Supplemental Material

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

†These authors contributed equally to this work.

1. Detailed Parameters of CeleX-V Event Camera

The camera information for the event camera we used in the experiments:

Table 1. Detailed parameters of CeleX5-MIPI camera

Notation	Value and explanation
Camera	CeleX5-MIPI
Sensor	CeleX-5 Sensor Core
Drivers	CX3 USB3.0 driver (Windows / Linux)
Intensity grading	0-4095
Image size	5/8''
Module effective Pixels	1280×800
F Mood sampling frequency	100Hz

The CeleX5-MIPI event camera we used in the experiments is a multifunctional smart image sensor with 1Mega-pixels. The output of the sensor is not a frame, but a stream of asynchronous digital events. The speed of the sensor is not limited by any traditional concept such as exposure time and frame rate [1]. Event Intensity(EI) Mood is the mood for event-based data collecting, which we used to establish EM-NLOS data set, and Full-Picture(F) Mode is the mood for corresponding frame-based data collecting, which we used to build a calibrated counterpart FM-NLOS data set.
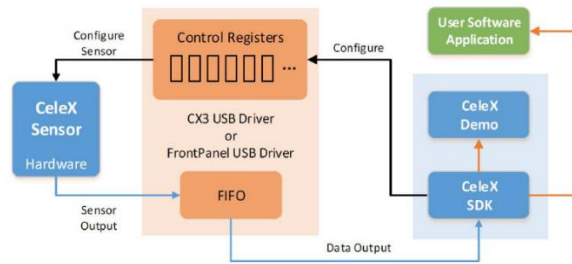

Fig. 1.   The working principle of CeleX-5 Chipset [1].

2. Structure of the Proposed R-UNet

The framework of Residual UNet (R-UNet) in our proposed event-based method is shown in Fig. 2. (a), and the layers of Residual Block is shown in Fig. 2. (b).
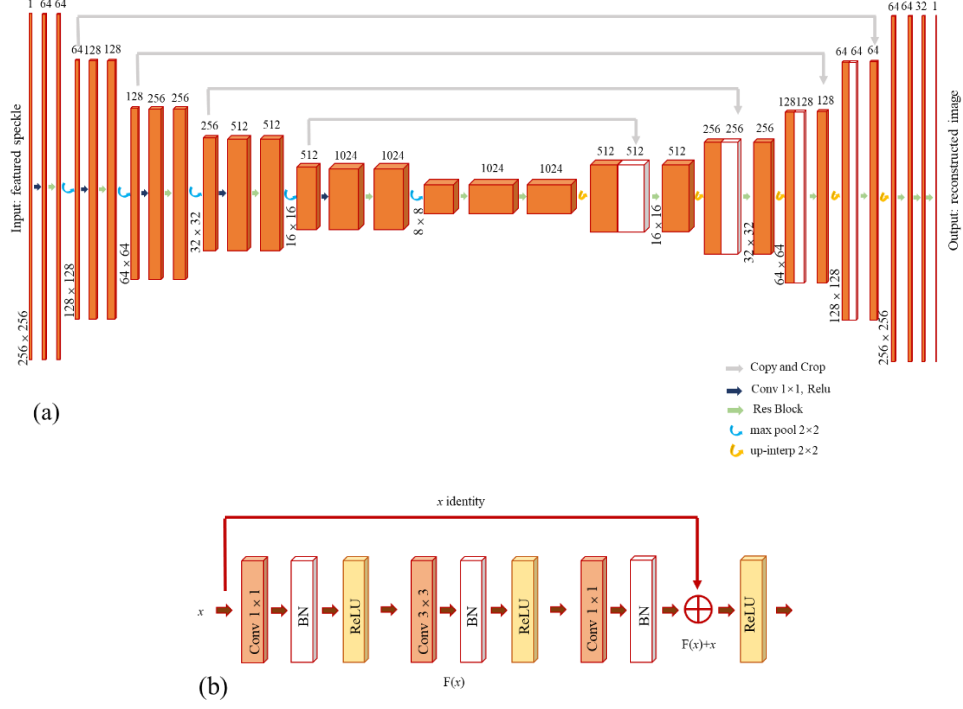

Fig. 2.   The structure of R-UNet and the Residual Block we used in event-based approach. (a) is the R-Unet framework, (b) is the layers in Residual Block.

