Fusing Reconstruction Algorithm for None-Line-of-Sight Phasor-field Imaging

CHAOYING GU[1] AND CHUANCHUAN YANG[1,*]

[1] *Electronics Department, Peking University, Beijing, 100871, China*
[*] *Corresponding author: yangchuanchuan@pku.edu.cn*

Compiled December 21, 2021

Non-line-of-sight (NLOS) imaging allows seeing through the corner by analyzing the scattered light off a relay surface. While Single Photon Avalanche Diode (SPAD) arrays can enable the imaging system to operate in a parallel way and significantly save data collection time, most of the existing demonstrations still use single-pixel SPADs due to arrays' lack of global gates and little fill-in factors, causing low signal-to-noise ratio (SNR). 32 × 32 SPAD array has been used to track hidden objects, but there is still a gap in 3D reconstruction using large-scale SPAD arrays. In this paper, we propose an algorithm to fuse the reconstruction results obtained by virtual illuminations of multiple wavelengths, which can improve the phasor-field algorithm's robustness to low SNR measurement. We demonstrate that the proposed algorithm can solve the trade-off between reconstruction SNR and resolution on synthetic datasets.

http://dx.doi.org/10.1364/ao.XX.XXXXXX

1. INTRODUCTION

NLOS imaging has been gaining interest as this technology enables the remote access of hidden or unreachable scenes around corners. Therefore, NLOS imaging can be adopted in rescuing scenarios, clinical endoscopic analysis and autonomous vehicles applications.

Single-pixel gated SPADs with a scanning laser spot are currently preferred ways to acquire hidden 3D scenes, with most set-ups using either in single pixel or array format [1]. According to existing algorithms, such as back projection [2], light-cone transform (LCT) [3], f-k migration [4] and virtual wave optics [5], the employment of single-pixel SPAD requires the laser to scan a grid on the relay wall to collect the data. The subsequent long acquisition time makes NLOS imaging currently unsuitable for applications outside laboratories [6].

To imaging moving objects, the acquisition time need to be cut down while maintaining the reconstruction SNR. This improvement can be realized by: (a) increasing the laser power and repetition rate, (b) adopting SPAD arrays to detect photons at multiple points at once, and (c) improving the algorithm's robustness to low SNR data.

Since (a) can lead to eye safety problems, some studies have been following the idea of (b). Ref.[7] provided a phasor-field algorithm which enables the non-confocal NLOS reconstruction. This algorithm allows collecting photons simultaneously using SPAD, but in experiment they still used a single-pixel SPAD. The main difficulty of using SPAD arrays in NLOS imaging is the small fill-in factor of SPAD arrays (generally 10% for second generation SPAD arrays, 13.4% for the mentioned latest megapixel SPAD array). Therefore, the active area per pixel is very limited, and each pixel collects light from a very limited area unless the relay surface is far away. When the relay surface is at the same distance with single-pixel SPAD circumstance, the signal recieved by each array pixel is significantly weaker.

Recently, Renna et al.[8] first demonstrated the possibility of incorporating a SPAD array for NLOS reconstruction using a non-confocal data acquisition scheme combined with a phasor-field method. They used a specially-designed 16 × 1 SPAD line array, but their experiment did not get rid of galvo mirrors and still scanned the laser over a grid of 150 × 150 positions. Though large-format time-gated SPAD array with up to 1 Mpixel has been available [9], to our knowledge, the largest SPAD array which has been demonstrated in experiments is 32 × 32 [10] due to large arrays' low temporal resolution or detection rate.

To loosen the restrictions on hardware, we require the reconstruction algorithm to be more robust under low SNR circumstances to compensate the SPAD arrays' small fill-in factors, which is what (c) states. Some work has been done to improve the back projection algorithm [11]. However, back projection algorithm runs more than a hundred times slower than the phasor-field algorithm using Rayleigh Sommerfeld diffraction (RSD). The latter one can reconstruct room-sized scene in seconds, thus more suitable for future real-time imaging.

In this work, we focus on improving the phasor-field algorithm to prepare for the practical application of SPAD arrays as (b) and (c) indicate. To be specific, we look into the theory and properties behind selecting the virtual illumination wavelength for reconstruction. We use the proposed criterion to reconstruct the 3D hidden scene with different virtual wavelengths. By fusing these results, we use synthetic data to demonstate that it is able to reconstruct the scene when the SNR is extremely low (the peak value of photons is below 10) still in seconds.

