A. Appendix

A.1. Proof of Theorem 1

To simplify notation, we omit below pixel location $\boldsymbol{X} \in \mathbb{Z}^2$ from events $\{t_i, p_i\}$ whenever it is unambiguous from the context. In a noise-free neuromorphic camera hardware, thresholding in (2) imply equality at threshold:

$$|J(\boldsymbol{X}, t_i) - J(\boldsymbol{X}, t_{i-1})| = \varepsilon. \tag{18}$$

Substituting a Taylor series expansion of the form

$$J(\boldsymbol{X}, t_{i-1}) \approx J(\boldsymbol{X}, t_i) + (t_i - t_{i-1}) J_t(\boldsymbol{X}, t_i), \tag{19}$$

where $J_t(\boldsymbol{X}, t) = \frac{\partial}{\partial t} J(\boldsymbol{X}, t)$, the relation in (18) may be rewritten as:

$$\begin{aligned} |J(\boldsymbol{X}, t_i) - (J(\boldsymbol{X}, t_i) + (t_i - t_{i-1}) J_t(\boldsymbol{X}, t_i))| \\ = |(t_i - t_{i-1}) J_t(\boldsymbol{X}, t_i)| = \varepsilon. \end{aligned} \tag{20}$$

Or equivalently,

$$t_i - t_{i-1} = \frac{\varepsilon}{|J_t(\boldsymbol{X}, t_i)|}. \tag{21}$$

In other words, $\frac{\varepsilon}{|J_t(\boldsymbol{X}, t_i)|}$ is the "rate" at which events are generated. Hence the probability that an event falls within a time interval $[t, t + \tau)$ is

$$M(\boldsymbol{X}) = \begin{cases} \tau \frac{|J_t(\boldsymbol{X}, t)|}{\varepsilon} & \text{if } \tau < \frac{\epsilon}{|J_t(\boldsymbol{X}, t)|} \\ 1 & \text{else.} \end{cases} \tag{22}$$

Intuition here is that if the rate $t_i - t_{i-1}$ is smaller than the window size τ, the event (t_i, p_i) and/or (t_{i-1}, p_{i-1}) will have taken place within this time interval. However, if the rate $t_i - t_{i-1}$ is larger than the window size τ, then events do not necessarily occur within this time interval. As is obvious from Figure 3, $M(\boldsymbol{X})$ scales proportionally to the time window τ and inverse proportionally to the rate $t_i - t_{i-1}$.

To compute $J_t(\boldsymbol{X}, t)$ from APS and IMU, we draw on the well established principles of optical flow. Known as "brightness constancy constraint," spatial translation of pixels over time obeys the following rule [23]:

$$J(\boldsymbol{X} + \Delta\boldsymbol{X}, t + \Delta t) = J(\boldsymbol{X}, t), \tag{23}$$

where $\Delta\boldsymbol{X} = (\Delta x, \Delta y)$ refers to the pixel translation occurring during the time interval Δt. By Taylor expansion, we obtain the classical "optical flow equation":

$$J_t(\boldsymbol{X}, t) \approx -\nabla J(\boldsymbol{X}, t) \boldsymbol{V}(\boldsymbol{X}, t), \tag{24}$$

where

$$\begin{aligned} \nabla J(\boldsymbol{X}, t) &= (J_x(\boldsymbol{X}, t), J_y(\boldsymbol{X}, t)) \\ &:= \left(\frac{\partial}{\partial x} J(\boldsymbol{X}, t), \frac{\partial}{\partial y} J(\boldsymbol{X}, t) \right) \end{aligned} \tag{25}$$

$$\boldsymbol{V}(\boldsymbol{X}, t) = (v_x, v_y) := \left(\frac{\Delta x}{\Delta t} \frac{\Delta y}{\Delta t} \right)^T \tag{26}$$

denotes the spatial gradient and the flow field vector of log intensity $J : \mathbb{Z}^2 \to \mathbb{R}$ at pixel $\boldsymbol{X} \in \mathbb{Z}^2$ at time $t \in \mathbb{R}$. We obtain the pixel velocity $\boldsymbol{V}(\boldsymbol{X}, t)$ from the IMU; and spatial gradient $\nabla J(\boldsymbol{X}, t)$ from APS.

