Passive non-line-of-sight imaging for moving targets with an event camera

Conghe Wang (王丛赫)[1,†], Yutong He (贺与同)[2,†], Xia Wang (王 霞)[1*], Honghao Huang (黄泓皓)[2], Changda Yan (闫昌达)[1], Xin Zhang (张 鑫)[1], and Hongwei Chen (陈宏伟)[2**]

[1] Key Laboratory of Photoelectronic Imaging Technology and System of Ministry of Education of China, School of Optics and Photonics, Beijing Institute of Technology, Beijing 100081, China

[2] Beijing National Research Center for Information Science and Technology (BNRist), Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

*Corresponding author: angelniuniu@bit.edu.cn
**Corresponding author: chenhw@tsinghua.edu.cn

Non-line-of-sight (NLOS) imaging is an emerging technique for detecting objects behind obstacles or around corners. Recent studies on passive NLOS mainly focus on steady-state measurement and reconstruction methods, which show limitations in recognition of moving targets. To the best of our knowledge, we propose a novel event-based passive NLOS imaging method. We acquire asynchronous event-based data of the diffusion spot on the relay surface, which contains detailed dynamic information of the NLOS target, and efficiently ease the degradation caused by target movement. In addition, we demonstrate the event-based cues based on the derivation of an event-NLOS forward model. Furthermore, we propose the first event-based NLOS imaging data set, EM-NLOS, and the movement feature is extracted by time-surface representation. We compare the reconstructions through event-based data with frame-based data. The event-based method performs well on peak signal-to-noise ratio and learned perceptual image patch similarity, which is 20% and 10% better than the frame-based method.

Keywords: non-line-of-sight imaging; event camera; event-based representation.
DOI: 10.3788/COL202321.061103

1. Introduction

Non-line-of-sight (NLOS) imaging has attracted great attention with its widespread potential applications in object detection, autonomous driving, and anti-terrorist reconnaissance[1–3]. According to whether a controllable light source is used, NLOS imaging is classified as active NLOS imaging[4,5] and passive NLOS imaging[6,7].

Passive NLOS imaging shows promising application and research prospects due to its simple device and convenient data acquisition. However, the NLOS problem is known as an inverse problem in mathematics, and we need to perform blind deconvolution, which is time-consuming and computationally burdened. Consequently, light-cone transform theory using matrix inverse[8], back projection algorithm based on photon time-of-flight information[9], and wave-based phasor field approach[10] are proposed successively. But few of these methods perform well in passive NLOS moving target reconstruction because the steady-state detection mode of passive NLOS suffers from serious degradation of the diffusion spot on the relay surface[11] and the superposition effect of isotropic diffuse reflection

by close pixels[12]. Currently, speckle coherence restoration[13] and intensity-based data-driven reconstruction methods[14,15] are used to solve the ill-posed passive NLOS imaging dilemma. Since the movement of the target induces motion blur to the intensity distribution on the relay surface and superposes with the obscurity caused by diffusion[16], the current end-to-end deep-learning approach[11,17] performs well only on static NLOS targets[18] but shows defects in reconstruction quality for moving targets. In contrast, to realize high quality and efficient reconstruction, we deduce the event form detection-forward model of passive NLOS and establish the event-based inverse problem, on the basis of which we first put forward the event cues for passive NLOS moving target reconstruction. In this way, the dynamic information of the intensity diffusion is precisely captured by the event detection paradigm.

2. Principle and Methods

In this section, we present the working principle of the event camera and then explain the inverse problem setup by derivation of the forward model in passive NLOS imaging.

2.1. Event-based vision

The event camera[19], known as a novel neuromorphic vision device, only responds to brightness changes per pixel asynchronously, while traditional frame-based cameras measure absolute brightness at a fixed rate. The record paradigm of event-based vision provides high temporal resolution, high dynamic range, and low power consumption[20]. Therefore, it finds great potential in challenging scenarios for standard cameras, such as high speed, high dynamic range imaging[21], or object detection[22] in a slightly changing optical field.

The working principle of the event camera is illustrated in Fig. 1. When the logarithmic intensity of the brightness changes reaches the trigger threshold, the pixel will be triggered and recorded as an event. Each pixel records the logarithmic form of the intensity when an event is fired and continuously monitors for sufficient amplitude changes based on this stored value. The event is described by four characteristic parameters: t describes the time stamp of the fired event, which records the moment when the logarithmic value of the intensity changes exceeds the threshold; x, y describes the spatial address of the pixel where the event is fired, respectively; p describes the reason for the excitation of the event, and increase or decrease in brightness, respectively.

