Dynamic non-line-of-sight imaging system based on the optimization of point spread functions

CHENGQUAN PEI,[1,7] ANKE ZHANG,[2,7] YUE DENG,[3] FEIHU XU,[4]
JIAMIN WU,[1] DAVID U-LEI LI,[5] HUI QIAO,[1] LU FANG,[6,8] AND
QIONGHAI DAI[1,2,9]

[1]*Department of Automation, Tsinghua University, Beijing 100084, China*

[2]*Tsinghua-Berkeley Shenzhen Institute, Tsinghua University, Shenzhen 518000, China*

[3]*School of Astronautics, Beihang University, Beijing, China*

[4]*Department of Modern Physics, University of Science and Technology of China, Hefei 230026, China*

[5]*Faculty of Engineering, University of Strathclyde, Glasgow G4 0NW, Scotland, UK*

[6]*Department of Electronic Engineering, Tsinghua University, Beijing 100084, China*

[7]*These authors contributed equally to this work.*

[8]*fanglu@sz.tsinghua.edu.cn*

[9]*qhdai@tsinghua.edu.cn*

Abstract: Non-line-of-sight (NLOS) imaging reveals hidden objects reflected from diffusing surfaces or behind scattering media. NLOS reconstruction is usually achieved by computational deconvolution of time-resolved transient data from a scanning single-photon avalanche diode (SPAD) detection system. However, using such a system requires a lengthy acquisition, impossible for capturing dynamic NLOS scenes. We propose to use a novel SPAD array and an optimization-based computational method to achieve NLOS reconstruction of 20 frames per second (fps). The imaging system's high efficiency drastically reduces the acquisition time for each frame. The forward projection optimization method robustly reconstructs NLOS scenes from low SNR data collected by the SPAD array. Experiments were conducted over a wide range of dynamic scenes in comparison with confocal and phase-field methods. Under the same exposure time, the proposed algorithm shows superior performances among state-of-the-art methods. To better analyze and validate our system, we also used simulated scenes to validate the advantages through quantitative benchmarks such as PSNR, SSIM and total variation analysis. Our system is anticipated to have the potential to achieve video-rate NLOS imaging.

1. Introduction

Imaging hidden objects from a camera's view is an emerging research topic with a broad range of applications, including robotics vision, remote sensing, medical diagnosis and autonomous vehicles. In general, a non-line-of-sight (NLOS) imaging system (including a short-pulse laser and a transient imaging device) sequentially performs the following steps: 1) the pulsed-laser shines on the relay wall, 2) the scattered light is bounced on the hidden object and diffusely reflected to the relay wall, and 3) a transient imaging device records the scattering response from the relay wall. NLOS imaging can reveal hidden objects around corners or obstructions due to the excellent temporal and spatial resolvability of time-resolved sensors, such as streak cameras, gated sensors, or SPADs.

　Over the years, many image acquisition methods have been proposed to achieve NLOS imaging [1], including time-correlated single-photon counting (TCSPC) [2–9], acoustic [10], passive imaging [11–13], interferometry [14,15], thermal imaging [16,17] methods, and wave

methods [18–20]. Moreover, several image reconstruction methods, such as occlusion aware light-inverse methods [21–26], deep-learning methods [27], heuristic solutions [28–31], linear inverse methods [32–34], data driving algorithms [35–37], back-projection methods [8,9,30,38–41], and deconvolution methods [2,5,11], were also adopted to solve NLOS imaging problems [42].

Among these attempts, TCSPC approaches are the most widely used in NLOS imaging due to their superior temporal resolutions. The first time-resolved NLOS system was implemented using a streak camera with a scanning laser [2]. A scanned 5D data cube was used for reconstructing three-dimensional (3D) images with back-projection methods [2]. A laser imaging, detection, and ranging (LIDAR) system with a confocal imaging setup was proposed to achieve fine-grained 3D image reconstruction using uniform spatial point-spread functions (PSF) [2]. NLOS imaging can also be formulated as a diffractive wave propagation problem called the phase-field method. Mathematical operators can model wave propagations in a conventional line-of-sight imaging system to implement NLOS imaging [18–20]. Among all introduced methods, the back-projection plays an essential role in reconstructing objects from transient images. However, back-projection methods cannot discriminate signals from acquisition noise [8,30,38–41]. Conventionally, camera sensor is the main noise source. However, in the NLOS system, the light dispersion and noise is also generated from light propagation [1]. To minimize noise from affecting the reconstruction results, oversampling is usually required. The repetitive acquisition drastically reduces the reconstruction frame rate, not suitable for dynamic NLOS imaging. The inverse filtering methods, such as confocal systems, require laser scanning sequentially over the scattering wall, and each illumination point is single-pixel imaged [2,3]. The scanning process and single-pixel imaging methods prevent the system from single-shot reconstruction. The passive and interferometry systems require an occlusion aperture with a known shape beforehand in the reconstruction zone [11–15]. However, such prior information is hard to acquire when solving realistic problems.

Although the systems mentioned above can image static scenes, they are slow and complicated due to lengthy scanning processes. The acquisition system requires a particular illumination pattern from a laser scanning process, limiting the achievable frame rate for dynamic imaging. Recently, a novel method develops to implement dynamic NLOS imaging system that utilizes sparse illumination pattern to reduce physical scanning time [6]. The dynamic NLOS imaging was implemented by two 16 by 1 pixel SPAD array. However, the scanning operation is also necessary. In Ref. [43], a novel method with a SPAD array was adopted to track the position of several moving objects. However, it is unable to image 3D scenes. In Ref. [44], the streak camera with a micro-lens array was used to acquire transient images, but the system also requires laser scanning to reconstruct 3D scenes. In the single-shot dynamic mode, the system can only reconstruct simple scenes such as tracking the position of two moving pads. We propose to utilize a 32×32 SPAD array [45,46] with a fixed laser to acquire all time-resolved information in a single shot. The proposed system is reciprocal to a fixed single-pixel detector with a laser scanning 32×32 position. However, the confocal setup (with simultaneous scanning laser and detector system) does not apply to the global acquisition process [3], where PSFs are non-uniform. Also, compared with a single-pixel system, SPAD array detectors have more substantial noise due to the crosstalk between pixels. Thus, we proposed an optimization-based approach by utilizing global PSF patterns. By discriminating PSF features from acquisition noise, the optimization-based method can achieve detailed results in low signal-to-noise (SNR) data from the SPAD array. With the enhanced collection efficiency from an imaging lens, the acquisition time and system noise significantly decrease compared to small aperture imaging. We can achieve NLOS reconstruction with a 32×32 SPAD array with a 55ps time resolution [45]. Compared with the 64×64 confocal NLOS approach with a 4ps temporal resolution [2], our system only consumes a 1.8% space time-bandwidth product.