2. PROPOSED FUSING RECONSTRUCTION METHOD

A. Geometry Setup

We first illustrate the NLOS measurement setup in Fig.1. A single temporal impulse response measurement with illumination point $\mathbf{x_p}$ and detection point $\mathbf{x_c}$ is denoted as $d(\mathbf{x_p}, \mathbf{x_c})$, where $\mathbf{x_p}$ is the laser illumination point on the relay wall, and $\mathbf{x_c}$ is the sensor detection point on the relay wall. Reconstruction of the hidden scene need a set of different measurements. The NLOS measurement can be divided to con-focal and non-confocal by whether $\mathbf{x_p}$ and $\mathbf{x_c}$ are located at the same position.

As Fig.1 (a) plots, confocal measurement is usually implemented with single-pixel SPADs and beam splitters. This kind of setup can result in high SNR ratio especially when applying retroreflective paint. However, this setup requires a mechanical scanning system to get different $d(\mathbf{x_p}, \mathbf{x_c})$. Fig.1 (b) shows the non-confocal measurement where the laser and SPAD sensor look at different points on the relay wall. With a proper lens projecting the SPAD array's pixels to different $\mathbf{x_c}$ on the relay wall, a non-confocal setup allows acquiring different $d(\mathbf{x_p}, \mathbf{x_c})$ at once when fixing. Hence, the application of SPAD arrays requires the non-confocal measurement.

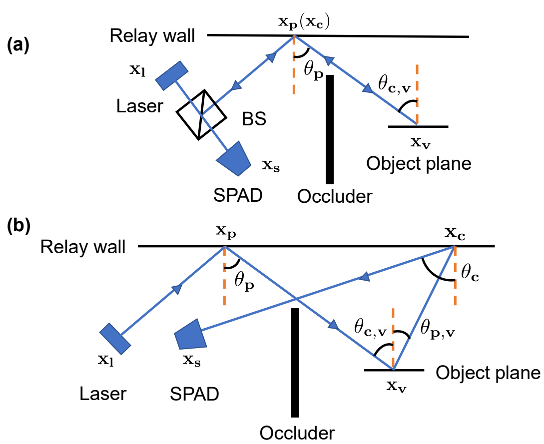

Fig. 1. NLOS measurement setup. (a) Confocal measurement. (b) Non-confocal measurement.

Furthermore, when considering the spatial broadening due to field of view (FoV), $\mathbf{x_p}$ and $\mathbf{x_c}$ should be 2D Gaussian spots rather than ideal 0-dimensional points. When the distance between $\mathbf{x_p}$ and $\mathbf{x_c}$ is smaller than the spatial divergence, the non-confocal measurement $d(\mathbf{x_p}, \mathbf{x_c})$ should act as the confocal setup. For confocal measurement, we can calculate the intensity ratio between the first bounce and third bounce for a specific spatial point $\mathbf{x_v}$ of the hidden scene as:

$$\frac{I_{\text{Third Bounce}}}{I_{\text{First Bounce}}} = \left[\frac{\sigma_x \sigma_y}{2\pi \left| \mathbf{x_v} - \mathbf{x_c} \right|^2} \right]^2 \cos\theta_{\mathbf{c,v}} \cos\theta_{\mathbf{p}} \tag{1}$$

where c is the speed of light and σ_x and σ_y are the standard deviations of the 2D Gaussian function. For room-sized geometry and typical sensor parameters, this ratio is around 10^{-7} to 10^{-9}. Since currently available SPAD sensor pixels all go though an avalanch process after a detection event happens, they have a dead period (the typical value is 30 ns to 60 ns) before then can detect the next arriving photon. Thus, the first bounce of high intensity can cause strong pile-up effect and overshadow the desired third bounce. For most of the commertial SPAD arrays, the pixels do not have a global shutter to gate out the first bounce simultaneously. Thus, the laser point $\mathbf{x_p}$ should be outside the total FoV of the SPAD array to avoid pile-up.