Data collected by an event camera are recorded in the form of four characteristic parameters, time stamp t, spatial address x, y, and polarity flag p. The ith event is noted as ev_i,

$$\mathrm{ev}_i = [t_i, x_i, y_i, p_i]^T, \quad i \in N. \tag{1}$$

2.2. Forward model of passive NLOS

The forward model of NLOS is a modeling expression of the light transport in data collecting, which could be regarded as the inverse process of NLOS imaging. It makes up of the theoretical basis of NLOS reconstruction from the scope of steady-state imaging and derivation.

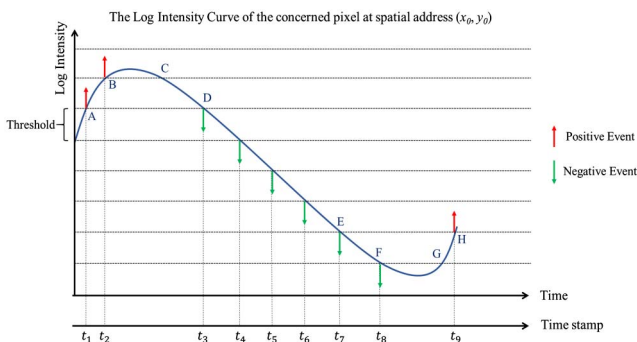The intensity of the diffusion spot on the relay surface[23] could be expressed as

The Log Intensity Curve of the concerned pixel at spatial address (x_0, y_0)

Fig. 1. Schematic of events excitation with changes in brightness at pixel level. The red arrow stands for a fired positive event, while the green arrow stands for the negative one.

$$I(p_y) = \iint_{p_f \in F} A(p_f, p_y) \cdot I(p_f) \mathrm{d}p_f + N, \tag{2}$$

where $I(p_y)$ represents the intensity distribution in the field of view (FoV) on the relay surface, $I(p_f)$ represents the intensity of the self-luminous target, N represents the summation of the ambient light noise and random noise in detection, and $A(p_f, p_y)$ stands for the point-to-point transmission process[23], which could be specialized as

$$A(p_f, p_y) = \frac{\cos[\angle(\vec{p_y} - \vec{p_f}, \vec{n_{pf}})] \cdot \cos[\angle(\vec{p_f} - \vec{p_y}, \vec{n_{py}})]}{\|\vec{p_y} - \vec{p_f}\|_2^2} \cdot \mu(p_f, p_y), \tag{3}$$

where $\angle(\vec{p_y} - \vec{p_f}, \vec{n_{pf}})$ and $\angle(\vec{p_f} - \vec{p_y}, \vec{n_{py}})$ donate the angle between the vector formed by the monitor pixel and the light source with the normal vector of each of their located plane, correspondingly. The cosine function of these two angles characterizes the effect on optical transmission by geometric relations. The Euclidean norm is used to quantify the attenuation of range. We assume that the relay surface can be approximately modeled as isotropic diffuse reflection, and coefficient μ, which describes the effect of the bidirectional reflectance distribution function (BRDF), is a constant. We substitute Eq. (3) into Eq. (2), and present the pixel-wise detection function on the relay surface[11],

$$I(p_y) = \iint_{p_f \in F} \frac{\cos[\angle(\vec{p_y} - \vec{p_f}, \vec{n_{pf}})] \cdot \cos[\angle(\vec{p_f} - \vec{p_y}, \vec{n_{py}})]}{\|\vec{p_y} - \vec{p_f}\|_2^2}$$
$$\cdot \mu \cdot I(p_f) \mathrm{d}p_f + N, \tag{4}$$

which could be simplified as

$$y = Af + N. \tag{5}$$

The detection function of diffusion on the relay surface could be written in the form of the matrix,

$$\begin{bmatrix} I_{y,1} \\ \vdots \\ I_{y,h \times w} \end{bmatrix} = \begin{bmatrix} A_{1,1} & \cdots & A_{1,H \times W} \\ \vdots & & \vdots \\ A_{h \times w,1} & \cdots & A_{h \times w,H \times W} \end{bmatrix} \begin{bmatrix} I_{f,1} \\ \vdots \\ I_{f,H \times W} \end{bmatrix} + N, \tag{6}$$

where the intensity of pixel i in the FoV is expressed by the elements $I_{y,i}$, while $I_{f,i}$ represents the intensity of pixel i of the self-luminous target. The transmission matrix A is discretized as $A_{i,j}$, and $h \times w$, $H \times W$ are the pixel range for FoV and target, respectively.

