CodedEvents: Optimal Point-Spread-Function Engineering for 3D-Tracking with Event Cameras

Supplementary Material

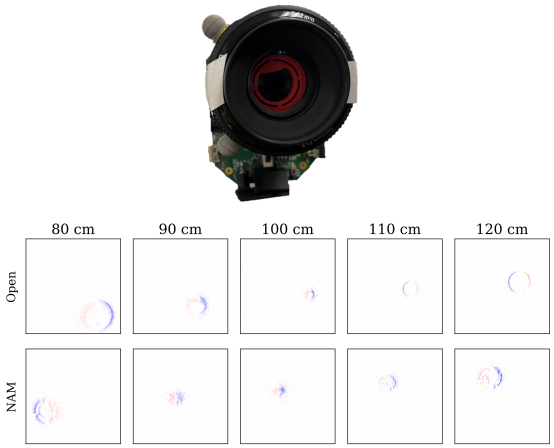

Figure 1. **Prototype.** Top: The fabricated mask is placed at the aperture plane of an event camera with a 50mm focal length lens. Bottom: Sample captured event frames for a point source.

S1. Hardware Prototype

We performed a real-world experiment for tracking a point light source at meter scale using a binary amplitude mask and a Prophessee EVK3 event camera. Specifically, we fabricated the NAM mask at 20mm diameter scale on a Creality Ender 3 S1 Pro using 1.75mm PLA filament (see Figure 1). Then, we captured an event dataset by moving a point source at discrete depth planes ranging between 75cm and 125cm with and without our coded aperture. For all measurements, the camera was focused at 100cm. We binned events in 1ms intervals to achieve an effective frame rate of 1000 FPS and trained a CNN to estimate the event frame's depth. Results in Figure 2 demonstrate improved tracking performance compared to an open aperture, particularly at depths where the point source is defocused.

S2. Accumulation Time

Cutting-edge event cameras offer 10kHz fresh rates; even with 16-frame accumulation, the camera effectively operates at 625FPS — much faster than conventional CMOS sensors. We also retrained our CNN-based tracking algorithm on 'pure' event frames with no accumulation. Overall performance degraded: NPM by $+45\%$ RMSE and NAM by $+54\%$ RMSE. Alternative architectures such as Spiking Neural Networks designed for sparse binary measurements may be better suited for processing 'pure' events.

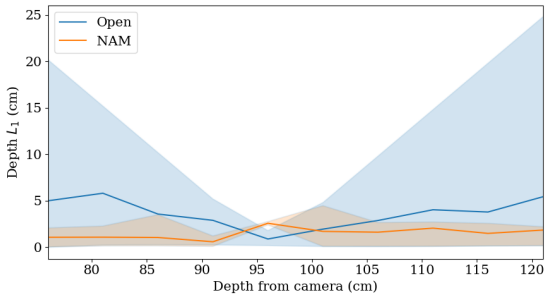

Figure 2. **Real-world 3D tracking.** Comparison between NAM and Open apertures for depth estimation at 1000FPS. Error bars show the 90% interquartile range.

S3. The Effects of Particle Speed

We have shown CRB depends on particle speed; a natural question is does the optimal design change with respect to speed. We optimize our neural phase mask using the CRB objective function with fixed particle speeds—$\{50, 100, 500, 1000\}$nm per time step. Our learned designs are shown in Figure 3. When a particle moves quickly relative to the binned interval, the optimal design resembles the Fisher phase pattern found for traditional CMOS sensors.

One can explain this collapse to the original Fisher mask design as follows. As a particle moves faster, the captured binned event frame looks more similar to the composition of a negative PSF at the start location and a positive PSF at the end location (Figure 4). This suggests that single-point event tracking mirrors two-point CMOS tracking.

S4. Log-Intensity Difference Approximation

In this section, we prove the log-intensity difference approximation we consider when deriving the Cramér Rao Bound is proportional to binned event frames.

Assume an idealized event camera model, where an event is triggered as soon as the log-intensity change between the reference and the current intensity equals some threshold, $\mathcal{T}$. Consider producing a binned event frame for a time interval $[t_{\text{start}}, t_{\text{end}}]$. For a single pixel, let the sequence of events over this interval occur at times $t_1, t_2, \dots t_n$ and have polarities $p_1, p_2, \dots, p_n \in \{-1, 1\}$. Let $f(t)$ be the log-intensity at time t for the same pixel and be continuous over the interval.

Lemma S4.1. *The log-intensity difference, $f(t_{end}) - f(t_{start})$, is proportional to the binned event pixel value,*

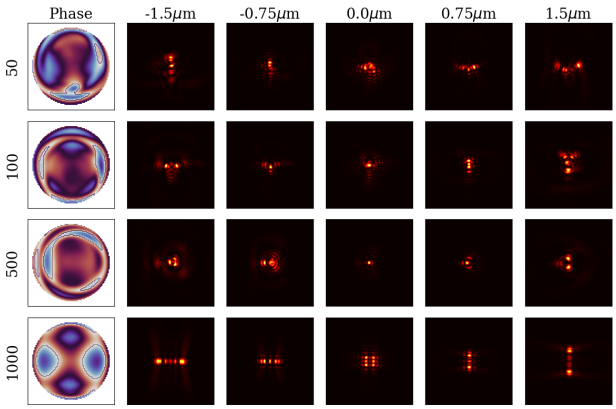

Figure 3. **Designed Phase Masks and corresponding PSFs for specific speeds.** Each row visualizes the neural phase mask designed for tracking particles moving at N nanometers per time interval. Observe that the optimal design for 'fast' moving particles is the Fisher design.