Consider the camera configuration in Figure 5, where a camera on a rotational gimbal is observing a stationary scene. Let $\boldsymbol{\theta}(t) = (\theta_x(t), \theta_y(t), \theta_z(t))^T$ represent the instantaneous 3-axis angular velocity of camera measured by IMU's gyroscope. Then the instantaneous pixel velocity $\boldsymbol{V}(\boldsymbol{X}, t)$ stemming from *yaw*, *pitch*, and *roll* rotations of the camera is computable as

$$\begin{pmatrix} v_x \\ v_y \\ 0 \end{pmatrix} = \boldsymbol{K} \begin{pmatrix} 0 & -\theta_z(t) & \theta_y(t) \\ \theta_z(t) & 0 & -\theta_x(t) \\ -\theta_y(t) & \theta_x(t) & 0 \end{pmatrix} \boldsymbol{K}^{-1} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}, \tag{27}$$

where the camera intrinsic matrix

$$\boldsymbol{K} = \begin{pmatrix} f & \kappa & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{pmatrix} \tag{28}$$

is characterized by focal length f, principal point c_x, c_y, and skew parameter κ ($\kappa = 0$ when the image sensor pixels are square). Hence, the pixel velocity is now entirely determined by the angular velocity and is decoupled from the scene content.

On the other hand, let $A : \mathbb{Z}^2 \times \mathbb{Z} \to \mathbb{R}$ be a synchronous APS output. APS makes measurements on intensity video $I : \mathbb{Z}^2 \times \mathbb{R} \to \mathbb{R}$ as follows:

$$A(\boldsymbol{X}, k) = \alpha \int_{k\eta}^{k\eta + \tau} I(\boldsymbol{X}, t) dt + \beta \tag{29}$$

where $k \in \mathbb{Z}$ denotes the frame number; $1/\eta$ is the frame rate; and α and β are gain and black offset, respectively. The APS exposure time is denoted $\tau < \eta$ hereto also correspond to the time window τ hypothesis in (4) and EPM in (5). In absence of pixel motion, substituting (29) into (1), $I : \mathbb{Z}^2 \times \mathbb{Z} \to \mathbb{R}$ and $J : \mathbb{Z} \times \mathbb{R} \to \mathbb{R}$ yields the relationship

$$J(\boldsymbol{X}, t) = \log \left(\frac{a}{\alpha\tau} A(\boldsymbol{X}, t) - \frac{a\beta}{\alpha\tau} + b \right) \tag{30}$$

at time $t = k\eta$. Taking its spatial gradient yields

$$\begin{aligned} \nabla J(\boldsymbol{X}, t) &= \frac{\nabla A(\boldsymbol{X})}{A(\boldsymbol{X}, t) - O} \\ O &= \beta + \alpha\tau b/a. \end{aligned} \tag{31}$$

