

Coded-E2LF: Coded Aperture Light Field Imaging from Events

Supplementary Material

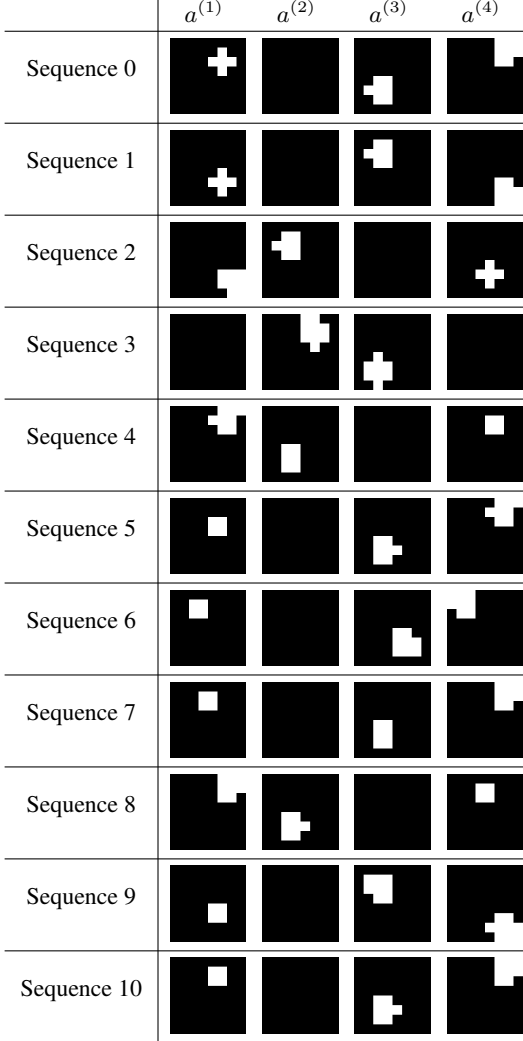


Figure 7. Coding patterns learned with baseline algorithm.

6. Additional Results and Discussion

6.1. Coding patterns

In Fig. 7, we present examples of coding patterns learned with our baseline algorithm. Note that at least one black pattern was included in each coding-pattern sequence.

The coding patterns have limited light efficiency, as the number of white elements is less than ten. This phenomenon results from a trade-off between the signal-to-noise ratio (SNR) and defocus blur; higher brightness leads to better SNR but increases blurring.

Table 2. Summary of ablation study with $N = 4$. All reported scores are averaged in BasicLFSR test dataset.

Method	Total event count	PSNR/SSIM
Baseline	9.492	29.81/0.8302
Baseline+RA	10.785	29.87/0.8309
Baseline+BF	6.522	29.69/0.8321
Baseline+BF+RA	7.175	30.48/0.8413

Method	$a^{(1)}$	$a^{(2)}$	$a^{(3)}$	$a^{(4)}$
(total events)	$ E^{(1,2)} $	$ E^{(2,3)} $	$ E^{(3,4)} $	
Baseline+RA				
(10.785)	6.413	1.704	2.668	

Figure 8. Learned coding patterns ($N = 4$) with event counts in $E^{(1,2)}$, $E^{(2,3)}$, and $E^{(3,4)}$, and their total. Event counts are per pixel, averaged in BasicLFSR test dataset.

6.2. Additional ablation study

In the main paper, we report three variants of our method (Baseline, Baseline+BF, Baseline+BF+RA). Here, we additionally report the result with Baseline+RA, where we initialize $I_{x,y}^{\text{ref}} = I_{x,y}^{(1)}$ for all (x,y) s at $n = 1$ and use Eqs. (9) and (10) for the event generation and reference update processes. The learned coding patterns are presented in Fig. 8. As shown in Table 2, Baseline and Baseline+RA were almost equivalent in terms of the reconstruction quality. The best reconstruction quality was achieved with Baseline+BF+RA, which is the full implementation of our method.

The difference in the total event counts between Baseline+BF (6.522) and Baseline+BF+RA (7.175) is mainly attributed to the difference in the event generation models. Baseline+BF was trained and evaluated using the baseline event generation model (Eq. (8)), which underestimated the number of events. Meanwhile, Baseline+BF+RA was trained and evaluated using the reference-aware event generation model (Eqs. (9) and (10)). If the coding patterns of Baseline+BF was evaluated using Eqs. (9) and (10), the total event count increased to 7.273 events/pixel.

6.3. Comparison with intensity-based methods

For simulation experiments, we strictly followed the protocols of Habuchi et al. [14]; thus, our results can be directly compared with those in their paper [14]. Some results with intensity-based methods are summarized in Table 3 to show

Table 3. Comparison with intensity-based methods. CA: coded aperture, LA: lens-array, @I: number of intensity images, @E: number of event images. * indicates quotation from Habuchi et al.’s [14]. All six methods share the same RecNet architecture as Habuchi et al.’s [14].

Methods	@I	@E	PSNR[dB]/SSIM (ALL)
CA + RecNet*	4	–	35.39/0.9346
CA + RecNet*	2	–	34.09/0.9210
CA + RecNet*	1	–	27.62/0.8139
LA + RecNet*	1	–	26.22/0.7218
Habuchi [14]	1	3	31.83/0.8729
Ours	–	3	30.48/0.8413

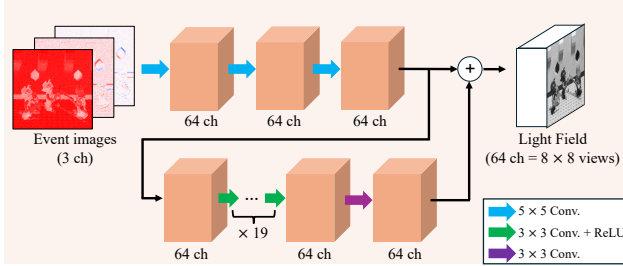


Figure 9. Network architecture of RecNet

the position of our method. Intensity-based coded aperture imaging (CA + RecNet) offers a clear advantage in terms of the reconstruction fidelity. However, our method has the potential to reduce the measurement time for light fields even in low light conditions.

6.4. Data rate analysis

Unfortunately, it is difficult to claim the advantage of our method in terms of data rate or bandwidth compared to its intensity-based counterpart. Since the event count depends on the brightness/complexity of the target scene, we discuss the average on the test dataset. Our method captures 7.175 events/pixel at the sensor, which corresponds to 0.112 events/pixel with respect to all the pixels of a light field (8×8 views). If each event requires 29 bits in the COO format, the data rate is 208 bits/pixel at the sensor and 3.25 bits/pixel with respect to light field pixels. This is more efficient than direct acquisition of 8×8 views, but less efficient than the intensity-based coded aperture method (e.g., 8 bits/pixel \times 2 – 4 frames).

7. Network architecture of RecNet

For RecNet, we used the same architecture as Habuchi et al. [14], which is illustrated in Fig. 9.