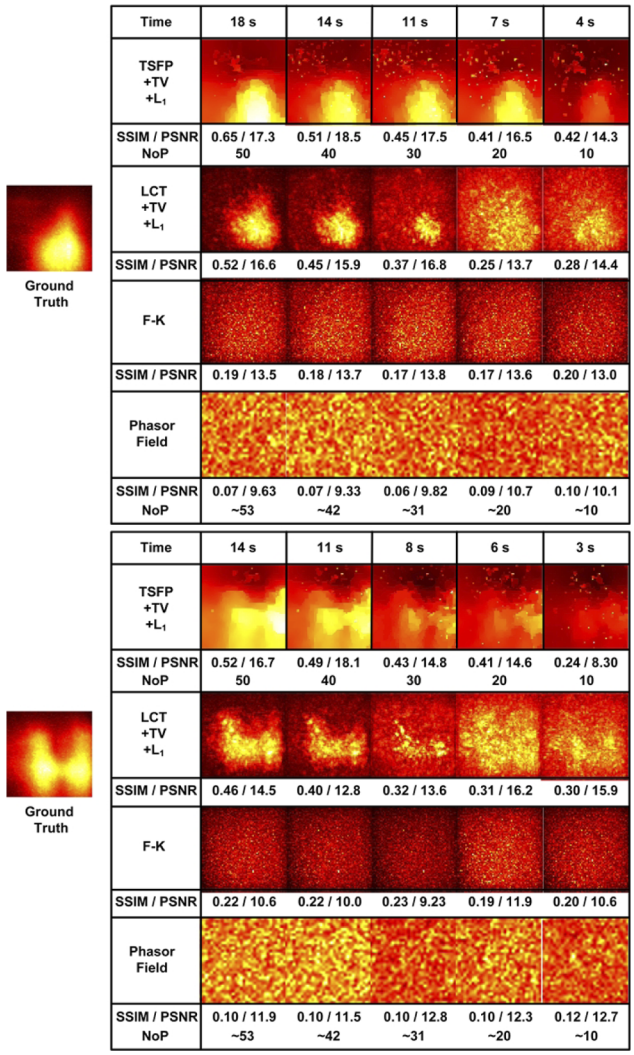

Fig. 4. Experimental reconstructions of the triangle (top) and a letter H (bottom).

and USAF. For the TSFP method, we use all the first-arriving photon events from several time bins to reconstruct the scene. For the PEH method, we apply the LCT solver [4], f-k solver [6], and phasor field solver [12] for reconstruction. To achieve fair comparisons, we do not include regularizations in all methods.

To evaluate the reconstruction quality, the structural similarity (SSIM) [20] and signal-to-noise ratio (PSNR) are introduced for maximum intensity projection of the reconstructions. In Fig. 3, the reconstruction results for the simulated data are shown in order of descending number of photon events per scanning position (NoP) from left to right. One can see from Fig. 3 that the TSFP method has much better reconstruction quality than the PEH methods at the same acquisition time, especially in the photon-starved situation. To put it another way, the TSFP method requires far less acquisition time than the PEH methods to achieve the same reconstruction quality.

Next, we investigate the experimental performance of the presented method. The NLOS experimental setup adopts confocal laser scanning, and each scanning position on the visible wall serves as a relay of the temporal signal. The TCSPC device is PicoHarp300 (PicoQuant), with a minimum time resolution of 4 ps. The temporal response of the overall system is measured as about 200 ps full width at half maximum (FWHM). A 515 nm pulsed laser (ROI) with a pulse width of 17 ps, a pulse energy of 0.8 nJ, and a repetition rate of 10 MHz scans a 50×50 cm^2 area on the visible wall at equal intervals. The reflected photons from the visible wall and the hidden scene are collected by a single-pixel SPAD (Micro Photon Devices). The hidden objects we use are custom-made paperboard in the shape of a triangle and a letter H.

The ground-truth images used for SSIM and PSNR evaluations are generated with the LCT + TV + L_1 method [4] using the overall measured PEH data, with an accumulation time of 1 s for each scanning point. The reconstruction results from LCT + TV + L_1, f-k, the phasor field method, and TSFP + TV + L_1 are compared in Fig. 4. The parameters in the LCT + TV + L_1 method are fine-tuned. It is shown that the proposed method outperforms traditional PEH-based methods and has the best visual quality.