B. Virtual Wavelength Selection for Fusing Reconstruction

The phasor field algorithm uses a virtual illumination function which is similar to the exposure function in correlation-based Time-of-Flight (C-ToF) imaging[12]. A C-ToF imaging system illuminates the scene with temporally modulated light source. The radiant intensity is modulated by a sinusoid of a specific frequency. In NLOS circumstances, the illumination is convolved with a Gaussian pulse to perform temporal focusing[13]. We have the implemented virtual illumination function:

$$\mathcal{P}\left(\mathbf{x}_\text{p}, t\right) = e^{j\omega_0 t} \delta\left(\mathbf{x}_\text{p} - \mathbf{x}_\text{ls}\right) e^{-\frac{(t-t_0)^2}{2\sigma^2}} \tag{2}$$

where λ_0 is the given reconstruction wavelength and $\omega_0 = 2\pi c/\lambda_0$, $\mathbf{x}_\text{ls}$ is the virtual light source position, and t_0 is the focused time. In general, $\mathbf{x}_\text{ls}$ is one of the laser illumination points on the relay wall, and t_0 corresponds to the imaging depth of the hidden scene.

When the laser shoots at $\mathbf{x}_\text{p}$ and the SPAD pixel points at $\mathbf{x}_\text{c}$, the measured temporal data of arriving photons are denoted as $H\left(\mathbf{x}_\text{p} \to \mathbf{x}_\text{c}, t\right)$. To perform the reconstruction, the RSD algorithm picks a certain illumination function $\mathcal{P}\left(\mathbf{x}_\text{p}, t\right)$ first, and convolve it by the measured data to get the virtual captured wavefront on the relay wall as:

$$\mathcal{P}\left(\mathbf{x}_\text{c}, t\right) = \int_\text{P} \left[\mathcal{P}\left(\mathbf{x}_\text{p}, t\right) * H\left(\mathbf{x}_\text{p} \to \mathbf{x}_\text{c}, t\right)\right] \mathrm{d}\mathbf{x}_\text{p} \tag{3}$$

where the $*$ operator indicates convolution in time.

According to Rayleigh criterion, the spatial resolution is $\Delta = 0.61\lambda L/d$ where L is the imaging distance and d is the virtual aperture diameter. This criterion shows shorter wavelength should be able to bring better reconstruction results. However, existed literature [5] chooses proper reconstruction frequency manually for each measured dataset, the only restriction is $\lambda \geq 2\Delta x$ where Δx is the distance between detection points $\mathbf{x_c}$ on the relay wall. In practice, existed work applies $\lambda = 4\Delta x$ or $\lambda = 6\Delta x$ to achieve a better SNR. The guideline mentioned above only applies to regular sampling grid, and ignores the influence of the temporal resolution of the measurement system, thus not applicable to different experiment setups.

The temporal measured impulse response $H\left(\mathbf{x}_\text{p} \to \mathbf{x}_\text{c}, t\right)$ can be written as:

$$\begin{aligned} H\left(\mathbf{x}_\text{p} \to \mathbf{x}_\text{c}, t\right) &= \left[f(\mathbf{x}_\text{p} \to \mathbf{x}_\text{c}, t) * g(t)\right] * \ell(t) \\ &= f(\mathbf{x}_\text{p} \to \mathbf{x}_\text{c}, t) * \left[g(t) * \ell(t)\right] \end{aligned} \tag{4}$$

where $f(\mathbf{x}_\text{p} \to \mathbf{x}_\text{c}, t)$ is the photon probability density function for measurement $d(\mathbf{x}_\text{p} \to \mathbf{x}_\text{c})$, $g(t)$ is the gating window profile, and $\ell(t)$ is the laser impulse profile.We model the $g(t)$ and $\ell(t)$ as gaussian functions, thus the full width at half maximum (FWHM) of the measurement system can be estimated by:

$$\text{FWHM} = \sqrt{t_p^2 + t_d^2} \tag{5}$$

where t_p is the pulse width of laser, and t_d is the timebin resolution of SPAD.

Therefore, it is crutial to adjust the reconstruction wavelength for differnet experiment setups to obtain a suitable SNR. To decide the proper wavelength, we first pull Eq.2 into Eq.3 and this generates:

$$\mathcal{P}_{\mathcal{F}}(\mathbf{x}_c,\omega) = (2\pi)^{\frac{3}{2}}\sigma \cdot H_{\mathcal{F}}(\mathbf{x}_{ls} \to \mathbf{x}_c,\omega) \cdot e^{-\frac{\sigma^2}{2}(\omega-\omega_0)^2} e^{-i\omega t_0}. \quad \textbf{(6)}$$

Then we pull in Eq.4, and denotes $h(t) * \ell(t)$ as $e^{-t^2/2\sigma_t^2}$ where $\sigma_t = \mathrm{FWHM}/2.355$. The actual virtual captured wavefront's frequency intensity will satisfy:

$$|\mathcal{P}_{\mathcal{F}}(\mathbf{x}_c,\omega)| \propto e^{-\frac{\sigma_t^2}{2}\omega^2} e^{-\frac{\sigma^2}{2}(\omega-\omega_0)^2} \quad \textbf{(7)}$$

where the former term works as an attenuator and the latter term works as a selector. $\lambda_0 = 2\pi c/\omega_0$ should be set larger than a threshold so that the central frequency components of the latter term can be large enough to perform the reconstruction.