2.3. Event-based reconstruction method

In our work, we adopt the event cues to extract the movement information and texture feature of the targets. The most intuitive

representation of the dynamic information on the relay surface is the origin event data. However, as addressed in Eq. (1), an event is composed of the triggered time stamp, spatial address, and polarity flag, which belong to three different data formats, respectively. As a result, it is critical to convert the sparse 4D event points into a featured tensor that contains both temporal and spatial characteristics. Therefore, we adopt the time-surface[24] method to represent the event-based data and extract the featured diffusion spot, which contains rich information on the target movement.

We visualize the event-based data to demonstrate their representations. The relay surface is selected as a mirror for clear visualization. The target "A" is moving from left to right in the FoV, as shown in Fig. 2.

In Fig. 2(b), the frame-based image is captured by a traditional camera, while Fig. 2(c) shows the asynchronous events captured by an event camera, consisting of 3D scatter points, which represent the parameters x, y, and t, in which red stands for positive polarity and blue stands for negative polarity. We select a short time interval as the temporal length of the voxel grid and plot the time-surface 3D map with the time-surface calculation[24], which contains both temporal and spatial correlations of the selected events, as shown in Fig. 2(e). If we normalize the time-surface value of each spatial address, project it into xoy plane, and display it in the form of gray-scale intensity, we could express the spatiotemporal correlation information in the form of an event address map[25,26], as shown in Fig. 2(d).

Based on the detection function in Eq. (6), we represent the intensity of pixel i in the FoV at the current moment by elements $I_{y,i}$ in the matrix, while $I'_{y,i}$ represent the intensity of pixel i at the center time stamp of the adjacent voxel grid. According to the working principle of the event camera, we establish the event-based detection function by performing a difference operation to $I'_{y,i}$ and $I_{y,i}$, as shown,

$$
\begin{bmatrix}
I'_{y,1} - I_{y,1} \\
I'_{y,2} - I_{y,2} \\
\vdots \\
I'_{y,h\times w} - I_{y,h\times w}
\end{bmatrix}
= A
\begin{bmatrix}
I'_{f,1} - I_{f,1} \\
I'_{f,2} - I_{f,2} \\
\vdots \\
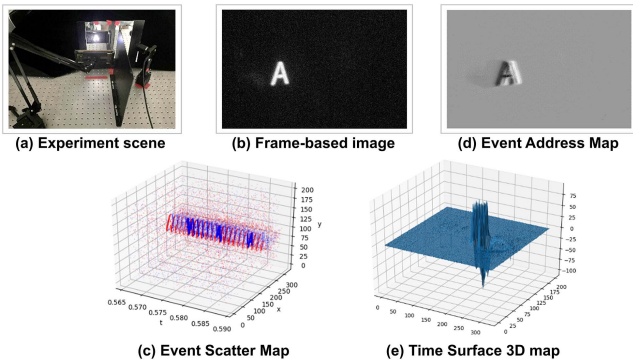I'_{f,H\times W} - I_{f,H\times W}
\end{bmatrix}
+ N' - N, \quad (7)
$$

and the intensity of the brightness changes at pixel i is expressed as

$$
I'_{y,i} - I_{y,i} = \begin{bmatrix} A_{i,1} & A_{i,2} & \cdots & A_{i,H\times W} \end{bmatrix} \lim_{\Delta t \to t_0} \cdot \Delta f + N_r
$$

$$
= \begin{bmatrix} A_{i,1} & A_{i,2} & \cdots & A_{i,H\times W} \end{bmatrix}
\begin{bmatrix}
I^m_{f,1} \\
I^m_{f,2} \\
\vdots \\
I^m_{f,H\times W}
\end{bmatrix}
+ N_r, \quad (8)
$$

where t_0 is the time interval between two adjacent voxel grids, and we use the limitation form to show the accumulation of the target intensity changes at pixel i during t_0, which can be written as $I^m_{f,i}$, containing the movement information on the target. The noise caused by ambient light is also counteracted by the sampling principle of event-based detection, and only random noise remains, i.e., $N_r = N' - N$.

Then, we utilize the step function $\text{Fire}(I'_y - I_y)$ to discriminate whether an event is fired. As shown in Fig. 3, when the difference value of intensity exceeds the threshold and falls in the red or blue area, an event is recorded at that pixel and outputs $[t_i, x_i, y_i, p_i]$.

After extracting the dynamic information on the diffusion spot movements, we put forward an event-embedded framework that fuses the extracted event features with a UNet structure to solve the inverse problem. As shown in Fig. 4, we display a video containing a parallel moving target with a smartphone, and leverage the event-based vision to record the dynamic diffusion spot on the relay surface.