In presence of pixel motion, however, APS image $A : \mathbb{Z}^2 \times \mathbb{Z} \to \mathbb{R}$ in (29) is blurred. Assuming constant velocity $\boldsymbol{V}(\boldsymbol{X}, t) = \boldsymbol{V}$ within time $t \in [k\eta, k\eta + \tau)$, the spatial gradient $\nabla A(\boldsymbol{X})$ is attenuated by $\tau|\boldsymbol{V}(\boldsymbol{X}, t)|$ (proof below).
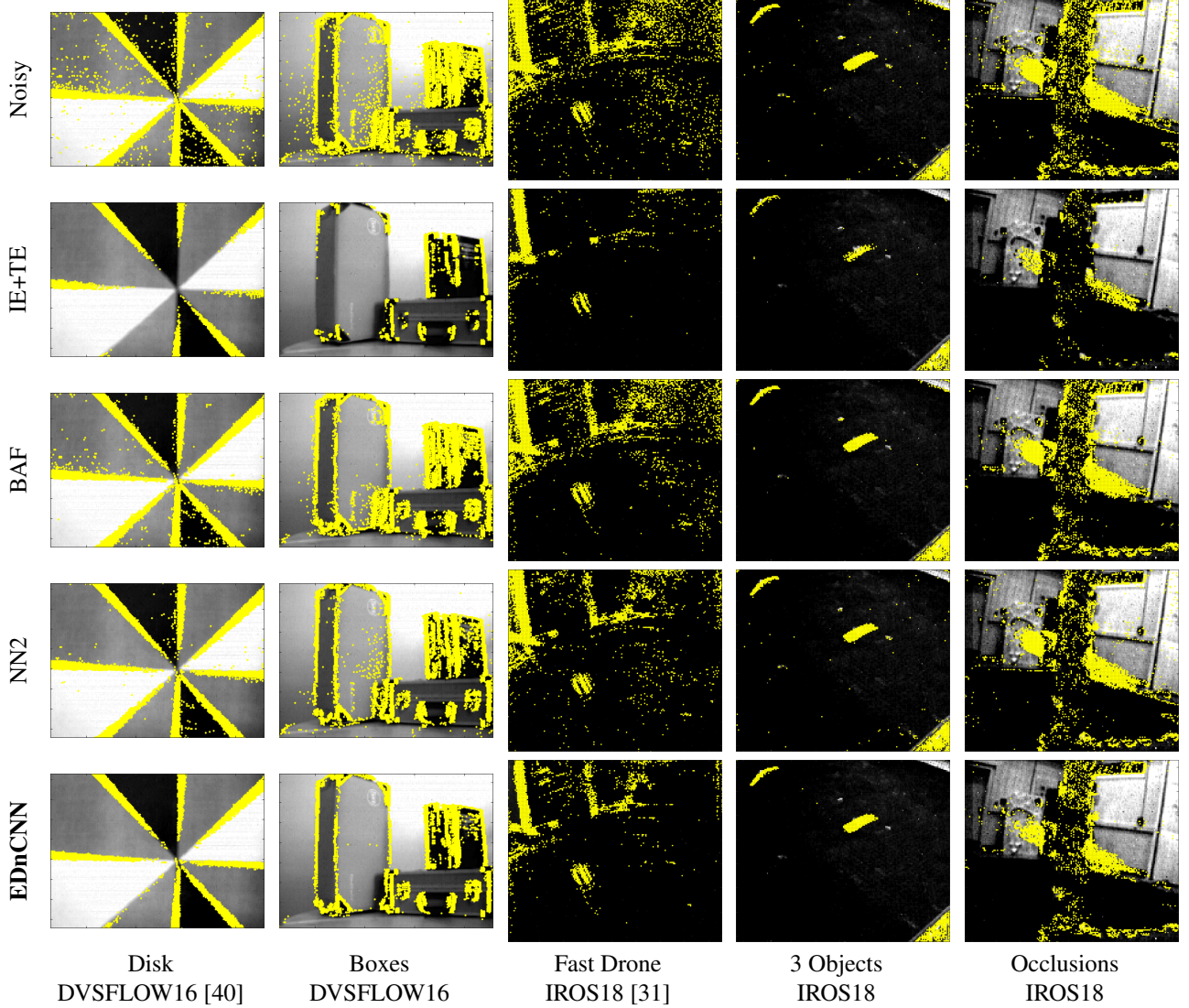
Figure 11. Additional qualitative results from DVS Optical Flow and IROS18. (First Row) A single APS frame from each dataset. The remaining rows show the APS frame overlayed with denoised DVS events from each algorithm. The APS images for columns "Fast Drone" and "3 Objects" have been contrast enhanced but remain dark due to limited signal. Limited APS signal does not impact DVS event generation.