2. Methods

2.1. Optimization of spatial point spread functions

As shown in Fig. 1(a), the NLOS imaging system consists of a 32×32 SPAD array, a 50MHz, 70fs femtosecond laser, and a relay wall. The laser light first shines the wall, travels to the hidden scene, and propagates back to the relay wall. A SPAD array camera captures transient waves containing the scattered geometric information of hidden objects. The laser's illumination position and the SPAD array's field of view (FoV) are fixed throughout the acquisition process. The transient image can be derived from a linear combination of each hidden point's response Ψ in Fig. 1(a). The linear function is showed in Fig. 1(b), where τ is the NLOS system's response captured by the SPAD camera, ρ is the shape of the NLOS object, Ψ is the NLOS PSF matrix, and each column of Ψ represents the PSF for a fixed point on the target. Figure 1(c) shows how each object voxel contributes to a transient image. The entire transient image is generated from the linear sum of all the images. If there is no occlusion, we can turn the forward model into a linear system and generate a corresponding matrix representation. Figure 1(d) is a simple example to depict circular objects at different positions producing transient images with different curvatures and positions. A response captured by a SPAD array with 32×32 pixels for a hidden sample object in the bottom is also provided.
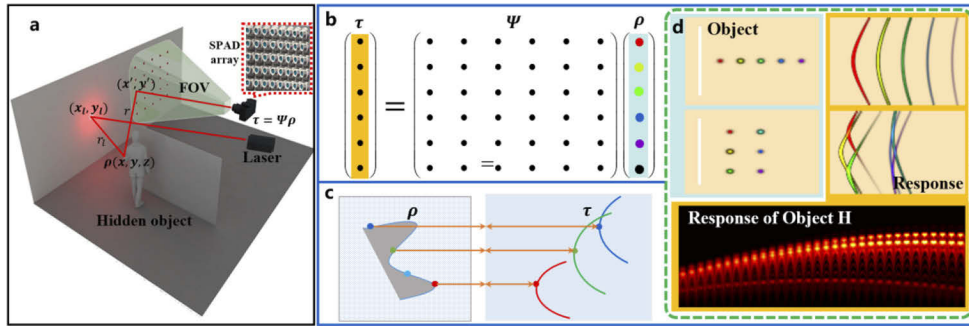

Fig. 1. The system and the algorithm. **a.** The femtosecond laser and the SPAD array record time-resolved data in a fixed position. **b.** time-resolved data can be illustrated as a linear combination of discrete points in a spatially variant PSF set. **c.** each object voxel contributes to a transient image; **d.** a simple example illustrates that when the object is a simple dot at various positions, the transient image's response varies according to its position.

The model traces the time-resolved response of a femtosecond laser interacting with scattering media and the hidden objects, and the transient information recorded by the sensor can be represented by [2,3],

$$
\begin{aligned}
\tau(x', y', t) = \iiint_{\Omega} \frac{1}{r_l^2 r^2} \rho(x, y, z) \times \\
\delta\left(\sqrt{(x'-x)^2 + (y'-y)^2 + z^2} + \sqrt{(x_l-x)^2 + (y_l-y)^2 + z^2} - tc\right) dxdydz
\end{aligned}, \tag{1}
$$

where ρ is the sought-after albedo of the hidden scene in the 3D half-space Ω when the distance from the scattering board. The Dirac Delta function δ describe the instant where the transient scattered light hit the relay wall. r_l is the distance from the light source (x_l, y_l, 0) to the imaging point (x, y, z), and r is the distance from the imaging point (x', y', 0) to the imaging point (x, y, z). The transient image is recorded when the light source illuminates the position (x_l, y_l) on the wall. Every measurement sample τ (x', y', t) captures the photon flux at the position (x', y') and the time t relative to an incident pulse scattered by the laser position at $t = 0$. With a Gaussian-shaped

laser pulse alongside the Poisson noise of the time-resolved camera, the system response can be modeled as:

$$
\begin{aligned}
\tau(x', y', t) = {} & \iiint_\Omega \frac{1}{r_l^2 r^2} \rho(x, y, z) \times \\
& P(\delta\left(\sqrt{(x'-x)^2 + (y'-y)^2 + z^2} + \sqrt{(x_l-x)^2 + (y_l-y)^2 + z^2} - tc\right)) dx dy dz
\end{aligned}
\tag{2}
$$

where P is the Poisson noise from the camera acquisition process. In the SPAD camera, Photon counters detect the single photon event within a certain time bin with some probability. In the extreme low light condition of the NLOS experiment, the sampling process of SPAD becomes a Bernoulli experiment that it repeats many times [2]. Since we assume the laser position is fixed in this system, we can formulate Eq. (2) to be the correlation between the intensity of a voxel in the hidden scene ρ and the encoding function ψ:

$$
\tau(x', y', t) = \iiint_\Omega \rho(x, y, z) \times \psi(x, y, z, x', y', t) dx dy dz,
\tag{3}
$$

where

$$
\psi(x, y, z, x', y', t) = P(\delta\left(\sqrt{(x'-x)^2 + (y'-y)^2 + z^2} + \sqrt{(x_l-x)^2 + (y_l-y)^2 + z^2} - tc\right)).
\tag{4}
$$

Equation (3) is similar to Eq. (2), and the function ψ is a known encoding function that corresponds to every target point ρ. However, because the function ψ varies spatially, we cannot directly convert the problem into the deconvolution form in a confocal setup. Instead, we aim to find the trial solution that can generate the closest system response

$$
\arg\min \left\{ \left\| \tau(x', y', t) - \iiint_\Omega \rho^*(x, y, z) \times \psi(x, y, z, x', y', t) dx dy dz \right\|_2^2 \right\},
\tag{5}
$$

where ρ^* is the trial solution. To avoid the problem being under-determined, the reconstruction resolution (ψ's column number) should be lower than the resolution of raw data (ψ's row number).

2.2. Discrete algorithm implementation

We need to discretize the forward model illustrated by Eqs. (3)-(5) to implement the optimization algorithm. The reconstruction zone $\rho(x, y, z)$ is separated into $\rho_{xyz} \in R^{M \times N \times K}$, where ψ_{xyz} is a corresponding independent PSF for the position *(x, y, z)* in ρ_{xyz}. Because each correlated response is linearly independent, we can vectorize ρ_{xyz} and ψ_{xyz} into ρ_i and ψ_i, where $i \in [1, M \times N \times K]$. Equations (3) and (4) can be simplified to

$$
\tau = \sum_{i=1}^{M \times N \times K} \psi_i \rho_i
\tag{6}
$$

$$
\psi_i(x', y', t) = \sum_{i=1}^{M \times N \times K} \frac{1}{rr_l} P\left(\delta\left(\sqrt{(x'-x_i)^2 + (y'-y_i)^2 + z^2} + \sqrt{(x_l-x_i)^2 + (y_l-y_i)^2 + z^2} - tc\right)\right),
\tag{7}
$$

where τ is the vectorized signal from the SPAD array camera, and the forward model becomes a linear system. *x', y'* represents the SPAD imaging position on the scattering wall, and t is the TCSPC time stamp. The Poisson noise P is still added in the discrete forward model. By concatenating vectorized ψ_i into a matrix A, Eq. (6) can be illustrated as,

$$
\tau = A\rho.
\tag{8}
$$

In our experiment, each PSF has a dimension of $R^{32 \times 32 \times 256}$. If the number of PSFs is lower than 32×32×256, the system is over defined. We used a 16×16×64 NLOS reconstruction domain.

Thus, the dimension of matrix A is $R^{262144 \times 16384}$. To reconstruct the NLOS scene is to find the best fit ρ, leading to a typical optimization problem,

$$\arg \min_{\theta} ||\tau - A\rho||_2^2 + ||\rho||_2^2. \tag{9}$$

We used the normal standard gradient descent with learning rates 1×10^{-4} and 1×10^{-6} for 2000 iterations for the static scene to optimize Eq. (9). In a dynamic scene, because the difference is minimal between adjacent frames, we can use previous frames to optimize the current frame to reconstruct images faster. We used 2000 iterations with a learning rate of 1×10^{-4} for the first frame and 100 iterations with a learning rate of 1×10^{-6} for each consecutive frame. We utilized one GTX 2080Ti graphics card to run the optimization method; the runtime for the first frame is around 20 seconds to optimize 2000 iterations. It took about 0.9 seconds and 100 iterations to reconstruct each consecutive frame.