In RSD method, each frequency component is propogated back into the scene seperately to generate the reconstruction, thus the decay of these high frequency components will cause the reconstruction to be blurrier and dimmer. we conclude that the virtual reconstruction wavelength should satisfy that

$$\lambda_0 > \alpha \cdot (c \cdot \mathrm{FWHM}) \quad \textbf{(8)}$$

to obtain suitable SNR. The parameter α has relation to do with the exposure time and laser pulse energy, which decides the intrinsic SNR. The typical value of α is $2 \sim 3$ when the peak value of third bounce among all array pixels is around 20.

C. Fusion Alogorithm

When using the SPAD array in NLOS imaging, the data collection time can be saved by parallel measurement, but the detection rate and FoV of SPAD array pixels are limited. Improvements in the fabricating can improve the detection rate, but for now SPAD arrays suffer from low detection rate which leads to ill-posed SNR. When using single-pixel SPAD for detection, we usually choose the smallest available wavelength to get the best resolution. When the instrinsic SNR goes down in array circumstances, Eq.7 indicates that longer wavelength (smaller ω_0) should be employed to suppress the noise. However, this suppression would cause a worse resoslution according to Rayleigh criterion.

To solve this contradiction, we propose a simple but effecitive algorithm to fusion the reconstruction results of short and long virtual wavelength. We show that this fusion algorithm is able to improve both the SNR and the resolution. Let $I(\mathbf{x_v},\lambda)$ denotes the reconstruction result using virtual wavelength $\lambda_0 = \lambda$, the fusioned result can be expressed a weighted summation for different wavelengths' results:

$$I(\mathbf{x_v}) = \sum_{n=1}^{N} w_n(\mathbf{x_v}) I(\mathbf{x_v},\lambda_n) \quad \textbf{(9)}$$

where the weighting factor $w_n(\mathbf{x_v})$ denotes the credibility of voxel $\mathbf{x_v}$ for wavelength λ_n.

Further, since we have concluded in Eq.8 that the SNR is roughly propotional to the virtual wavelength we use, we use the deviation from longer wavelength's reconstruction result to measure the credibility:

$$w_n = \mathcal{A}\left(|I(\mathbf{x_v},\lambda_{n+1}) - I(\mathbf{x_v},\lambda_n)|\right) \quad \textbf{(10)}$$

where $\mathcal{A}(\cdot)$ is a monotonic decreasing function in range $[0,1]$. As for $n = N$, we let $w_N = 0$, and only use the longest wavelength's result as a reference to get better SNR without sacrificing resolution.

We find that when the deviation is larger, this voxel's value is more likely to be noise. The function $\mathcal{A}(\cdot)$ works as a counterpart of activation function in neural networks. If the quality of a voxel supposed to be better than the others, it is more activated in the final results. Through numerical experiments, we find that the sigmoid function can achieve the most smooth and stable result:

$$\mathcal{A}(x) = 1 - \frac{1}{1+e^{-x}}. \quad \textbf{(11)}$$

The computation overhead of the fusion algorithm is negligible compared to the overhead of RSD algothims. Therefore, the time and memory overhead mainly depends on the number N of total wavelengths used in reconstruction. In practice, we find that $N = 2$ can improve the quality of reconstructed volume significantly.

3. NUMERICAL RESULTS

A. Simulation Setup

We setup the NLOS simulation using the idea of ray tracer, and build a light-weight renderer. We first discretize the surface of the hidden object as squares at 3D position $\mathbf{x_v}$ with area S_o and surface normal α. The photon probability density function $f(t)$ in Eq.4 is calculated as:

$$\begin{aligned} &f(\mathbf{x}_p \to \mathbf{x}_c, t) \\ &= \frac{I_0}{(2\pi)^3} \int \frac{S_o \cos\theta_{\mathbf{c,v}}}{|\mathbf{x_v}-\mathbf{x_p}|^2} \cdot \frac{\sigma_x\sigma_y \cos\theta_{\mathbf{p,v}}}{|\mathbf{x_v}-\mathbf{x_c}|^2} \cdot \frac{\sigma_x\sigma_y \cos\theta_{\mathbf{c}}}{|\mathbf{x_c}-\mathbf{x_s}|^2} dS_o \end{aligned} \quad \textbf{(12)}$$

where the geometry notations are plotted in Fig.1 (a), and the constant parameter I_0 is composed of the laser's wavelength and pulse energy and the SPAD pixels' detection rate. I_0 can be easily calibrated through one measurement.