We perform a time-surface calculation on the voxel grid[21] to extract the featured diffusion spot and represent event data at different time intervals by a series of 2D intensity images. Time-surface[24] is expressed by

$$
S_i(\rho, p) = e^{-[t_i - T_i(\rho,p)]/\tau}, \quad (9)
$$

where S_i is the time-surface value of ev_i, which is defined by applying an exponential decay kernel with time constant τ on the values of the context time stamp $T_i(\rho, p)$, where p is the polarity flag. T_i is defined to represent the time-context information around an incoming event ev_i as the array of most recent
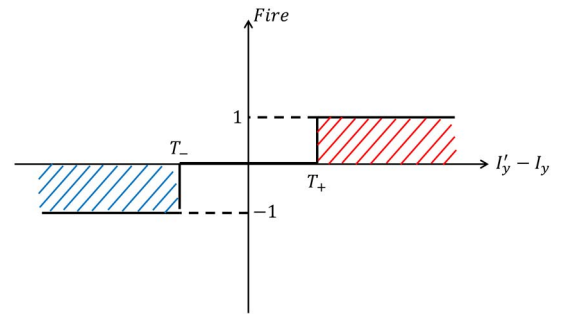

Fig. 2. Representations of event-based data.

Fig. 3. Discrimination function of fired event.
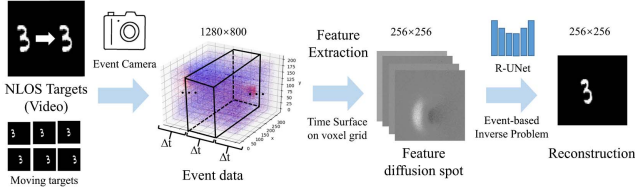
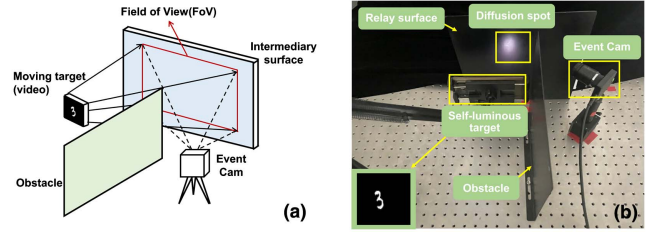**Fig. 4.** Flow chart of event-embedded passive NLOS imaging.

Fig. 5. Experimental setup. (a) Basic principle of our NLOS scene; (b) experimental settings; the self-luminous target is a video with moving digits.

events times at t_i for the neighboring pixels in the adjacent area with the radius of ρ,

$$T_i(\rho, p) = \max_{j \leq i}\{t_j | r_j = r_i + \rho, p_i = p\}. \tag{10}$$

According to the assumption that the fired events of the adjacent voxel grids share a similar spatial address, the spatial address of the fired event stream is highly correlated. As a result, when we perform a time-surface calculation on the acquired event data in a voxel grid, the context information on the movements is accumulated. Then, the temporal and spatial features of the moving target can be expressed by

$$\text{Emap} = S\left[\sum_{y \in \text{FoV}, \Delta t \to t_0} \text{Fire}(I'_y - I_y)\right] = B \cdot f, \tag{11}$$

where Fire($\cdot$) is referred to in Fig. 3, S stands for the time-surface representation, and B is the counterpart of A.

The event-based inverse problem is established by

$$f = B^{-1} \cdot \text{Emap}. \tag{12}$$

A typical solution of this reconstruction is using optimal methods to solve the matrix inverse of B, as shown,

$$f = \underset{f}{\arg\min}\|Bf - \text{Emap}\|_2^2 + J, \tag{13}$$

where J donates the *a priori*, which is obtained by the asynchronous sampling paradigm of the event camera.

However, the condition number of matrix B is relatively large due to the pixel-wise mutual interference, resulting in the low rank of matrix B. Also, iteratively solving the calculation of B^{-1} is time-consuming. For the last step of solving the inverse problem in NLOS reconstruction, we employ a UNet structure, use skip connection to perform multiscale feature fusion, and add a residual block to avoid gradient explosion when training.

3. Experimental Setups

For the experimental proof of the event-based approach, we constructed experimental setups, as shown in Fig. 5. We displayed a video that provided the self-luminous moving target in the NLOS region of the event camera, blocked by the obstacle. The moving diffusion spot on the relay surface is recorded by the event camera (CeleX-V) in Section 1 of the Supplementary Material and Visualization 1.

The targets used in the experiment contain characters of number digits selected from the MNIST training set, MNIST test set, and PRINT test set (Arial font numbers), with the size of 3 cm × 3 cm, which are placed 25 cm away from the frosted aluminum fender.