3. Comparisons based on synthetic data

We compared the proposed method with confocal and phasor field reconstruction methods through simulations of two scenes, "H" and "X", as shown in Fig. 2(a) and Fig. 2 (b), respectively. Both scenes were reconstructed with a 0.7m scanning aperture and a Gaussian laser in temporal and spatial dimensions. The spatial resolution is 32×32, and the time-resolution is 32 ps in time-resolved data.

To compare all the methods based on the same external condition, the forward process of time-resolved data is adjusted to the confocal setup. Although such a setup is different from our experimental system, the only difference is each response's curvature in the ideal synthetic data. The only difference in the spatial response should not affect reconstruction results. Our method's NLOS reconstruction still depends on the spatially varied response. Both the phasor field and our method yield similar results regardless of whether the setup is in the confocal mode. However, the confocal method cannot reconstruct spatially varied responses. Directly applying the confocal algorithm on non-confocal data results in significant artifacts in the non-confocal mode. We adjusted scene parameters for all methods' reconstruction processes according to simulated experiments, ensuring that all the reconstructed data is under the same condition. We can achieve better results from simulations than confocal methods, even though our method is designed for solving general NLOS problems.

We extensively conducted reconstruction experiments under different noise levels and compared the results in the peak signal-to-noise ratio (PSNR), the structural similarity (SSIM) and the total variation (TV) with confocal and phase-field methods. As Fig. 2 shows, our method has the highest PSNR and SSIM while maintaining the lowest TV. Our method shows excellent noise-rejection capacity. As the noise level increases, PSNR and SSIM decrease as expected. The TV increases due to rising random reconstruction noise. All reconstruction methods show a similar trend, and our method offers the best results.
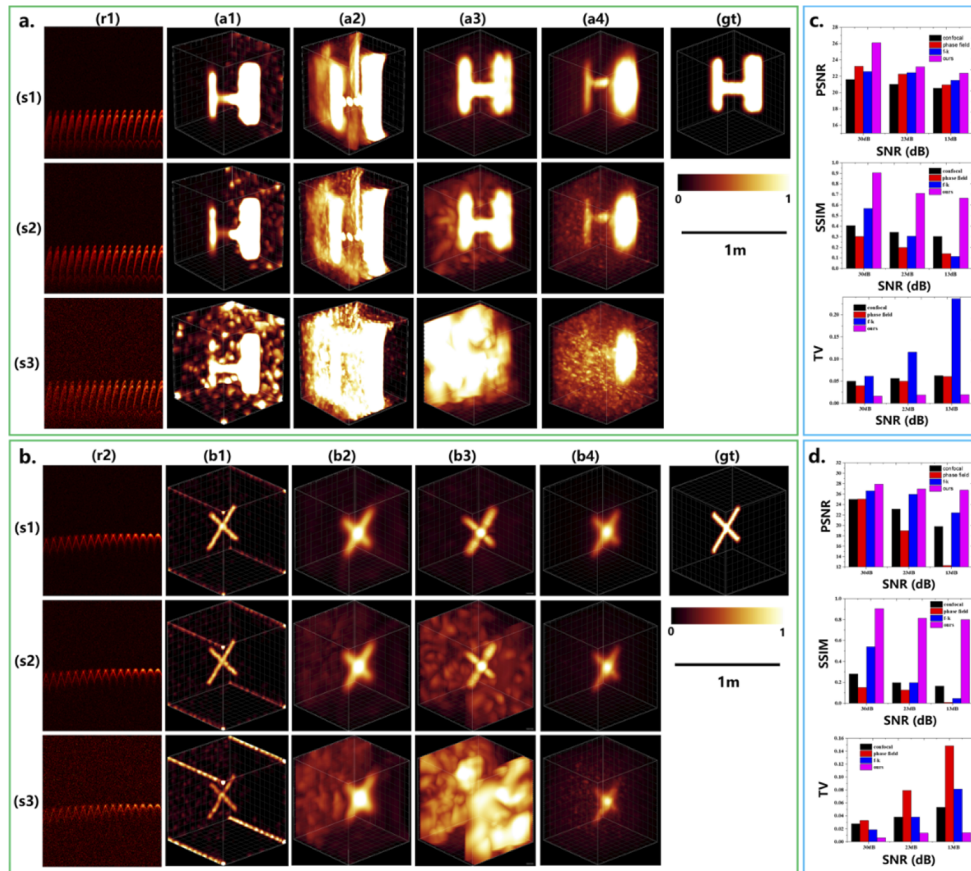
Fig. 2. The comparison of various methods in synthetic data. **a.** Reconstruction result of scene "H", column (r1) represents the raw data with various SNR setup, (s1) 30.01 dB, (s2) 22.99 dB (s3) 13.03 dB. The ground truth (gt) is compared with (a1) Confocal [2], (a2) Phasor Field [3], (a3) FK-Migration [4]. **b.** Reconstruction result of scene "X", column (r2) represents the raw data with various SNR value, (s1) 30.01 dB, (s2) 23.01 dB, (s3) 13.01 dB. The ground truth (gt) is compared with (b1) Confocal, (b2) Phasor Field, (b3) FK-Migration. **c. d.** PSNR, SSIM and TV plots at different noise levels of scene "H" and "X".

4. Experiments

4.1. Pixel synchronization

As pixels' TCSPCs in the SPAD array operate independently, we need to synchronize the time stamp of each detector. Before experimenting, the SPAD array can image the relay wall orthogonal to the sensor without any object in the reconstruction zone. The laser repetition rate is 50MHz, and all SPAD pixels are illuminated simultaneously. After more than 5 seconds of the TCSPC integration, the dark current and noise are negligible compared to a strong reflected signal from the scattering board. In the experiments, we chose the time stamp corresponding to the maximum count as the start timestamp for every TCSPC pixel.

4.2. noise2noise (N2N) validations

To benchmark the performance of the N2N denoising method [47] quantitatively, we trained the network model over a synthetic dataset. The dataset was formed by 8000 time-resolved images,

and each image was corrupted with 3 dB Gaussian noise. The model parameters were obtained by training the model 30 epoch with batch size of 4. We discussed the model details further in Appendix 2.

In Fig. 3(a), We compared reconstruction results visually in noise level of 2.78 dB and 11.8 dB. The random noise was drastically reduced when applied with N2N method, and the reconstruction results by N2N was more visually satisfactory. We tested the method over various noise level from 2.78 to 11.8 dB. As the Fig. 3(b) showed, the N2N method generally brings around 2 dB of PSNR gain and 0.1 SSIM gain.

We conducted the following tasks to evaluate the proposed method. (1) Data acquisition & hardware configuration: Our system consists of a femtosecond amplified diode laser (500 mW at 780 nm with a pulse width of 70fs at 50 MHz repetition rate) and a 32×32 SPAD-TCSPC array (PF32, Photon Force, Ltd.) with a 55ps time-resolution and a 100ns dead time. (2) NLOS setup: The system is located 2 m from the relay wall, with the scenes hidden from the direct view. The FoV on the relay wall is 82cm×82cm. The scattering wall is made of standard white Styrofoam. To avoid pile-up effects, we set the laser position out of the FoV on the relay wall.

