Supplementary Document:
Time-Efficient Light-Field Acquisition Using Coded Aperture and Events

Shuji Habuchi† Keita Takahashi† Chihiro Tsutake† Toshiaki Fujii† Hajime Nagahara‡
† Nagoya University, Japan ‡ Osaka University, Japan

A. Learned Aperture Patterns

We present several sets of aperture patterns leaned with different configurations (flexible-τ, event only†, event only) in Fig. S.1. The event only case without the second loss term resulted in the patterns with significant brightness differences, which were incompatible with real event cameras.

To clarify the advantage of the learned aperture patterns, we also tested a set of fixed coding patterns (fixed codes) shown in Fig. S.1, each pattern of which had exactly 50% brightness. The network pipeline was retrained/tested with this set of patterns. See "ours (flexible-τ, fixed codes)" in Table S.1 for the result, which is worse than that obtained with the learned patterns ("ours (flexible-τ)").

B. CA with Short Exposures

In the main text, we assumed that the exposure was constant for all the methods; Therefore, CA ($N = 4$) took four times the total exposure time of our method. Here, we also considered a case where CA ($N = 4$) took the same total exposure time as our method. In this case, the exposure for each image was reduced to 1/4 $\times$ the original exposure, which led to noisier images. Assuming that the relative noise level for an image (σ) was inversely proportional to $\sqrt{\text{exposure time}}$[1], we increased the noise level from $\sigma = 0.005$ to $\sigma = 0.01$; the network pipeline was retrained/tested with $\sigma = 0.01$. See "CA ($N = 4$, short exposure)" in Table S.1 for the result; the PSNR/SSIM scores for this case are located between those of "ours (fixed-τ-low)" and "ours (flexible-τ)".

C. Details around Eq. (12)

The left hand-side of Eq. (12) is an event stack (sum of events at each pixel over a transient time). Therefore, it can take $0, \pm1, \pm2, \pm3, \ldots$. Function Q is designed to quantize the input in the forward process but let the gradient pass through as it is in the backward process;

[1]This was simply derived from the statistical property of Gaussian noises. The noise model for a camera is more complicated in reality, but it was simplified as an additive Gaussian throughout our experiments.

Table S.1. Additional results. Rows with "*" are cited from Tables 1 and 2 for reference. Scores are PSNR/SSIM over all test data.

Method	Test τ	ALL
* CA ($N = 4$)	–	35.39/0.9346
CA ($N = 4$, short exposure)	–	34.05/0.9154
* Ours (fixed-τ-low)	0.075	34.45/0.9219
* Ours (flexible-τ)	0.15	33.79/0.9149
Ours (flexible-τ, fixed codes)	0.15	33.14/0.9004

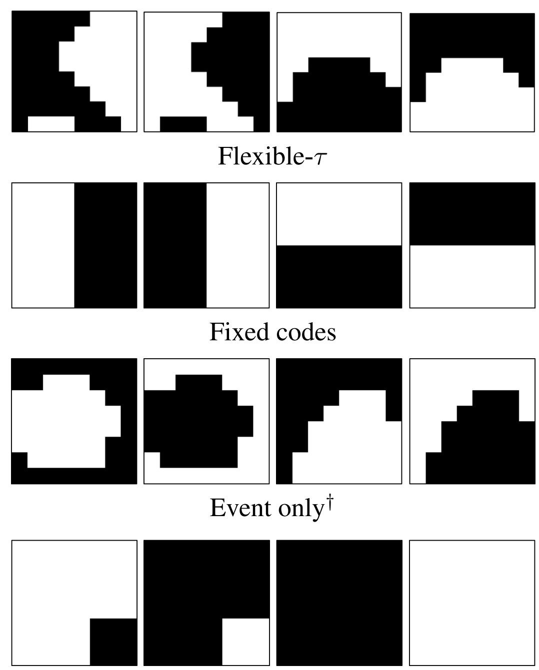

Figure S.1. Learned and fixed aperture coding patterns ($a^{(1)}$ to $a^{(4)}$ from left to right). † indicates that model was trained with second loss term.

```
forward (input):
    return sign(input).floor(abs(input))
backward (grad):
    return grad
```