Our algorithm requires iteration and takes longer than the direct imaging method [4]. In our iteration scheme, the majority of the time is spent on the gradient descent step, which calls for forward model computation. To reduce the reconstruction time, it is possible to optimize the code with time-efficient

implementations of the forward operator. Combining our TSFP approach with deep-learning techniques [13] for fast and accurate reconstructions would be an interesting direction for future research.

In conclusion, we have proposed a new detection method for NLOS imaging in a photon-limited environment. Compared with traditional PEH methods that use a fixed measurement time, the TSFP method is a detection framework that has a higher photon event efficiency and requires less time. In NLOS imaging, we carried out several synthetic and practical experiments and verified that the TSFP method leads to a significant reduction in acquisition time cost, compared with the PEH method, to achieve the same reconstruction quality. Our method only requires a small number of photon events for NLOS reconstruction, and the acquisition time may be further reduced with SPAD arrays. These advantages of the TSFP framework and its possible extensions have great potential to facilitate the development of practical NLOS imaging systems for real-time and photon-starved applications.

Funding. National Natural Science Foundation of China (11971258, 12071244, 61975087).

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

REFERENCES

1. T. Hutchison Kirmani, J. Davis, and R. Raskar, in *IEEE 12th International Conference on Computer Vision (ICCV)* (2009), pp. 159–166.
2. T. W. Velten, O. Gupta, A. Veeraraghavan, M. G. Bawendi, and R. Raskar, Nat. Commun. **3**, 745 (2012).
3. M. Buttafava, J. Zeman, A. Tosi, K. Eliceiri, and A. Velten, Opt. Express **23**, 20997 (2015).
4. M. O'Toole, D. B. Lindell, and G. Wetzstein, Nature **555**, 338 (2018).
5. S. Xin, S. Nousias, K. N. Kutulakos, A. C. Sankaranarayanan, S. G. Narasimhan, and I. Gkioulekas, in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019), pp. 6793–6802.
6. D. B. Lindell, G. Wetzstein, and M. O'Toole, ACM Trans. Graph. **38**, 1 (2019).
7. X. Liu, I. Guillén, M. La Manna, J. H. Nam, S. A. Reza, T. H. Le, A. Jarabo, D. Gutierrez, and A. Velten, Nature **572**, 620 (2019).
8. C.-Y. Tsai, A. C. Sankaranarayanan, and I. Gkioulekas, in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019), pp. 1545–1555.
9. S. I. Young, D. B. Lindell, B. Girod, D. Taubman, and G. Wetzstein, in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020), pp. 1404–1413.
10. F. Heide, M. O'Toole, K. Zang, D. Lindell, S. Diamond, and G. Wetzstein, ACM Trans. Graph. **38**, 1 (2019).
11. V. Arellano, D. Gutierrez, and A. Jarabo, Opt. Express **25**, 11574 (2017).
12. X. Liu, S. Bauer, and A. Velten, Nat. Commun. **11**, 1645 (2020).
13. W. Chen, F. Wei, K. N. Kutulakos, S. Rusinkiewicz, and F. Heide, ACM Trans. Graph. **39**, 1 (2020).
14. X. Liu, J. Wang, Z. Li, Z. Shi, X. Fu, and L. Qiu, Light: Sci. Appl. **10**, 1 (2021).
15. D. V. Kirmani, D. Shin, A. Colaço, F. N. C. Wong, J. H. Shapiro, and V. K. Goyal, Science **343**, 58 (2014).
16. F. Heide, S. Diamond, D. B. Lindell, and G. Wetzstein, Sci. Rep. **8**, 17726 (2018).
17. T. Goldstein and S. Osher, SIAM J. Imaging Sci. **2**, 323 (2009).
18. M. Beck and Teboulle, IEEE Trans. on Image Process. **18**, 2419 (2009).
19. M. Galindo, J. Marco, M. O'Toole, G. Wetzstein, D. Gutierrez, and A. Jarabo, in *Proceedings of SIGGRAPH '19 Posters* (2019), paper 73.
20. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, IEEE Trans. on Image Process. **13**, 600 (2004).