The experiments included image of hidden objects "H", "S", and "U" in 0.1s acquisition time. The sample was placed at a random position and angle of about 0.6 meters away from the scattering media. Under a short exposure from the SPAD array in Fig. 5, the proposed method fails to reconstruct the hidden object when noise is higher than a certain level. Thus, we developed a novel 3D noise2noise method (N2N, see Appendices) to improve transient images' SNR. Figure 5 shows the comparison between raw data and denoised data. N2N can effectively suppress the random noise of the time-resolved response while preserving the texture fidelity.

The proposed method was extensively tested on N2N noise reduction, as shown in Fig. 3(c). Under the extreme low exposure case, the proposed method fits the PSF dictionary to noise instead of time-resolved information. Thus, reconstructed images cannot show actual structures and contain significant artifacts that block the observation. When reconstructing denoised raw data, the original structure can be restored more comprehensively. Significantly reduced background noise and artifacts result in more explicit images.

4.3. Dynamic experiments

We also imaged dynamic hidden scenes with the same experimental setup as the N2N validation. The first scene is a rotating statue, the laser illuminated the scattering wall at a fixed position, and the SPAD camera recorded the reflected photons with 100ms exposure time. The other dynamic scene is a rotating "U", and the experiment parameters are identical with the first scene. We recorded a frame every 200ms, and the results are shown in Fig. 4(a).

To assess the highest temporal resolution of the system, we tested with a varying exposure time of 1000ms, 500ms, 100ms and 50ms, as shown in Fig. 4(b). The intensity of the reconstructed data decreases accordingly. However, the results fail to show details when the exposure time is shorter than 50ms. Therefore, experiment results confirm the reliability of the proposed system and the reconstruction method. With the proposed NLOS imaging system, we can observe dynamic scenes at about 20 frames per second. Moreover, we also performed a contrast analysis for the proposed system (see Appendices). By varying the virtual aperture size and the scene's depth, we can observe factors other than the exposure time that affect the reconstruction quality. The contrast analysis results show that the spatial resolution of reconstructed images is positively related to the virtual aperture size of the scattering wall and negatively related to the scene's depth. Objects far from the scattering media are hard to reconstruct due to the weak signal and indiscernible response. The detailed analysis is illustrated in the Appendices.
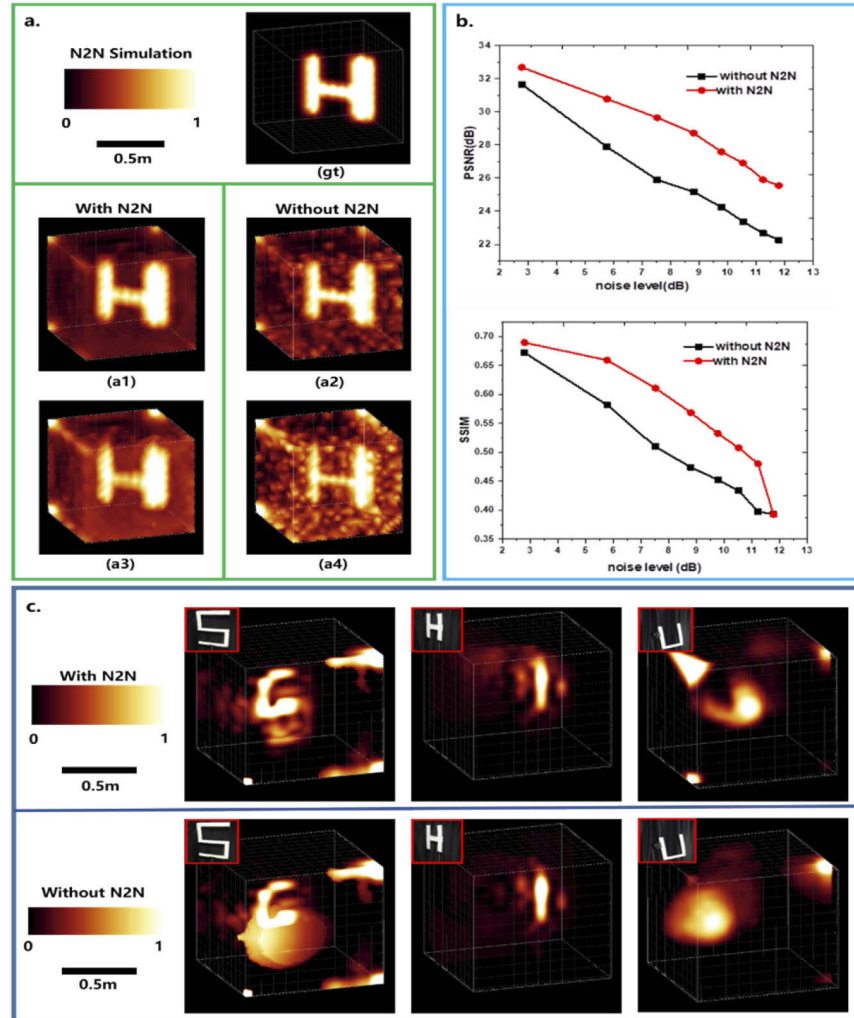
Fig. 3. Comparison of the results using the Noise2noise (N2N) method. **a.** Comparison with N2N denoising method in simulation data. The ground truth (gt) is provided for visual evaluation. Four results are showed for visual comparison, (a1) PSNR is 32.6 dB when noise level was 2.78 dB, (a2) PSNR is 31.7 dB when noise level is 2.78 dB, (a3) PSNR is 22.3 dB when noise level is 11.8 dB, (a4) PSNR is 25.5 dB when noise level was 11.8 dB. **b.** The quantitative comparison of PSNR and SSIM reconstruction in different noise level. When noise level is smaller than 12 dB, the N2N method can improve around 2 dB in PSNR and 0.1 in SSIM in reconstruction results. **c.** The Visual comparison of N2N method in experiment. The denoising method is able to ameliorate the reconstruction results visually in detailed structure.
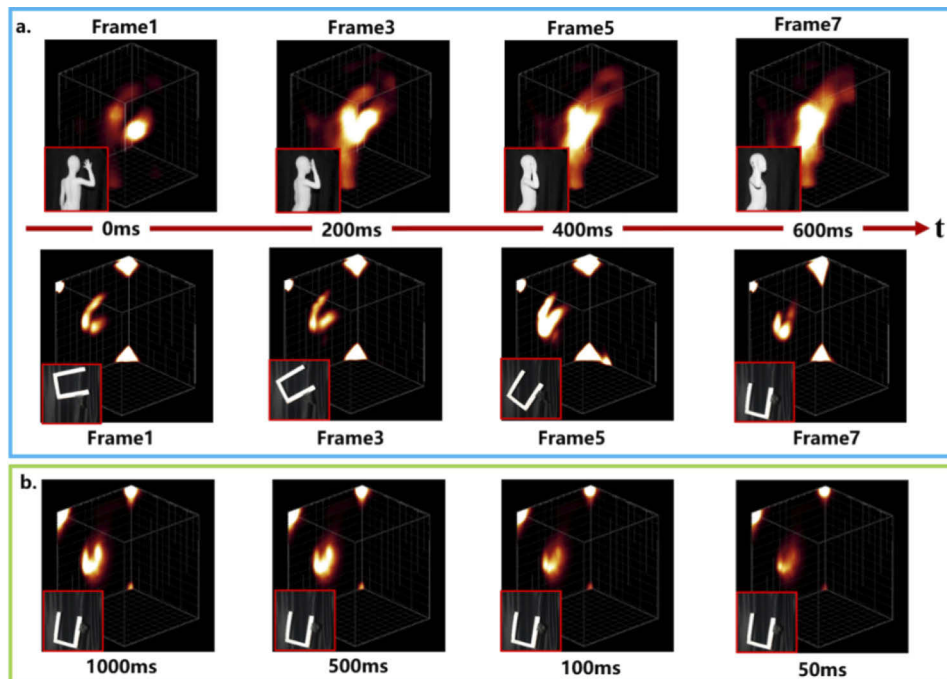
Fig. 4. Dynamic results and the highest achievable frame rate of the imaging system. **a.** Dynamic result of a rotating statue and a rotating "U". **b.** Test results of "U" with exposures of 1000ms, 500ms, 100ms and 50ms. The NLOS reconstruction results can preserve the structure if the exposure time is higher than 50ms.