We select UNet structure as our baseline to fit with our limited data set EM-NLOS and FM-NLOS data sets, since the data collection and data set making procedural is time consuming. Due to the well performance of UNet on small dataset, and our aim is to verified the advantage of event-based approach compared with frame-based approach in reconstructing NLOS moving targets, the size and diversity of our data set are not required to be so large. We just need to ensure the equivalent quota of EM-NLOS and FM-NLOS for fair comparison. We introduce the residual block with bottleneck layer to fuse the multi-scale information fully and avoid the gradient explosion when training.

The composition of our EM-NLOS and FM-NLOS data set are listed as shown in Table 2 and Table 3.

Table 2. Composition of NLOS-ES data set

Class	Notation	Group	Amount
Train	MNIST train set 0-9	13groups	3950
Validation	MNIST train set 0-9	13groups	130
Test	MNIST test set 0-9	1 group	100
	PRINT set 0-9	1 group	110

Table 3. Composition of NLOS-FS data set

Class	Notation	Group	Amount
Train	MNIST train set 0-9	13groups	3870
Validation	MNIST train set 0-9	13groups	130
Test	MNIST test set 0-9	1 group	100
	PRINT set 0-9	1 group	90

3. Application of the Proposed Event-Based NLOS Framework

To demonstrate the application of the proposed event-based NLOS framework, we perform several experiments with the same settings on MNIST and fashion MNIST. The feature representation difference under different target movements based on the proposed framework is discussed and the stability of the proposed method is verified under ambient environment interference.

3.1 Generalization on more complex targets

We perform a set of experiments on fashion MNIST data set with the same experimental settings of experiments on MNIST data set. We acquire the event-based data of 500 different targets in fashion MNIST and generate a data set (Event fashion MNIST NLOS, EfM-NLOS) with featured diffusion spot calculated by time-surface representation at 20 different positions (10000 images). We assign the featured spot (featured TS map) of 490 targets at 20 different positions (9800 images) as training set, and the rest as test set. The reconstructions on test set are shown in Fig. 3., where the approximate shape of the targets are restored correctly.

Featured TS map   Reconstruction   Ground Truth   Featured TS map   Reconstruction   Ground Truth
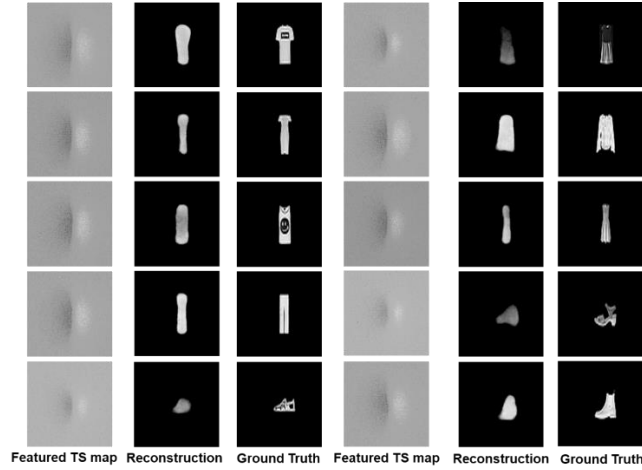
Fig. 3. The reconstruction results on fashion MNIST test set.

We trained our Residual-UNet (R-UNet) on EfM-NLOS train set by Adaptive Moment estimation (Adam) optimizer with Nvidia RTX 3090 GPU for 600 epochs. The batch size is set to 32, and the start learning rate is assigned as 0.00001. We adjust the learning rate every 50 epochs by multiplying a decay factor of 0.9 for better optimization.

The results indicate that our framework has certain generalization on more complex targets and

avoid overfitting to some extent.

Although the performance is not such ideal as that on MNIST test set, it could be a possible settlement for future researches in fusion of event-based cue with traditional passive NLOS imaging methods.

3.2 Feature representation difference under different target movements

Since event camera responses to brightness changes in the scene, and depicts the type of variation by polarity flag (1 for brightness increase and 0 for brightness decrease), it is sensitive to the direction of target movement. We plot the address map of event scatters and featured diffusion spot after time-surface representation of raw event data. As shown in Fig. 4., it is obvious that the polarity flag indicates the moving direction, which could not be expressed by frame-based data. As a result, we could distinguish the movement direction by extracting the distribution of polarity flag in the proposed event-based method. We flip the featured spot horizontally for a better exhibition, since the diffusion spot moves mirroring to the target movement.
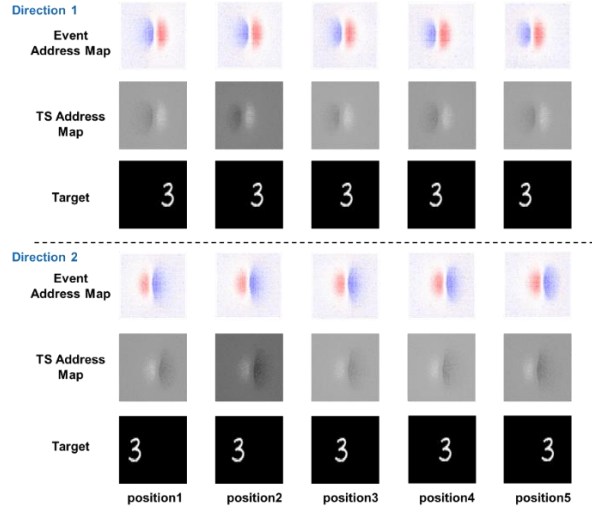