Additionally, in experiment the detection efficiency can be extremely low (less than one third bounce photon per 10k frames). We also take the poisson distribution of arriving photons into consideration, the final expression of simulated data is given by a poisson random number with mean $H(\mathbf{x}_p \to \mathbf{x}_c, t)$ in Eq.4.

We use 100×100 SPAD array with 31.25 ps timebin width , and the pixels are pojected to 1 m × 1 m grids on the relay wall. The laser has 50 ps width and 500 kHz repetition rate. The hidden object size is 0.4 m × 0.4 m. The 3D setup of our simulation and the detected data is shown in Fig.2.

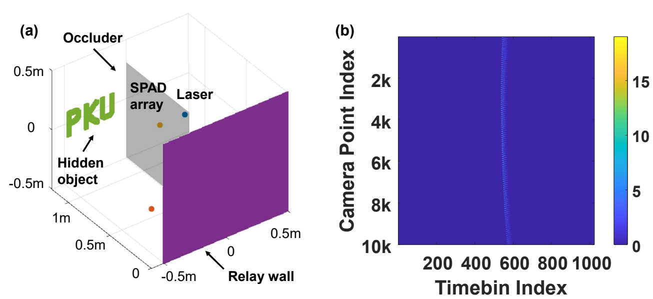

Fig. 2. Simulation details. (a) 3D setup of simulation. (b) simulated histogram data.

B. Verification

In Fig.3, we fix the scene geometry parameters to verify the influence of the system's FWHM. We also fix the pulse energy to make sure in each measurement the sum of total detected photons is the same. However, the imaging results with the same wavelength $\lambda_0 = 4\,\text{cm}$ show significant difference in SNR, which proves the conclusion we get in Eq.8.

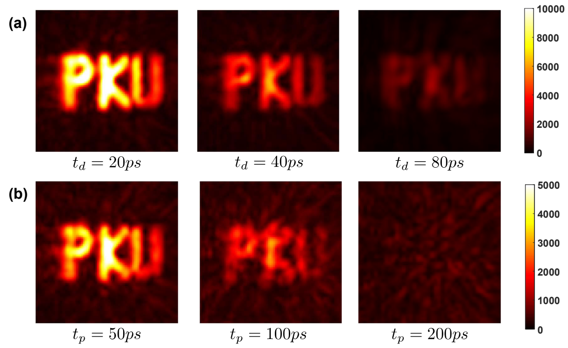

Fig. 3. Simulation with different experiment setups and the same virtual wavelength $\lambda_0 = 4cm$. (a) Fix laser pulse width $t_p = 50ps$, change SPAD's timebin resolution;(b): fix SPAD's timebin resolution $t_d = 31.25ps$, change laser pulse width.

In Fig.4, we set $t_p = 50\,\text{ps}$ and $t_d = 31.25\,\text{ps}$, then we can calculate that the proper λ_0 is 4 cm according to Eq.8. Fig.4 (b) and (c) proves that short virtual wavelength brings a good resolution, but the shot noise is affecting the reconstruction quality. By contrast, the background noise get suppresed but the letters are blurred in long virtual wavelength reconstruction.

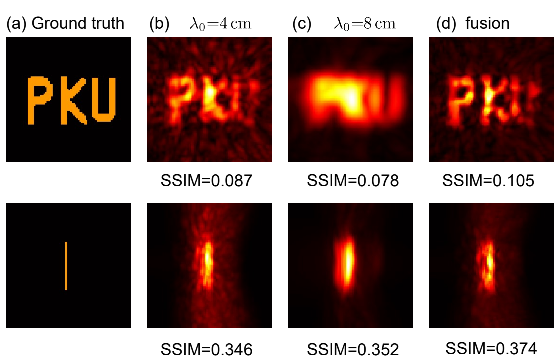

Fig. 4. Imaging using different methods. Top row: front view and second row: side view. (a) Ground truth. (b)-(c): RSD method using virtual illumination wavelength λ=4cm and λ=8cm. (d) Proposed fusing reconstruction method.