Therefore, to correct for the attenuation we revise (32) as follows:

$$\nabla J(\boldsymbol{X},t) = \frac{\tau \nabla A(\boldsymbol{X})}{A(\boldsymbol{X},t) - O} \begin{pmatrix} |V_x(\boldsymbol{X},t)| & 0 \\ 0 & |V_y(\boldsymbol{X},t)| \end{pmatrix}. \tag{32}$$

To understand the impact of the blur on derivatives, consider a canonical edge image $I(\boldsymbol{X},t) = U(\boldsymbol{X} + \boldsymbol{V}t)$ (where U is a unit step function in x direction) crossing pixel $\boldsymbol{X} = \binom{0}{0}$ at time $t = 0$ and frame $k = 0$:

$$I(\boldsymbol{X},t) = U(\boldsymbol{X} + t\boldsymbol{V}(\boldsymbol{X},t)). \tag{33}$$

In absence of motion, the APS spatial derivative $A_x(\boldsymbol{X},t) := \frac{\partial}{\partial x} A(\boldsymbol{X},t)$ has the following value at pixel

location $\boldsymbol{X} = \binom{0}{0}$ and frame $k = 0$:

$$\begin{aligned} A_x(0,0,0) &= \alpha \int_0^\tau \left. \frac{\partial}{\partial x} U(\boldsymbol{X}) \right|_{\boldsymbol{X}=(0,0)} dt + \beta \\ &= \alpha\tau + \beta. \end{aligned} \tag{34}$$

By contrast, $A_x(\boldsymbol{X},0)$ with non-trivial motion $\boldsymbol{V}$ has the

following form:

$$A_x(\boldsymbol{X}, k) = \alpha \int_{k\eta}^{k\eta+\tau} \frac{\partial}{\partial x} U(\boldsymbol{X}) \Big|_{\boldsymbol{X}=(0,0)} dt + \beta$$

$$= \alpha \int_{k\eta}^{k\eta+\tau} \delta(\boldsymbol{X} + t\boldsymbol{V}(\boldsymbol{X}, t))(1 + tV_x(\boldsymbol{X}, t))dt + \beta.$$

$$(35)$$

At $k = 0$ and $\boldsymbol{X} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$,

$$A_x(0, 0, 0)$$

$$= \alpha \int_0^\tau \delta((0,0) + t\boldsymbol{V}(0,0,t))(1 + tV_x(0,0,t))dt + \beta$$

$$= \frac{\alpha}{|\boldsymbol{V}|} + \beta.$$

$$(36)$$

Comparing (34) and (36), we confirm that the spatial derivatives are attenuated by $\tau|\boldsymbol{V}(\boldsymbol{X}, t)|$. The above analysis generalizes to any pixels, any time, any edge orientations.

A.2. Proof of Optimal Classifier

The mean of $E_{opt}(\boldsymbol{X})$ is $M(\boldsymbol{X})$ due to the Bernoulli probability in (5). By law of large numbers, (16) converges to (15). We use (16) in our work because implementation is far simpler than (14) or (15).

A.3. Calibration Optimization

Although this calibration process does not have to run in real-time (since EPM is only used during benchmarking or training), one can speed up the process by appealing to the fact that ϵ is only used in the final step of Theorem 1 in (6). Contrast this to O, which is used to find $J_t(\boldsymbol{X}, t)$ in (7). Hence searching for the optimal $\widehat{O}$ value is more computationally intensive than searching for $\widehat{\varepsilon}$, in general. Hence a cascading of two 1D search algorithm of the form:

$$(\widehat{\varepsilon}, \widehat{O}) = \arg\max_O \left(\max_\varepsilon \log P[E|\varepsilon, O] \right), \quad (37)$$

is more efficient than a literal implementation of a single 2D search algorithm in (37).

A.4. Additional Qualitative Evaluation

Figure 11 illustrates how previous denoising algorithms performed on the DVS Optical Flow and IROS18 datasets. IE+TE removes most of the noise, but also removes a large amount of real events. BAF and NN2 allow obvious noise through, but EDnCNN removes a large amount of the noise while retaining the signal in the DVS events.