5. Discussion

Nowadays, NLOS reconstruction can deliver excellent results through improved iterative back-projection solutions using a new rendering model and analysis method. Revolutionary reconstruction systems and algorithms have been improving NLOS imaging quality. For example, colorful NLOS imaging has also been implemented by a single photomultiplier tube [48,49]. Many methods have been proposed to increase the spatial resolution of the NLOS reconstruction, such as real-time transient imaging for amplitude modulated continuous wave LIDAR applications [50], analysis of missing features based on time-resolved NLOS measurements [51], convolutional approximations to incorporate priors into filtered back projection (FBP) [52], occlusion-aided NLOS imaging using SPADs [21,22], Bayesian statistics reconstruction for accounting for random errors [53], temporal focusing methods [31], compressed methods with different acquisition schemes [54]. The reconstruction time for these methods ranges from minutes to hours. To our knowledge, none of them has been applied successfully to the SPAD array camera without laser scanning. Also, the NLOS resolution is limited by the system temporal resolution (contributed from the SPAD and the laser). For a time-resolved imaging system, the time resolution, including the laser pulse width and the TDC temporal resolution around 100 ps leading to a theoretically achievable grid resolution of 3 cm in the hidden scene.

The proposed NLOS system with a SPAD array camera can globally record the hidden object's reflected waves. Thus, it can achieve NLOS reconstruction with a low exposure time without redundant spatial sampling. It shows great potential for real-time NLOS imaging due to its scanning-free mechanism. The novel reconstruction method based on the forward model of NLOS converts the reconstruction problem into an optimization problem. We divided the reconstruction

zone into grids where each voxel corresponds to an independent time-resolved response. We can reconstruct the hidden object from spatial-variant response through linear regression. Because the model can separate structural data from random noise, the reconstruction method shows anti-noise advantages compared with back-projection methods. Also, the reconstruction is further enhanced by neural network denoising tools, described in Appendix 2. By utilizing the fixed noise model of the SPAD camera, training a specific denoising model neural networks can drastically improve SNR. We can extract high SNR signals, assuming the time-resolved data is similar between two consecutive frames.

There are two approaches to improve construction performance. First, transient images are captured by a 32×32 SPAD camera, which corresponds to the same reconstructed resolution. Thus, sparse sampling highly restricts the spatial resolution. In future, we will design a SPAD camera with 240×160 pixels to improve the spatial resolution. We also plan to develop a novel reconstruction method to obtain high-quality NLOS images with 32×32 or fewer pixels based on compressed sensing. Second, the FoV of our camera system is only 82cm×82cm, limited by the off-the-shelf lens. With a custom-design lens, we will enlarge the FoV in the scattering wall to improve the reconstruction quality. Up to the present, the reconstruction method is a linear optimization based on gradient descent to optimize the spatial albedo. However, the calibration error of wall reflectivity and imperfect laser deviation is not considered. In the future, we will develop a novel reconstruction method further to uniformly optimize the system forward model and the spatial albedo for hidden objects. To further facilitate the reconstruction compatibility, we can also implement an NLOS imaging system from a low temporal resolution at the nanosecond level, paving the way for a practical and low-cost NLOS imaging system. Moreover, a remote NLOS imaging system is also meaningful for some practical applications.

In future works, we will adopt a SPAD array camera with low readout noise and develop a novel anti-noise method to improve the SNR of transient images to enhance the speed of NLOS imaging.

6. Conclusions

To implement dynamic NLOS imaging, we designed a novel NLOS imaging system based on a SPAD array camera. The new system can globally acquire high-quality transient images from hidden objects. A new reconstruction method was also proposed by utilizing an optimization-based approach. A linear regression algorithm was developed by converting the descattering problem into a spatial-varied correlation, achieving good results. Some simulations and analyses were made to compare PSNR, SSIM and TV performances with confocal and phase-field methods. Results show that our method has good PSNR, SSIM and TV performances, and our method can reconstruct hidden objects under a low SNR. Moreover, we also conducted experiments to image static and dynamic scenes, and images were reconstructed by the proposed method. Results show that our system can implement dynamic NLOS imaging at 20 fps. Hidden objects can be reconstructed under a short exposure time. Future developments will enable NLOS imaging with even higher spatial and temporal resolutions with rapid advances in semiconductor manufacturing and imaging methods.

Appendices

1. Synchronization

SPAD Array usually adopt time independent TCSPC, and the direct imaging causes strong tearing caused between pixels. Such drawback generally doesn't affect back-projection based method where each histogram is projected independently. However, the teared response is detrimental for pattern-based reconstruction. However, we need a calibration process so that a random shift of photon counting can be aligned.

2. 3D noise2noise method

One of the main drawbacks from Single Photon Avalanche Detector (SPAD) array imaging is the strong noise brought by the histogram-based imaging method. The response is based on probability of the avalanche at a particular time stamp, leading to strong mixed noise with high sparsity when the acquisition time is low. However, such noise model consists of many sources such as dark current, cross talking, jitter of laser, weak response signal, dead time, broken pixel and TDC counting error, which hinders building the noise model for measuring the expectation value. Since inaccurate noise estimation causes aliasing error when removing noise, it is very difficult to compute the real value under such complicate noise model. Also, Acquisition of the real value is also difficult in such system through long exposure. The long-exposed image still contains obvious sparse noise even when the memory is depleted for TCSPC counting. Thus, we have recovered the noised data without ground truth as reference.

Fortunately, the noise model is mostly time-invariant. If we sample noised time-resolved image in different time, all the data is based on the same noise model. Considering the complexity of the noise model and lack of supervisory, we adopt the noise2noise (N2N) unsupervised deep learning method for noise reduction for robustness. The network utilizes the fact that the expectation of multiple noised samples of the same scene is assumed to be the ground truth. By utilizing the high frame rate of SPAD array and assuming the content from adjacent frame is similar enough, we relaxed the same scene constraint into the similar scenes between the consecutive frames from SPAD array. By averaging the multiple video sequence pair through N2N, we are able to acquire smooth time-resolved data for NLOS reconstruction.