$\sum_{i=1}^{n} p_i$, *with error* $|\epsilon| < 1$.

$$f(t_{end}) - f(t_{start}) \propto \epsilon + \sum_{i=1}^{n} p_i \tag{24}$$

Proof. By assumption, the magnitude of the change corresponding to each event is $\mathcal{T}$. Notice that $\mathcal{T}p_i$ is the log-intensity difference between the previous event time (the reference) and the current event time.

$$\sum_{i=1}^{n} p_i = \frac{1}{\mathcal{T}} \sum_{i=1}^{n} f(t_i) - f(t_{i-1}) \tag{25}$$

The right-hand side is a telescoping sum,

$$\sum_{i=1}^{n} f(t_i) - f(t_{i-1}) = f(t_n) - f(t_0). \tag{26}$$

$t_0 = t_{\text{start}}$ because the first event must occur $t_1 - t_0$ after the start of the interval. Then, the binned event frame is

$$\sum_{i=1}^{n} p_i = \frac{1}{\mathcal{T}} \left(f(t_n) - f(t_{\text{start}}) \right). \tag{27}$$

Finally, $|f(t_n) - f(t_{\text{end}})| = |\delta| < \mathcal{T}$ because if the quantity exceeded the threshold, an additional event would be triggered. Substitute t_{end} for t_n.

$$\sum_{i=1}^{n} p_i = \frac{1}{\mathcal{T}} \left(f(t_{\text{end}}) - f(t_{\text{start}}) + \delta \right) \tag{28}$$

$$= \frac{1}{\mathcal{T}} \left(f(t_{\text{end}}) - f(t_{\text{start}}) \right) + \epsilon \tag{29}$$

Thus, a binned event frame can be approximated as log-intensity difference divided by $\mathcal{T}$ with error $|\epsilon| < 1$. □

As an event camera becomes more sensitive to change ($\mathcal{T}$ decreases), the approximation's percent error decreases because the magnitude of the binned event frame increases but the total absolute error is fixed at most 1.

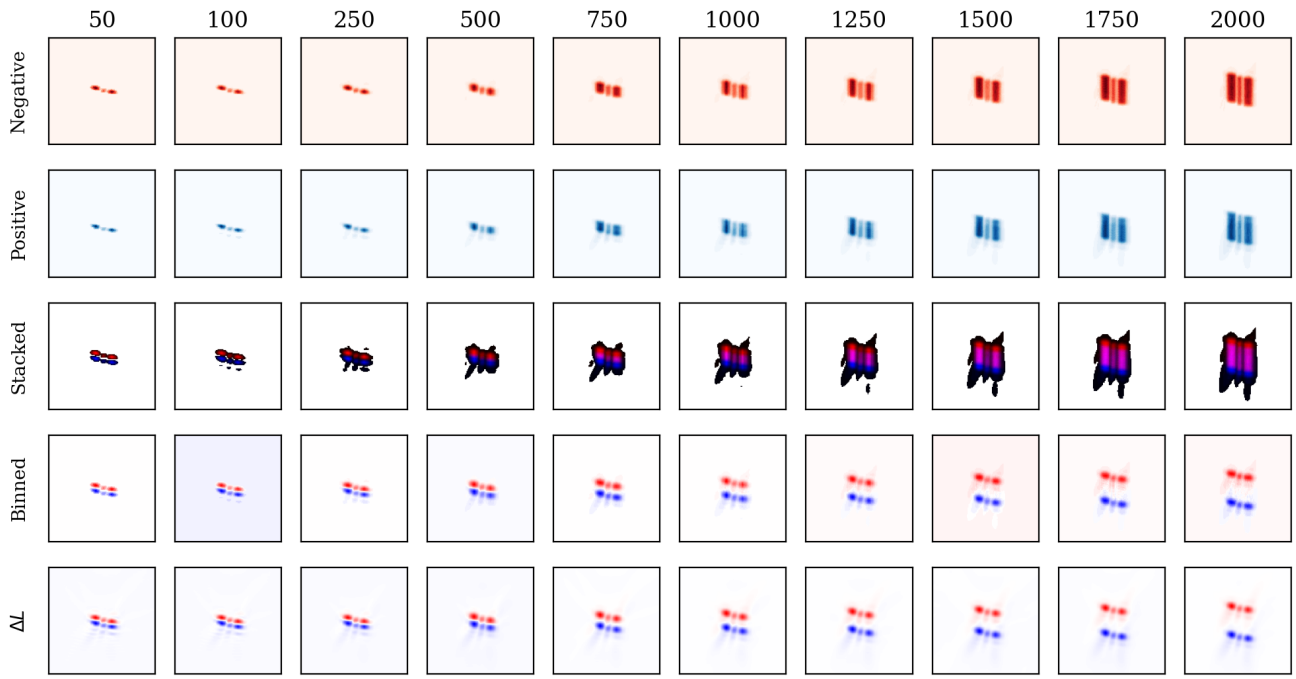

Figure 4. **Event camera measurements of a moving particle with the Fisher mask.** Motion is simulated over a fixed time interval with 100 event samples. Observe a 'fast' moving particle produces an event frame with two copies of a regular PSF: a negative copy at the start location, and a positive copy at the end location. *Row 1*: negative event count over the time interval. *Row 2*: positive event count over the time interval. *Row 3*: the red channel visualizes negative events and the blue channel visualizes positive events. The pink regions represent where the events cancel in a binned measurement. *Row 4*: binned event frame $pos - neg$. *Row 5*: log-intensity difference ΔL.