We select 14 different kinds of characters for each digit (0–9) in the MNIST training set and test set, and then acquire both event-based data and frame-based data with different modes of the CeleX-V camera. The self-luminous target displayed by the smartphone translates from left to right in the FoV at a preset speed of 2.5 cm/s. When recording the moving diffusion spot in full-picture (F) mode, we get series of screenshots in different positions with the frame rate of 100 frames per second, while in event-intensity (EI) mode, we get a stream of event-based data of the diffusion spot movement. Data collected by these two modes are calibrated to the ground truth by the time stamp and made into image format data sets event MNIST NLOS (EM-NLOS) and frame MNIST NLOS (FM-NLOS), correspondingly.

To the best of our knowledge, we first established the EM-NLOS data set, which contains 4080 images in the training set and validation set and 210 images in the test set. The training set is made up of 3950 featured events. The time-surface map has 130 targets (13 groups, 0–9) at different positions, while the validation set contains 130 images. The test set consists of 110 images with 10 digits (0–9) selected from the MNIST test set and 100 images with 10 digits (0–9) in Arial font. As the counterpart, FM-NLOS contains the corresponding frame-based intensity diffusion spot movement, with 4180 images in total. We compare the training results on EM-NLOS and FM-NLOS, which are noted as an event-based method (E method) and a frame-based method (F method), respectively. The reconstruction accuracies of the F method and the E method on the MNIST test set and the PRINT test set are shown and compared in Fig. 6.

We trained our residual-UNet (R-UNet) on the EM-NLOS training set with an adaptive moment (Adam) estimation optimizer with Nvidia RTX 3090 GPU for 800 epochs. To achieve fair comparisons, we trained the R-UNet on the FM-NLOS training set with the same configuration[17]. The structure of our R-UNet and training parameters are given in Section 2 of the Supplementary Material.
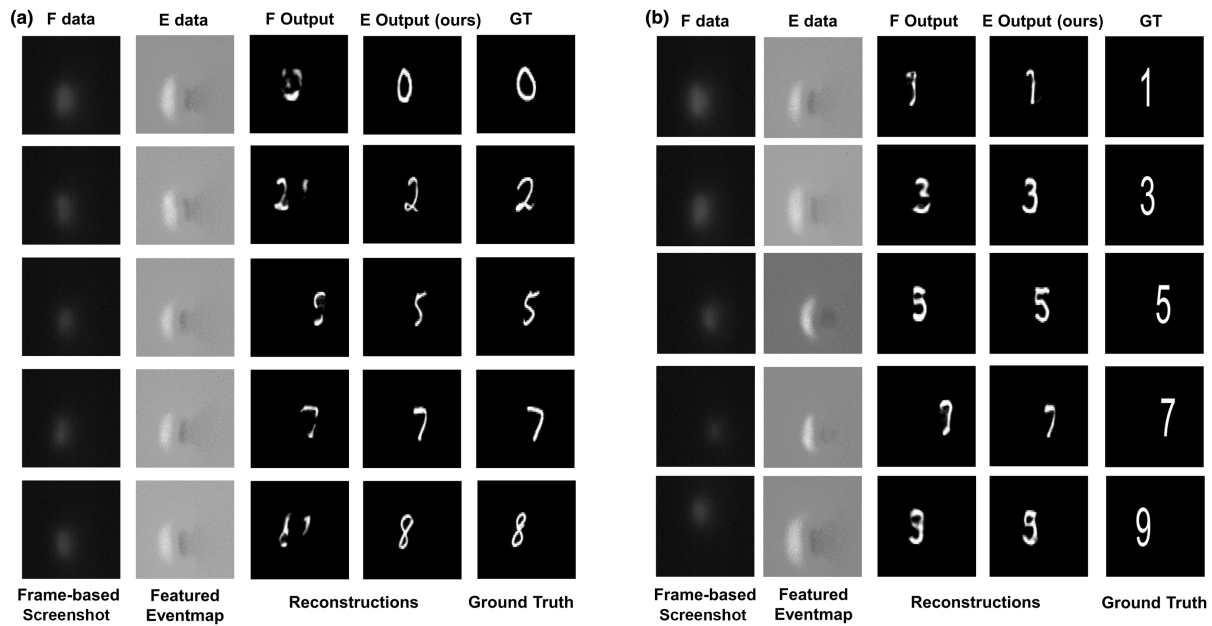
Fig. 6. (a) Part of the reconstruction results for the PRINT test set in both EM-NLOS and FM-NLOS; (b) part of the reconstruction results for the MNIST test set in both EM-NLOS and FM-NLOS.