In Fig.4 (d), we fusion the results of $\lambda_1 = 4\,\text{cm}$ and $\lambda_2 = 8\,\text{cm}$ using the aforementioned weighting factor. Implemented on general CPU (@ 3.30 GHz), the computation time of the proposed fusing reconstruction method is 0.015 s for $150 \times 150 \times 100$ voxels, while the RSD reconstruction takes 8.8 s and 4.5 s each for different wavelengths.

We use structural similarity (SSIM) to measure the quality of reconstruction. Fig.5 shows that our proposal can improve the SSIM obviously. We further show in Fig.5 that the fusion result of only two wavelengths we chose can outperform the ones of all of the single virtual wavelength results.

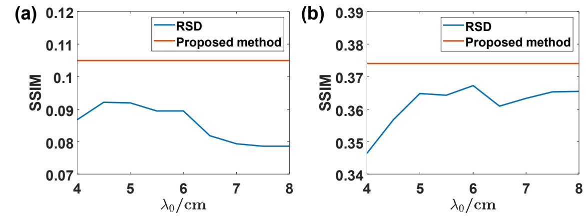

Fig. 5. The SSIM of RSD method using different single wavelength and the proposed fusing reconstruction method using two chosen wavelengths. (a): Front view; (b): side view.

4. CONCLUSION

In conclusion, we develop a fusing reconstruction method for NLOS imaging using SPAD array. This method is based on the proposed criterion for the selection of virtual wavelength, and is proved to improve the the imaging quality when the SNR of measured data is incredibly low. From the trade-off property between SNR and resolution, the proposal improves the phasor-field algorithm without increasing the computational overhead. This algorithm is simple and practical enough to facilitate the use of SPAD arrays in NLOS imaging.

Funding. National Key Research and Development Program of China (2019YFB1802904).

Acknowledgments. We thank the Computational Optics Group at University of Wisconsin-Madison for providing their RSD algorithm code, and Andreas Velten for his kind guidance.

Disclosures. The authors declare no conflicts of interest.

Data Availability Statement. Data underlying the results presented in this paper are available in Ref.[14].

REFERENCES

1. D. Faccio, A. Velten, and G. Wetzstein, Nature Reviews Physics **2**, 318 (2020).
2. A. Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, M. G. Bawendi, and R. Raskar, Nat. communications **3**, 1 (2012).
3. M. O'Toole, D. B. Lindell, and G. Wetzstein, Nature **555**, 338 (2018).
4. D. B. Lindell, G. Wetzstein, and M. O'Toole, ACM Transactions on Graphics (TOG) **38**, 1 (2019).
5. X. Liu, I. Guillén, M. L. Manna, J. H. Nam, S. A. Reza, T. H. Le, A. Jarabo, D. Gutierrez, and A. Velten, Nature **572**, 620 (2019).
6. M. Buttafava, J. Zeman, A. Tosi, K. Eliceiri, and A. Velten, Optics express **23**, 20997 (2015).
7. X. Liu, I. Guillén, M. L. Manna, J. H. Nam, S. A. Reza, T. H. Le, A. Jarabo, D. Gutierrez, and A. Velten, Nature **572**, 620 (2019).
8. M. Renna, J. H. Nam, M. Buttafava, F. Villa, A. Velten, and A. Tosi, Instruments **4**, 14 (2020).
9. K. Morimoto, A. Ardelean, M.-L. Wu, A. C. Ulku, I. M. Antolovic, C. Bruschini, and E. Charbon, Optica **7**, 346 (2020).
10. G. Gariepy, F. Tonolini, R. Henderson, J. Leach, and D. Faccio, Nat. Photonics **10**, 23 (2016).
11. X. Feng and L. Gao, Opt. letters **45**, 3921 (2020).
12. M. Gupta, S. K. Nayar, M. B. Hullin, and J. Martin, ACM Transactions on Graph. (ToG) **34**, 1 (2015).
13. J. Marco, A. Jarabo, J. H. Nam, X. Liu, M. Ángel Cosculluela, A. Velten, and D. Gutierrez, "Virtual light transport matrices for non-line-of-sight imaging," (2021).
14. Chaoying Gu, "Synthetic Dataset," https://github.com/ArianaGu/NLOS-fusing.