Fig. 4. The feature difference under different moving direction of the target, direction 1 donates the target movement from right to left, and direction 2 shows its movement from left to right.

The difference of featured diffusion spot between different target speeds are not notable because we perform time-surface calculation to raw event data and extract the spatio-temporal information instead of a simple mapping of spatial intensity.

3.3 Performance under ambient environment interference

To verify the ability of observing the dynamic diffusion spot in NLOS scenarios with ambient interference, we conduct two additional experiments and perform statistical analysis to visualize the superiority of event camera in capturing slight changes of brightness.

As for steady ambient light environment, we conduct an experiment under uniform daylight lamp illumination with the power of 50Hz alternating current electricity to form a steady ambient light environment, as shown in Fig. 5.
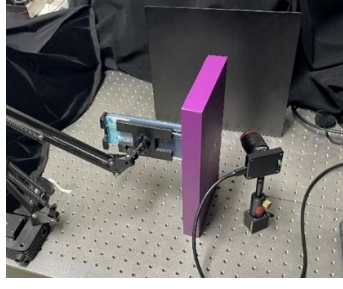
Fig. 5. Experiment under steady ambient light environment.

To demo our experiment more explicitly, we visualize the data acquisition process in supplementary materials (Visualization 2.mp4). It is obvious that the signal of dynamic diffusion spot is observed more clearly by event-based mode.

We conduct another experiment under additional illumination of a strong light flashlight, which could be regard as harsh ambient light interference environment, as shown in Fig. 6.
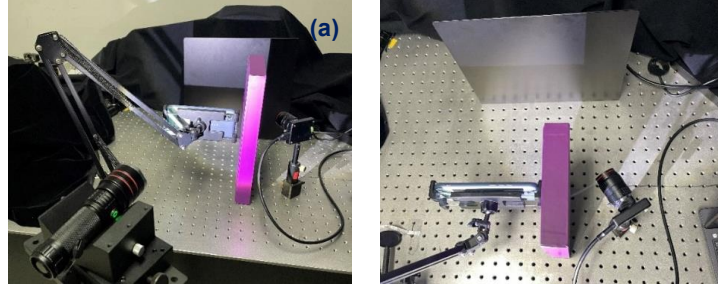

Fig. 6. Experiment under harsh ambient light interference environment, (a) is the front view, (b) is the vertical view.

The dynamic diffusion spot is not easily observed under such strong artificial disturbance, as shown in supplementary materials (Visualization 2.mp4), but the slight change in brightness is detected by event camera to some extent.

In addition, we perform statistical analysis to visualize the superiority of event camera in capturing slight changes of brightness. We calculated the distribution of event-based data acquired under steady ambient light environment in temporal and spatial dimension, as shown in Fig. 7 and Fig. 8.
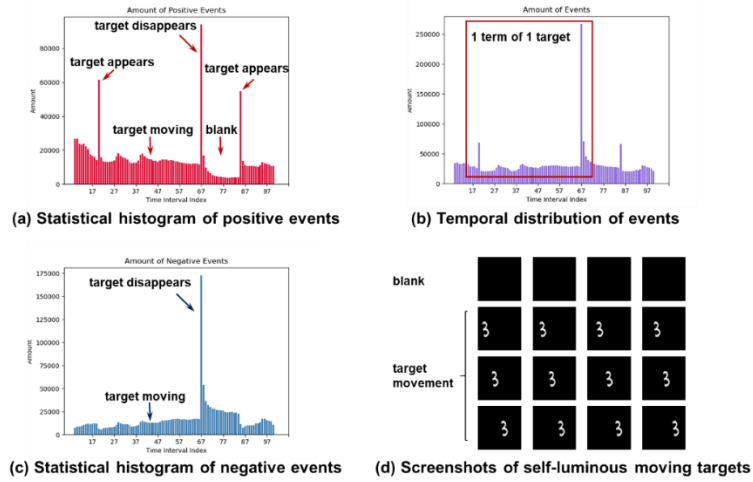

Fig. 7. Statistical histogram of events in temporal dimension.