4. Experimental Results and Discussions

The reconstruction quality of the moving target is assessed from two perspectives: the visual reconstruction quality and the position accuracy. For the former, we introduce the peak signal-to-noise ratio (PSNR) and learned perceptual image patch similarity (LPIPS)[27] to evaluate the reconstructions. The reconstruction accuracies of the F method and the E method on the MNIST test set and the PRINT test set are shown and compared in Figs. 6(a) and 6(b), respectively. It is obvious that the proposed E method with event-based data shows much better reconstruction quality than the F method, especially in recognizing the digits.

As for the position accuracy, we define the index contour distance (Cd) to measure the position of the reconstructions. The Cd value is evaluated by the average distance between the left edge and the digit left contour (made up by first pixels with a gray scale of 255 in each row after image binarization). Digit 7 in the MNIST test set and digit 3 in the PRINT test set are demonstrated as examples in this Letter. As shown in Fig. 7, the reconstruction by the E method is closer to ground truth than that of the F method. The average Cd deviation of the E method is far smaller than that of the F method, as shown in Fig. 8.

Furthermore, the visual reconstruction accuracy of the E method is also intuitively higher than that of the F method. One can see from Fig. 9 that the E method evidently performs better on both of the two metrics, indicating that the event-based approach exceeds the frame-based method under the same data set size and network structure.

We statistically analyze the reconstruction accuracy indexes of 10 digits in both the MNIST test set and the PRINT test set at different positions. The average LPIPS and PSNR of each

reconstructed frame for different test digits with the E method and the F method are shown in Table 1, respectively. One can see from Table 1 that the reconstruction LPIPS obtained by the E method is 38% and 11% lower than by the F method on the two test sets, respectively, which demonstrate the higher quality
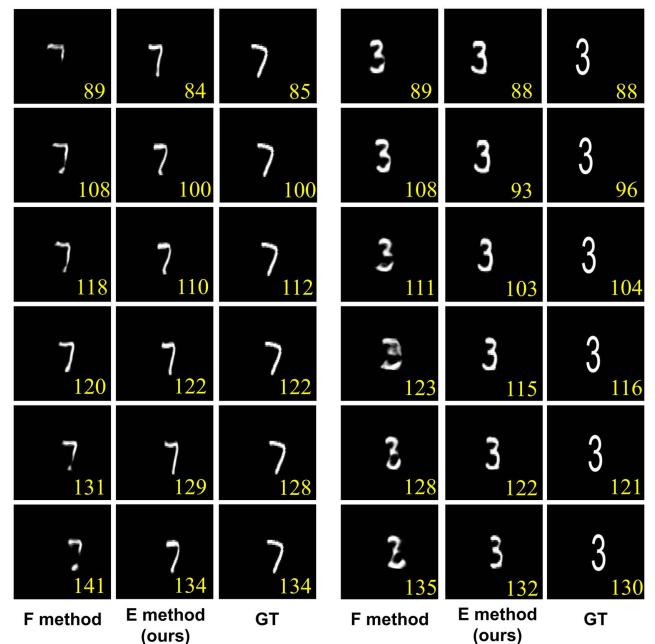

Fig. 7. Reconstructions of NLOS moving target at different positions through the E method and the F method. Six different positions of digit 7 (MNIST test set) and digit 3 (PRINT test set) are displayed as examples. The Cd value (pixel) is labeled at the corner of each frame.
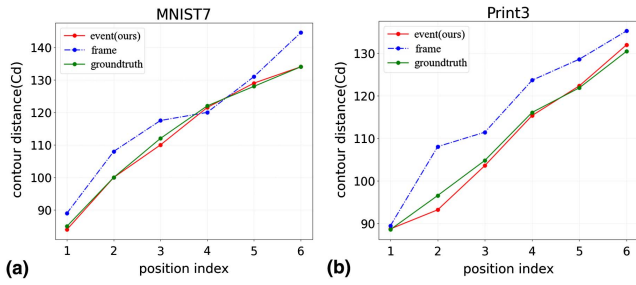
Fig. 8. Cd value of NLOS reconstructions at different positions. (a), (b) are the Cd values of reconstructions shown in Fig. 7 (digit 7 and digit 3, respectively).
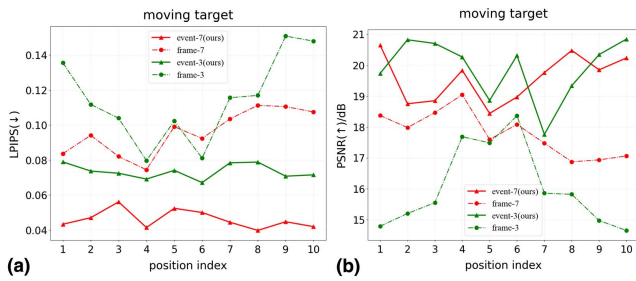

Fig. 9. Evaluation metrics LPIPS and PSNR for reconstructions of digit 7 (MNIST test) and digit 3 (PRINT test) at 10 different positions, respectively. The full line denotes the E method, while the dotted line denotes frame-based ones.

of human perception. As for the PSNR, the E method scores higher than the F method on every test target and performs about 10% better numerically.

