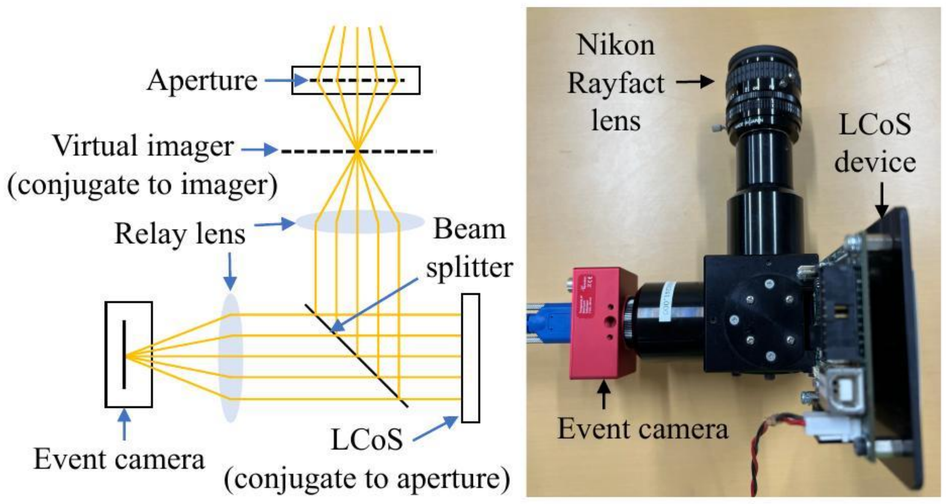

Figure 3. Hardware setup of our prototype camera.

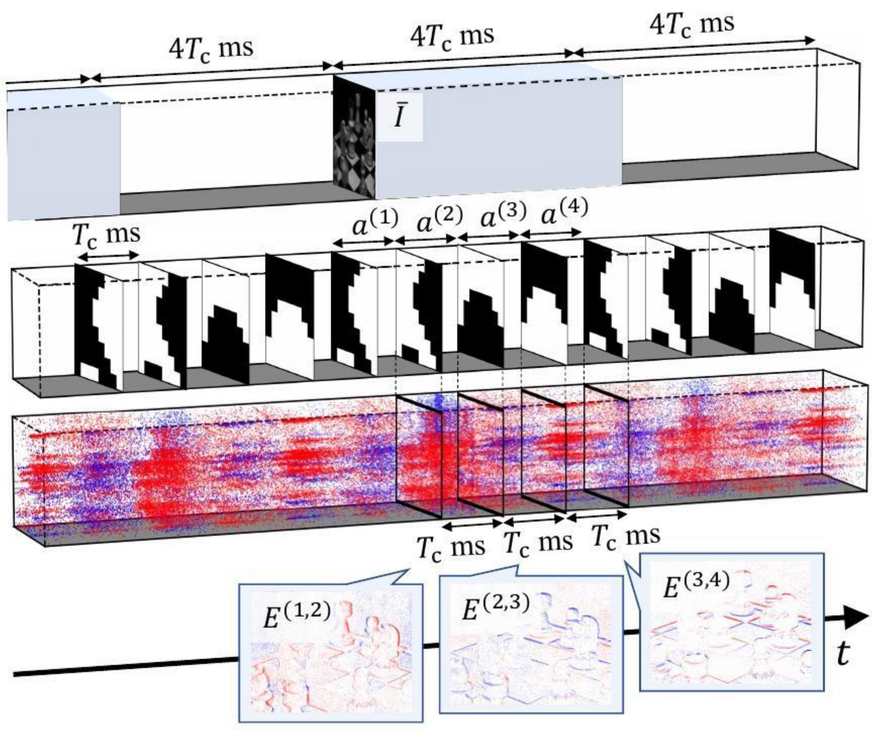

Figure 4. Timing chart. From top to bottom, exposure for image frames, aperture-coding patterns, and event stream are shown.

single cycle of the four coding patterns; See Fig. 4 for a timing chart. Although the aperture coding is not electrically synchronized with the event camera, we can identify the event stacks in the event stream because the cyclic bursts of events correspond to the changes of the coding patterns.

4. Experiments

4.1. Quantitative Evaluation

The evaluation was conducted on the test data of the BasicLFSR dataset [23], which includes 23 light fields categorized into five groups, all of which were reserved for evaluation. For each light field, we computationally carried out image/event acquisition (AcqNet) and reconstruction (RecNet) processes, and quantitatively evaluated the reconstruction accuracy. Tables 1 and 2 summarize the quantitative scores (PSNR and SSIM, where larger is the better for both) averaged for each group and over all the groups.