The N2N network removes the noise from a fixed noise model by utilizing "the average" of the two continues frames. When we only use two frames for training, the network finds the mapping from the input to the output. With paired data of continuous video, the noise is usually sparse and vastly varying while the scattered time-resolved signal varies slowly. Thus, to denoised the video series, we are finding the expectation model that fits every pair of consecutive frames formulated as:

$$\arg\min_{\theta} E_v\{L(f_\theta(v_i), v_{i+1})\} \tag{10}$$

where, f is the neural network, v_i and v_{i+1} is two consecutive time resolved data frame imaged by SPAD array, L is the loss function and E_v is the expectation of the Loss over the whole video sequence. In the training process, we let two random consecutive frames of a scattered video sequence to be one training pair. With one being the input data and the other being the label, the trained U-Network (Unet) fits the same part between the training data to label while diminishing the noise. The two frames are assumed to be similar enough because the time-resolved response from non-line-of-sight (NLOS) scene contains mostly low frequency components with a small shifting due to the motion. Because the high noise contains mostly in high frequency, the averaging is much more effective when exerting on the noise component instead of data with minor motion.

Because the input and output image contain small difference, directly using the mean square error results in the motion blur in scatter image, which might affect the reconstruction results. We designed the objective function of reconstruction to be,

$$L(x_i, x_{i+1}) = \alpha||f_\theta(x_i) - x_{i+1}||_2 + \beta||f_\theta(x_i) - G(x_i)||_2 \text{ where } \alpha + \beta = 1 \tag{11}$$

where, f is the mapping process of the neural network, x the input and output image and G is the wiener filter operator. The first term coerce the averaging between the two frames while diminishing the noise, the second term constrain the output to be similar to and noise filter input using conventional method. The objective function constrains the output to be similar to inputs while removing a complex system noise.

As the figure showed, the result in Fig. 5(a) contains noise mainly comes from the stochastic imaging process. The Fig. 5(b) is the denoised results from N2N. We can hardly discern the noise from the signal in the raw data because low SNR and high sparsity. However, the data does contain useful signal of the time-resolved response in the time-resolved image. After noise filtered by the N2N network, the results are more visualization friendly in texture sharpness and smoothness. In the low SNR zone where the noise become very sparse, The N2N network is still able to reconstruct the time-resolved response and reject strong and sparse noise.
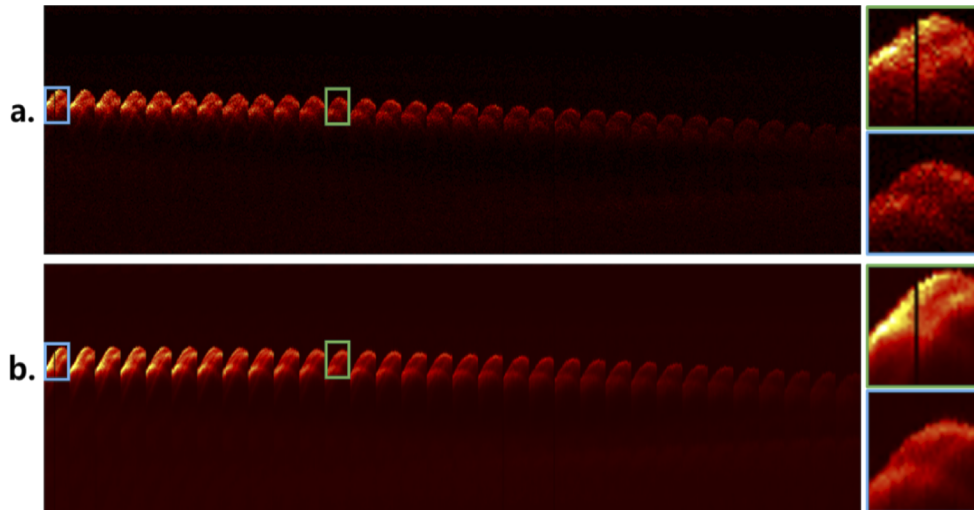

Fig. 5. N2N comparison of time-resolved raw data. **a.** The raw time-resolved data with exposure time 100ms in 50M laser repeating rate. **b.** Denoised data from N2N neural networks. The denoised data extract more clear texture in both strong and weak signal area (blue and green box) than raw data.

3. Experiment setup

We set up an experiment system shown in Fig. 6. The experiment mainly utilizes the photon-force "PF32" SPAD array and a femtosecond laser CFL-05RFF0 with 50M frame rate, 70fs pulse width, 780nm wavelength and 500mW power. The laser and detector are synchronized and delayed through a MPD Picosecond Delayer provided by the Photon Force. The camera is focus on the scattering relay wall with a 4mm CHIOPT-FA0401C lens. We used 2 reflection mirrors for controlling the ignition angle of laser. The side lobe from the output pulse is shaped through and the scattering wall is 1m × 1m, and the utilized virtual aperture is 0.82m × 0.82m. The reconstruction zone is 1.2m×1.2m×1.2m.

As the Fig. 6(a) shows, our NLOS system only has four simple components, a SPAD array camera, a femtosecond laser, a trigger relay and a pulse shaper. The pulse shaper is used to adjust the size of femtosecond spot avoiding the crosstalk in FOV of SPAD camera due to irregular laser speckles. In Fig. 6(b), the scattering wall in our system is made up of standard white Styrofoam.

4. Contrast analysis

We further analyzed the NLOS reconstruction ability through examine the contrast under extreme condition. By varying the field of view of the camera (aperture size) and the scene distance from the scattering wall, the calculated contrast is computed and compared to find the optimal working condition of the NLOS system. To test the reconstruction aptitude of our system, we set up a simple scene with 5 close points on a flat surface that is parallel to the scattering relay wall. The
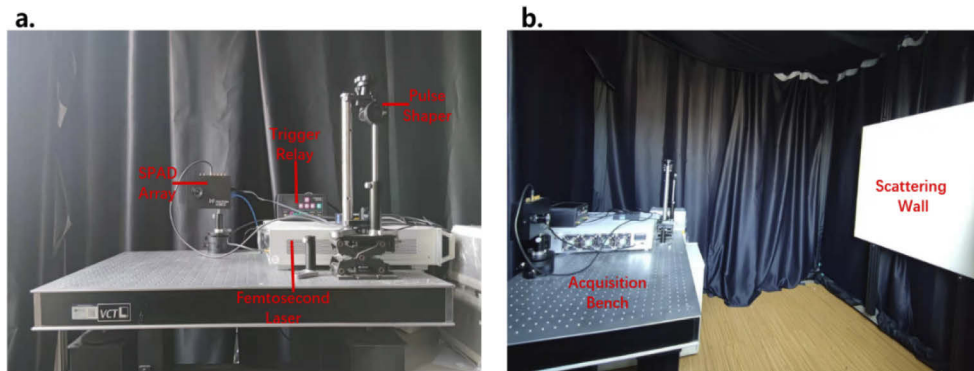
Fig. 6. Experiment setup. **a**. The acquisition bench contains: SPAD-Array, Femtosecond Laser, pulse shaping optics and a trigger delay **b.** The whole reconstruction site consists of the acquisition bench and the scattering wall.