As for the generalization, see Section 3 of the Supplementary Material for discussions about target movement and Visualization 2 for the possibility of applying event-based NLOS imaging in real-world circumstances.

5. Conclusion

In summary, we leverage the sampling specialty of an event camera and propose a new detection and reconstruction method for passive NLOS imaging. The E method extracts rich dynamic information from the diffusion spot movements and provides a physical foundation for passive NLOS imaging of moving targets. Compared with the deep-learning approach with a traditional camera, the event-based framework shows better performance when reconstructing NLOS moving targets. We carry out experiments on two types of targets with different distribution forms and verify that the reconstruction quality is significantly improved with the framework we addressed in both visual accuracy and position accuracy. The reconstruction quality on the PRINT test set indicates that our method has extracted more movement information on moving targets with the event detection paradigm compared with traditional frame-based detection. We believe that the event approach for the inverse problem, together with the EM-NLOS data set, is a big step and can inspire new ideas toward the development of feature-embedded passive NLOS imaging with multidetector information fusion[12] and

Table 1. Evaluation Metrics of Results with E Method and F Method[a].

	PSNR/dB (↑)				LPIPS (↓)			
	MNIST test		PRINT test		MNIST test		PRINT test	
Digit	E (ours)	F	E (ours)	F	E (ours)	F	E (ours)	F
0	17.98	17.02	18.25	16.67	**0.063**	0.088	**0.068**	0.084
1	**29.24**	26.57	**21.97**	18.66	**0.014**	0.036	**0.071**	0.083
2	19.09	19.04	17.29	15.96	**0.065**	0.092	0.085	0.087
3	**19.01**	16.48	**19.83**	16.42	**0.063**	0.129	**0.075**	0.099
4	19.33	18.13	**17.58**	15.36	0.067	0.076	0.101	0.114
5	20.84	19.56	17.42	15.96	**0.056**	0.075	0.082	0.090
6	20.48	19.64	**16.41**	14.63	**0.049**	0.070	0.101	0.106
7	**20.50**	17.22	17.78	17.26	**0.042**	0.106	0.081	0.086
8	19.66	18.47	**17.23**	15.04	**0.047**	0.086	0.088	0.095
9	**19.53**	16.86	17.35	17.17	**0.057**	0.085	0.084	0.086
Avg.	20.57	18.90	18.11	16.32	0.052	0.084	0.083	0.093

[a]The average LPIPS and PSNR of each reconstructed frame for 10 digits are compared. An improvement above 10% in PSNR and above 15% in LPIPS is in bold.

NLOS target tracking[28]. The event cues we demonstrated fuse the event paradigm information of NLOS moving target with end-to-end data-driven methods for solving the event-based inverse problem. In future work, we will take target movements and environment disturbance into consideration, and continually put forward the applications of event-based cues in practice by providing enhancement for methods based on other dimensions of the light field. The event-based vision utilized in this work has great potential to facilitate further research on feature-embedded passive NLOS and its applications.

Acknowledgement

†These authors contributed equally to this work.