As shown in Table 1, we configured four models of our method that are different with respect to the contrast threshold τ. During training, τ was fixed to 0.075/0.15/0.3 for the fixed-τ-low/mid/high models, respectively, while it was randomly (uniformly) drawn from $[0.075, 0.3]$ for the

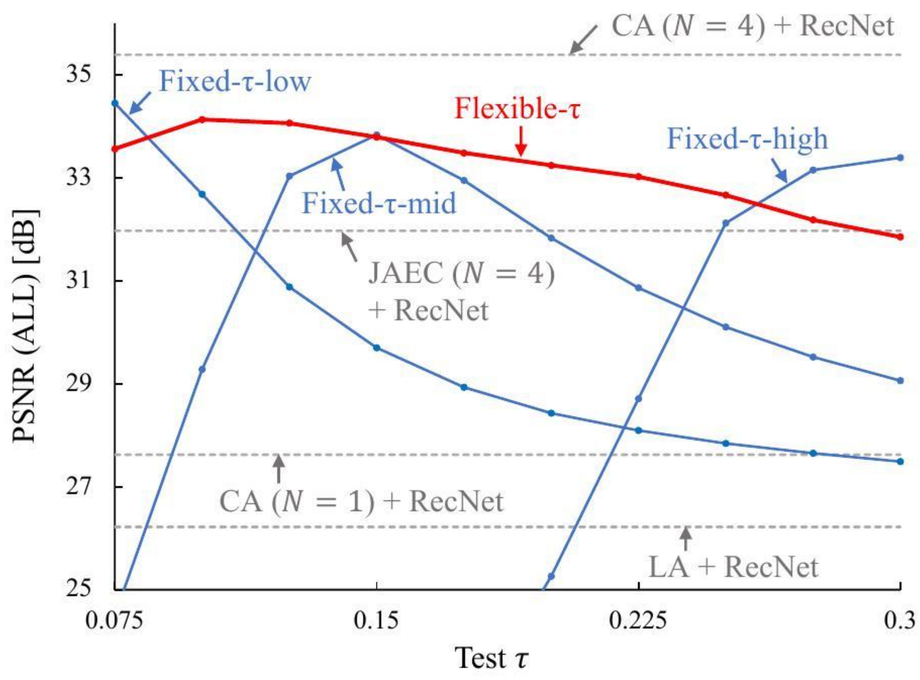

Figure 5. Reconstruction quality of our method (flexible-τ, fixed-τ-low/mid/high) against test-time τ.

flexible-τ model. At test time, τ was set to 0.075/0.15/0.3. When a fixed-τ model was trained and tested with an identical τ, a smaller τ resulted in better reconstruction scores. This is reasonable because with a smaller τ, a larger number of events were observed; thus, more abundant information was obtained from the target scenes. The flexible-τ model resulted in slightly inferior accuracy to the fixed-τ models tested with the corresponding τs. However, the flexible-τ model was adaptive to various τs. To make this point clearer, Fig. 5 shows the reconstruction quality of each model (the average PSNR over all the groups) against a wide range of the test-time τ. The fixed-τ models were sensitive to the test time τ; each of them had a narrow peak around the τ with which the model was trained. In contrast, the flexible-τ model maintained fine reconstruction quality over a wider range of τ, which is important to ensure the compatibility with real event cameras.

Table 1 also includes ablation models with which the measurement was limited to either the image frame (image-only) or the event stacks (event-only); for each ablation model, the same algorithm pipeline as ours was trained from scratch. The second term of the loss function (Eq. (13), with $\lambda = 0.00001$ and $\theta = 131{,}130$) was enabled only for the event-only model, which is indicated as "†". Without this term, the learned coding patterns of the event-only model had significant brightness differences and caused too many events. The poor reconstruction quality obtained with the ablation models indicates that both the image frame and events are necessary for accurate light-field reconstruction.

As summarized in Table 2, we also tested other imaging methods for comparison. As the baseline of our method, **coded-aperture imaging (CA)** [10, 14, 22, 30, 38] takes several (N) images following Eq. (1), which requires a longer measurement time than a single exposure. **Joint aperture-exposure coding (JAEC)** [28, 41, 42, 46] captures a single coded image following Eq. (2) with $N = 4$.