contrast of the reconstructed results is computed as $\frac{I_{\max} - I_{min}}{I_{\max} + I_{min}}$ where the I_{max} and the I_{min} is the reconstructed maximum and minimum value. For the experiment setup, the spacing between scene points is fixed to 4cm and 6cm in vertical and horizontal direction. To compare the system ability under different condition, we conducted 3 experiments by varying the aperture size and the depth of scene with following setup, depth 0.3m, 0.6m and aperture size 0.7m and 0.85m. All the result is shown in Fig. 7.
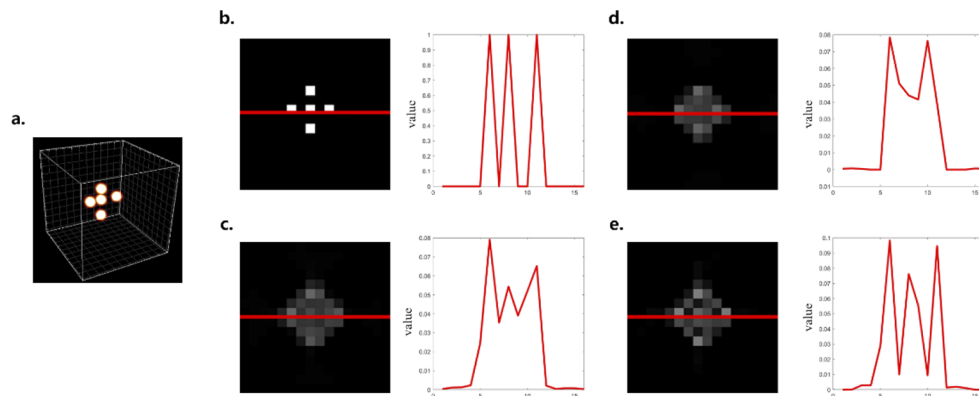

Fig. 7. Contrast analysis. **a.** The ground truth is 5 points on a vertical surface in 3d space parallel to the relay wall. The lateral spacing of the points is fixed to 4 cm and 6 cm, where the depth of the surface varies between 30 cm and 60 cm. The virtual aperture of the relay wall varies between 70 cm and 85 cm. **b.** The ground truth contrast response on the reconstruction surface. The contrast is 1. **c.** The reconstructed contrast response when depth is set to 0.3 m and aperture is set to 0.7 m. The contrast is 0.382. **d.** The reconstructed contrast response when depth is set to 0.6 m and aperture is set to 0.7 m. The contrast is not applicable since the reconstructed result is not correct in structure. **e.** The reconstructed contrast response when depth is set to 0.3 m and aperture is set to 0.85 m. The contrast is 0.82.

As shown in Fig. 7, the Fig. 7(a) is the ground truth in reconstruction zone and Fig. 7(b) is the contrast analysis for ground truth. In this situation, the contrast is 1. From Fig. 3, we found out the reconstruction results is highly depending on the aperture and the depth, when we keep the same aperture to be 0.7m, increasing the scene depth will decrease the resolution since the energy

decreases and the curvature of time-resolved response gradually decreased to plane surface. As Fig. 7(c) and Fig. 7(d) shows, the system loses the reconstruction ability when the scene distance is increased from 0.3m to 0.6m. The central point became not discriminable and leading the contrast calculation not applicable. Also, the contrast become higher when the aperture size enlarges, which also increase the curvature of the time-resolved response. As Fig. 7(c) and Fig. 7(e) shows, the contrast increased from 0.38 to 0.82 when the aperture size increased from 0.7m side length to 0.85m. The more curved response leads to a more discernible reconstruction during the reconstruction.

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. A. Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, G. Bawendi, and R. Raskar, "Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging," Nat. Commun. **3**(1), 745 (2012).
2. M. O'Toole, D. B. Lindell, and G. Wetzstein, "Confocal non-line-of-sight imaging based on the light-cone transform," Nature **555**(7696), 338–341 (2018).
3. X. Liu, I. Guillén, M. La Manna, J. Nam, S. Reza, T. Huu Le, A. Jarabo, D. Gutierrez, and A. Velten, "Non-line-of-sight imaging using phasor-field virtual wave optics," Nature **572**(7771), 620–623 (2019).
4. D. B. Lindell, G. Wetzstein, and M. O'Toole, "Wave-based non-line-of-sight imaging using fast f-k migration," ACM Trans. Graph. **38**(4), 1–13 (2019).
5. C. Wu, J. Liu, X. Huang, Z. Li, C. Yu, J. Ye, J. Zhang, Q. Zhang, X. Dou, V. Goyal, F. Xu, and J. Pan, "Non-line-of-sight imaging over 1.43 km," Proc. Natl. Acad. Sci. **118**(10), e2024468118 (2021).
6. B. Wang, M. Zheng, J. Han, X. Huang, X. Xie, F. Xu, Q. Zhang, and J. Pan, "Non-Line-of-Sight Imaging with Picosecond Temporal Resolution," Phys. Rev. Lett. **127**(5), 053602 (2021).
7. C. Jin, J. Xie, S. Zhang, Z. Zhang, and Y. Zhao, "Richardson-Lucy deconvolution of time histograms for high-resolution non-line-of-sight imaging based on a back-projection method," Opt. Lett. **43**(23), 5885–5888 (2018).
8. V. Arellano, D. Gutierrez, and A. Jarabo, "Fast back-projection for non-line of sight reconstruction," Opt. Express **25**(10), 11574–11583 (2017).
9. X. Liu, S. Bauer, and A. Velten, "Phasor field diffraction based reconstruction for fast non-line-of-sight imaging systems," Nat. Commun. **11**(1), 1645 (2020).
10. D. B. Lindell, G. Wetzstein, and V. Koltun, "Acoustic non-line-of-sight imaging," In *Proc. IEEE Conf. Computational. Vis. Pattern Recognition*, 6773–6782 (2019).
11. K.L. Bouman, V. Ye, A. Yedidia, F. Durand, G. Wornell, A. Torralba, and W. Freeman, "Turning corners into cameras: principles and methods," In *Proc. IEEE Int. Conf. Comput. Vis.*, 2289–2297 (2017).
12. C. Saunders, J. Murray-Bruce, and V. K. Goyal, "Computational periscope with an ordinary digital camera," Nature **565**(7740), 472–475 (2019).
13. J. Boger-Lombard and O. Katz, "Non-line-of-sight localization by passive optical time-of-flight," Nat. Commun. **10**, 3343 (2019).
14. J. Bertolotti, E. G. Van Putten, C. Blum, Ad. Lagendijk, W. L. Vos, and A. P. Mosk, "Non-invasive imaging through opaque scattering layers," Nature **491**(7423), 232–234 (2012).
15. O. Katz, P. Heidmann, M. Fink, and S. Gigan, "Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations," Nature Photon **8**(10), 784–790 (2014).
16. T. Maeda, Y. Wang, R. Raskar, and A. Kadambi, "Thermal non-line-of-sight imaging," In *Proc. IEEE Int. Conf. Comput. Photogr.*, 1–11 (2019).
17. M. Kaga, T. Kushida, K. Takatani, T. Tanaka, Y. Funatomi, and Mukaigawa, "Thermal non-line-of-sight imaging from specular and diffuse reflections," IPSJ Trans. Comp. Vis. Appl. **11**(1), 8 (2019).
18. J. A. Teichman, "Phasor field waves: a mathematical treatment," Opt. Express **27**(20), 27500–27506 (2019).
19. S. A. Reza, M. L. Manna, S. Bauer, and A. Velten, "Phasor field waves: experimental demonstrations of wave-like properties," Opt. Express **27**(22), 32587–32608 (2019).
20. J. Dove and J. H. Shapiro, "Paraxial theory of phasor-field imaging," Opt. Express **27**(13), 18016–18037 (2019).
21. F. Heide, M. O'Toole, K. Zang, D. Lindell, S. Diamond, and G. Wetzstein, "Non-line-of-sight imaging with partial occluders and surface normal," ACM Trans. Graph. **38**(3), 1–10 (2019).
22. F. Xu, G. Shulkind, C. Thrampoulidis, J. H. Shapiro, A. Torralba, F. Wong, and G. W. Wornell, "Revealing hidden scenes by photon-efficient occlusion-based opportunistic active imaging," Opt. Express **26**(8), 9945–9962 (2018).
23. S.W. Seidel, Y. Ma, J. Murray-Bruce, C. Saunders, W. Freeman, C. Yu, and V. Goyal, "Corner occluder computational periscopy: estimating a hidden scene from a single photograph," In *Proc. IEEE Int. Conf. Comput. Photogr.*, 1–9 (2019).
24. M. Baradad, V. Ye, A. Yedidia, F. Durand, W. Freeman, G. Wornell, and A. Torralba, "Inferring light fields from shadows," In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 6267–6275 (2018).