For the convenience of data acquisition, we produce formatted videos as the self-luminous targets. As shown in Fig. 7(d), each video begins with 2 seconds of blank before the target appearance, then

6 seconds of horizontal movement before the target twinkles and disappear within 10 milliseconds. It is obvious to see from Fig. 7(a)-7(c) that even under steady ambient light environment, the appearance and disappearance is reflected clearly by the temporal distribution of event stream.
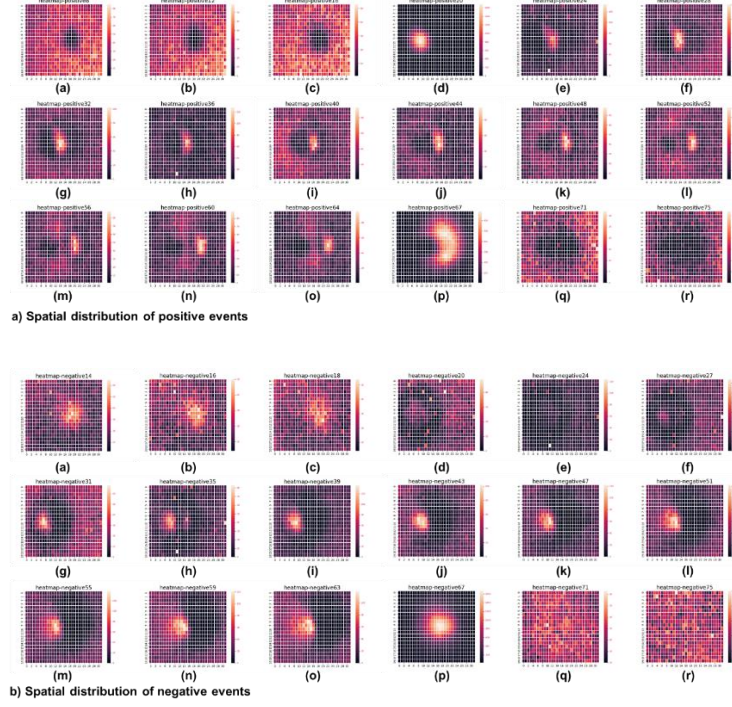

Fig. 8. Heatmap of events of each interval in spatial dimension.

As shown in Fig. 8, we calculate the spatial distribution of the passive and negative events in each time interval respectively and plot the heatmap for more explicit demo. It could be seen that the distribution translates with the diffusion spot formed by the target video, which manifest the sensitivity of event camera in detecting dynamic diffusion spot under ambient light interference.

Furthermore, we also test the data acquired under steady ambient light interference with the pre-trained model, as shown in Fig. 9. The approximate shape of the pre-trained targets could be reconstructed, which verifies certain generalization under interference.
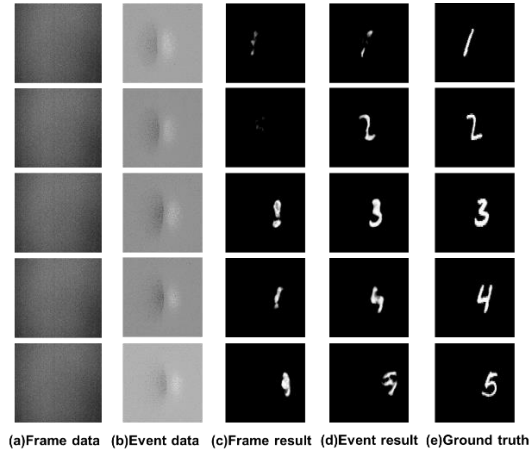

Fig. 9. Reconstructions of pre-trained targets under untrained ambient conditions, (a) frame data, (b) event TS map, (c) result of frame-based method, (d) result of event-based method, (e) ground truth.

Overall, the event-based framework in NLOS imaging provides a new approach for dynamic diffusion spot detection under weak Signal-to-Noise Ratio (SNR) and harsh diffuse reflection. By

exploiting the potentialities of bio-inspired sensors, it will provide a gateway for passive NLOS imaging step by step.

References

1. CelePixel CelePixel Technology Co. Ltd, https://github.com/CelePixel/CeleX5-MIPI.