

Supplementary Material: High-speed Image Reconstruction through Short-term Plasticity for Spiking Cameras

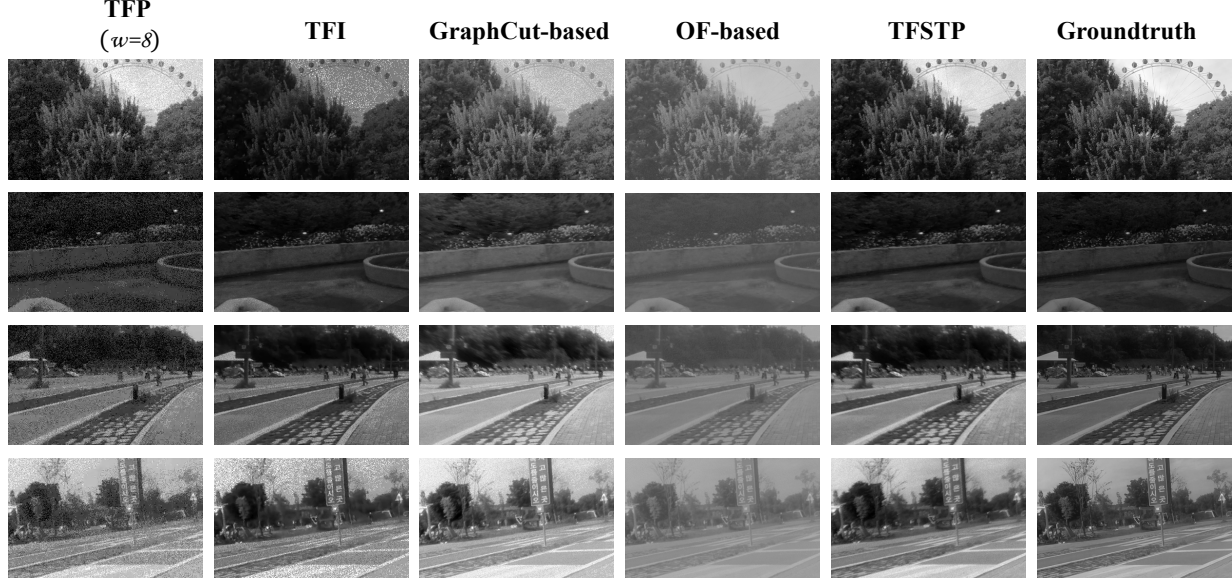


Figure S1. Comparison among different reconstruction methods on synthetic data.

A. Spiking Camera

Fovea-like Sampling Method. Inspired by the sampling mechanism of primate fovea [4, 3], spiking cameras take advantage of spike sequences to represent the brightness change in the spatial-temporal domain [1, 2]. Specifically, the photosensitive units continuously capture photons and increase the photodiode voltage. When the accumulated intensity exceeds a given threshold, a spike is generated and the photodiode voltage is reset to a predefined reset voltage. This process can be formulated as:

$$\text{A spike is generated at time } t^f \text{ if } \int_{t^{f-1}}^{t^f} I(t)dt \geq \phi, \quad (\text{S1})$$

where $I(t)$ denotes the scene radiance, ϕ denotes the predefined threshold, and t^{f-1} represents the firing moment of the last spike. The spikes generated by spiking cameras can be represented by a 3-tuple $\mathcal{S} : \{x, y, t\}$, where $\{x, y\}$ denotes the spatial coordinates of the spikes in the photosensitive units, and t is the spike firing timestamp.

Texture Reconstruction from Inter-spike-interval (TFI). Based on the sampling mechanism of spiking cameras, the photosensitive units receive different scene radiance will trigger spikes with different frequencies. The inter-spike-

interval (ISI) decreases as the scene radiance increases. Therefore, the pixel value (proportional to scene radiance) can be estimated by the interval between two neighboring spikes:

$$\hat{P}_{TFI} = \frac{C}{\Delta t}, \quad (\text{S2})$$

where C refers to the maximum dynamic range of the spiking camera, and Δt represents the ISI.

Texture reconstruction from Playback (TFP). The TFP method infers the pixel value by collecting the spikes in a moving time window. By counting these spikes, we have

$$\hat{P}_{TFP} = \frac{N_w}{w} \cdot C, \quad (\text{S3})$$

where C is the maximum dynamic range of the spiking camera, w is the size of the time window, and N_w is the total number of spikes collected in the time window.

B. Selection of the STP parameters

For the two reconstruction methods of TFSTP and TFM-STP, we have different preferences for parameter selection. The length of the time constants (τ_D and τ_F) and the relationship between them (e.g., $\tau_D < \tau_F$ or $\tau_D = \tau_F$) will

affect the convergent time of the STP dynamics, which implement some sort of adaptive time window. Larger time constants bring the higher contrast, less noise, and more texture details in the static area but more motion blur in the motion area. Smaller time constants bring a smaller convergent rate, leading to the dark area with lower contrast in the early stage of reconstruction (1 *ms*). This effect has been shown in Fig. 10 of the main text. In TFSTP, we need to balance the reconstruction quality between static and motion areas, so smaller time constants are chosen in $STP^{\theta 1}$. But in TFMSTP, we prioritize detecting the motion areas when choosing time constants, so larger time constants are chosen in $STP^{\theta 2}$.

C. Quantitative Evaluation on Simulated Data

In order to make referenced quantitative comparisons, we use the synthesized data¹ made by Zhao et al. [5]. Fig. S1 compares some reconstruction results of different methods on the simulated data. It can be noticed that although the GraphCut-based [6] method can obtain the best results in SSIM, its reconstruction images contain more noise than the results of TFSTP.

References

- [1] Siwei Dong, Tiejun Huang, and Yonghong Tian. Spike camera and its coding methods. In *Data Compression Conference*, 2017. 1
- [2] Siwei Dong, Lin Zhu, Daoyuan Xu, Yonghong Tian, and Tiejun Huang. An efficient coding method for spike camera using inter-spike intervals. In *Data Compression Conference*, 2019. 1
- [3] Richard H Masland. The neuronal organization of the retina. *Neuron*, 76(2):266–280, 2012. 1
- [4] Heinz Wässle. Parallel processing in the mammalian retina. *Nature Reviews Neuroscience*, 5(10):747–757, 2004. 1
- [5] Jing Zhao, Ruiqin Xiong, Hangfan Liu, Jian Zhang, and Tiejun Huang. Spike2imgnet: Learning to reconstruct dynamic scene from continuous spike stream. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 2
- [6] Lin Zhu, Siwei Dong, Jianing Li, Tiejun Huang, and Yonghong Tian. Retina-like visual image reconstruction via spiking neural model. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020. 2

¹The simulated data is publicly available at <https://cove.thecvf.com/datasets/517>.