25. S. Young, D. Lindell, and G. Wetzstein, "Non-line-of-sight surface reconstruction using the directional light-cone transform," In *IEEE Conf. Comput. Vis. Pattern Recognit.*, 1407–1416 (2020).
26. C.-Y. Tsai, A.C. Sankaranarayanan, and I. Gkioulekas, "Beyond volumetric albedo -a surface optimization framework for non-line-of-sight imaging," In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1545–1555 (2019).
27. C. A. Metzler, F. Heide, P. Rangarajan, M. M. Balaji, A. Viswanath, A. Veeraraghavan, and R. G. Baraniuk, "Deep-inverse correlography: towards real-time high-resolution non-line-of-sight imaging," Optica **7**(1), 63–71 (2020).
28. C.-Y. Tsai, K.N. Kutulakos, S.G. Narasimhan, and A.C. Sankaranarayanan, "The geometry of first- returning photons for non-line-of-sight imaging," In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 7216–7224 (2017).
29. A . K. Pediredla, M. Buttafava, A. Tosi, O. Cossairt, and A. Veeraraghavan, "Reconstructing rooms using photon echoes: a plane based model and reconstruction algorithm for looking around the corner," In *Proc. IEEE Int. Conf. Comput. Photogr.*, 1–12 (2017).
30. I. Starshynov, O. Ghafur, J. Fitches, and D. Faccio, "Coherent control of light for non-line-of-sight imaging," Phys. Rev. Appl. **12**(6), 064045 (2019).
31. A. Pediredla, A. Dave, and A. Veeraraghavan, "Snlos: Non-line-of-sight scanning through temporal focusing," In *Proc. IEEE Int. Conf. Comput. Photogr.*, 1–13 (2019).
32. O. Gupta, T. Willwacher, A. Velten, A. Veeraraghavan, and R. Raskar, "Reconstruction of hidden 3D shapes using diffuse reflections," Opt. Express **20**(17), 19096–19108 (2012).
33. D. Wu, G. Wetzstein, C. Barsi, T. Willwacher, M. O'Toole, N. Naik, Q. Dai, K. Kutulakos, and R. Raskar, "Frequency analysis of transient light transport with applications in bare sensor imaging," In *Proc. 12th Eur. Conf. Comput. Vis.* 7572, 542–555 (2012).
34. F. Heide, L. Xiao, W. Heidrich, and M.B. Hullin, "Diffuse mirrors: 3D reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors," In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 3222–3229 (2014).
35. M. Tancik, T. Swedish, G. Satat, and R. Raskar, "Data-driven non-line-of-sight imaging with a traditional camera," *Imaging Appl. Opt.* (Optical Society of America, 2018).
36. W. Chen, S. Daneau, F. Mannan, and F. Heide, "Steady-state non-line-of-sight imaging," In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 6790–6799 (2019).
37. P. Caramazza, A. Boccolini, D. Buschek, M. Hullin, CF. Higham, R. Henderson, R. Murray-Smith, and D. Faccio, "Neural network identification of people hidden from view with a single-pixel, single-photon detector," Sci Rep **8**(1), 11945 (2018).
38. J. Tasinkevych and I. Trots, "Circular radon transform inversion technique in synthetic aperture ultrasound imaging: an ultrasound phantom evaluation," Arch. Acoust. **39**(4), 569–582 (2015).
39. S. Moon, "On the determination of a function from an elliptical radon transform," J. Math. Anal. Appl. **416**(2), 724–734 (2014).
40. M. Buttafava, G. Boso, A. Ruggeri, A. D. Mora, and A. Tosi, "Time-gated single-photon detection module with 110 ps transition time and up to 80 MHz repetition rate," Rev. Sci. Instrum. **85**(8), 083114 (2014).
41. M. Laurenzis and A. Velten, "Feature selection and back-projection algorithms for non-line-of-sight laser-gated viewing," J. Electron. Imaging **23**(6), 063003 (2014).
42. D. Faccio, A. Velten, and G. Wetzstein, "Non-line-of-sight imaging," Nat. Rev. Phys. **2**(6), 318–327 (2020).
43. G. Gariepy, F. Tonolini, R. Henderson, J. Leach, and D. Faccio, "Detection and tracking of moving objects hidden from view," Nat. Photonics **10**(1), 23–26 (2016).
44. X. Feng and L. Gao, "Ultrafast light field tomography for snapshot transient and non-line-of-sight imaging," Nat. Commun. **12**(1), 2179 (2021).
45. J. Richardson, R. Walker, L. Grant, D. Stoppa, F. Borghetti, E. Charbon, M. Gersbach, and R. Henderson, "A 32×32 50ps resolution 10-bit time to digital converter array in 130 nm CMOS for time correlated imaging," *IEEE Custom Integrated Circuits Conference*, 77–80, (2009)
46. https://www.photon-force.com/pf32-time-resolved-single-photon-counting-camera-readout/.
47. J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2Noise: Learning Image Restoration without Clean Data," In *International Conference on Machine Learning*, 2965–2974 (2018).
48. G. Musarra, A. Lyons, E. Conca, Y. Altmann, F. Villa, F. Zappa, and M. J. Padgett, "Non-line-of-sight Three-dimensional imaging with a single pixel camera," Phys. Rev. Appl. **12**(1), 011002 (2019).
49. G. Musarra, A. Lyons, E. Conca, F. Villa, F. Zappa, Y. Altmann, and D. Faccio, "3D RGB Non-Line-Of-Sight single-pixel imaging," *Imaging and Applied Optics 2019 (COSI, IS, MATH, pcAOP), OSA Technical Digest (Optical Society of America)*, 2019.
50. C. Peters, J. Klein, M. B. Hullin, and R. Klein, "Solving trigonometric moment problems for fast transient imaging," ACM Transactions on Graphics **34**(6), 1–11 (2015).
51. X. Liu, S. Bauer, and A. Velten, "Analysis of feature visibility in non-line-of-sight measurements," In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 10140–10148 (2019).
52. B. Ahn, A. Dave, A. Veeraraghavan, I. Gkioulekas, and A. C. Sankaranarayanan, "Convolutional Approximations to the General Non-Line-of-Sight Imaging Operator," In *Proc. IEEE International Conference on Computer Vision*,7889–7899 (2019).
53. L. Huang, X. Wang, Y. Yuan, S. Gu, and Y. Shen, "Improved algorithm of non-line-of-sight imaging based on the Bayesian statistics," JOSA A **36**(5), 834–838 (2019).

54. J. T. Ye., X. Huang, Z. P. Li., and F. Xu., "Compressed sensing for active non-line-of-sight imaging," Opt. Express **29**(2), 1749–1763 (2021).