References

1. D. Faccio, A. Velten, and G. Wetzstein, "Non-line-of-sight imaging," Nat. Rev. Phys. **2**, 318 (2020).
2. A. Kirmani, T. Hutchison, J. Davis, and R. Raskar, "Looking around the corner using transient imaging," in *IEEE 12th International Conference on Computer Vision (ICCV)* (2009), p. 159.
3. T. Maeda, G. Satat, T. Swedish, L. Sinha, and R. Raskar, "Recent advances in imaging around corners," arXiv:1910.05613 (2019).
4. M. La Manna, F. Kine, E. Breitbach, J. Jackson, T. Sultan, and A. Velten, "Error backprojection algorithms for non-line-of-sight imaging," IEEE Trans. Pattern Anal. Mach. Intell. **41**, 1615 (2019).
5. M. O'Toole, D. B. Lindell, and G. Wetzstein, "Confocal non-line-of-sight imaging based on the light-cone transform," Nature **555**, 338 (2018).
6. T. Sasaki, C. Hashemi, and J. R. Leger, "Passive 3D location estimation of non-line-of-sight objects from a scattered thermal infrared light field," Opt. Express **29**, 43642 (2021).
7. J. Boger-Lombard and O. Katz, "Passive optical time-of-flight for non-line-of-sight localization," Nat. Commun. **10**, 3343 (2019).
8. C. Pei, A. Zhang, Y. Deng, F. Xu, J. Wu, D. U.-L. Li, H. Qiao, L. Fang, and Q. Dai, "Dynamic non-line-of-sight imaging system based on the optimization of point spread functions," Opt. Express **29**, 32349 (2021).
9. A. Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, M. G. Bawendi, and R. Raskar, "Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging," Nat. Commun. **3**, 745 (2012).
10. X. Liu, I. Guillén, M. La Manna, J. H. Nam, S. A. Reza, T. H. Le, A. Jarabo, D. Gutierrez, and A. Velten, "Non-line-of-sight imaging using phasor-field virtual wave optics," Nature **572**, 620 (2019).
11. R. Geng, Y. Hu, Z. Lu, C. Yu, H. Li, H. Zhang, and Y. Chen, "Passive non-line-of-sight imaging using optimal transport," IEEE Trans. Image Process. **31**, 110 (2021).
12. K. Tanaka, Y. Mukaigawa, and A. Kadambi, "Polarized non-line-of-sight imaging," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020), p. 2136.
13. O. Katz, P. Heidmann, M. Fink, and S. Gigan, "Non-invasive singleshot imaging through scattering layers and around corners via speckle correlations," Nat. Photonics **8**, 784 (2014).
14. M. Tancik, T. Swedish, G. Satat, and R. Raskar, "Data-driven non-line-of-sight imaging with a traditional camera," in *Imaging and Applied Optics* (2018), paper IW2B.6.
15. J. He, S. Wu, R. Wei, and Y. Zhang, "Non-line-of-sight imaging and tracking of moving objects based on deep learning," Opt. Express **30**, 16758 (2022).
16. 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**, 63 (2020).
17. C. Zhou, C.-Y. Wang, and Z. Liu, "Non-line-of-sight imaging off a phong surface through deep learning," arXiv:2005.00007 (2020).
18. A. Zhang, J. Wu, J. Suo, L. Fang, H. Qiao, D. D.-U. Li, S. Zhang, J. Fan, D. Qi, Q. Dai, and C. Pei, "Single-shot compressed ultrafast photography based on U-net network," Opt. Express **28**, 39299 (2020).
19. G. Gallego, T. Delbruck, G. M. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. J. Davison, J. Conradt, K. Daniilidis, and D. Scaramuzza, "Event-based vision: a survey," IEEE Trans. Pattern Anal. Mach. Intell. **44**, 154 (2020).
20. A. Amir, B. Taba, D. Berg, T. Melano, J. McKinstry, C. Di Nolfo, T. Nayak, A. Andreopoulos, G. Garreau, M. Mendoza, J. Kusnitz, M. Debole, S. Esser, T. Delbruck, M. Flickner, and D. Modha, "A low power, fully event-based gesture recognition system," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), p. 7388.
21. H. Rebecq, R. Ranftl, V. Koltun, and D. Scaramuzza, "High speed and high dynamic range video with an event camera," IEEE Trans. Pattern Anal. Mach. Intell. **43**, 1964 (2021).
22. S. Schaefer, D. Gehrig, and D. Scaramuzza, "AEGNN: asynchronous event-based graph neural networks," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2022), p. 12361.
23. C. Saunders, J. Murray-Bruce, and V. K. Goyal, "Computational periscopy with an ordinary digital camera," Nature **565**, 472 (2019).
24. X. Lagorce, G. Orchard, F. Galluppi, B. E. Shi, and R. B. Benosman, "HOTS: a hierarchy of event-based time-surfaces for pattern recognition," IEEE Trans. Pattern Anal. Mach. Intell. **39**, 1346 (2017).
25. C. Yan, X. Wang, X. Zhang, and X. Li, "Adaptive event address map denoising for event cameras," IEEE Sens. J. **22**, 3417 (2022).
26. C. Wang, X. Wang, C. Yan, and K. Ma, "Feature representation and compression methods for event-based data," IEEE Sens. J. **23**, 5109 (2023).
27. R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2018), p. 586.
28. S. Chan, R. E. Warburton, G. Gariepy, J. Leach, and D. Faccio, "Non-line-of-sight tracking of people at long range," Opt. Express **25**, 10109 (2017).