Supplementary Information:

Neuromorphic High-Throughput Imaging via Transport-Induced Differential Sensing

Chutian Wang[1], Shuo Zhu[1], Ni Chen[1], Jingqian Wu[1], Shuo Li[1], Yuxing Li[1], Yanmin Zhu[1,2], Rongzhou Chen[1], Yunping Zhang[1,3], and Edmund Y. Lam[1,4,*]

[1]Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam, Hong Kong, China
[2]School of Biomedical Engineering, The University of Hong Kong, Pokfulam, Hong Kong, China
[3]Department of Applied Data Science, Hong Kong Shue Yan University, North Point, Hong Kong, China
[4]School of Innovation, The University of Hong Kong, Pokfulam, Hong Kong, China
[*]e-mail: elam@eee.hku.hk

Contents

Supplementary Note 1: Throughput-constrained spatiotemporal sampling

We consider sensing over a spatiotemporal domain $\Omega \times T$ under a fixed throughput constraint. Let $w(\mathbf{x},t) \geq 0$ denote a spatiotemporal sampling density such that the total number of statistically independent measurements satisfies

$$\int_{\Omega}\int_{T} w(\mathbf{x},t)\,\mathrm{d}t\,\mathrm{d}\mathbf{x} = B, \tag{S1}$$

where B denotes the available sensing throughput.

For synchronous frame-based acquisition, the sampling density is signal-independent and separable,

$$w_F(\mathbf{x},t) = w_F(\mathbf{x})\,w_F(t). \tag{S2}$$

Here $w_F(\mathbf{x})$ represents spatial sampling on the sensor lattice and is proportional to the number of pixels N_{pix}, which defines the native spatial resolution, while $w_F(t)$ represents temporal sampling and is proportional to the frame rate f_{fps}. Substituting Equation (S2) into Equation (S1) yields

$$B \propto N_{\mathrm{pix}}\,f_{\mathrm{fps}}, \tag{S3}$$

which explicitly captures the coupling between spatial resolution and temporal sampling imposed by a fixed throughput budget in synchronous imaging.

In contrast, asynchronous event-based sensing induces a signal-dependent sampling density. Under locally rigid sweep motion and approximately stationary texture statistics within $\Omega \times T$, a separable approximation remains valid,

$$w_E(\mathbf{x},t) \approx w_E(\mathbf{x})\,w_E(t). \tag{S4}$$

The spatial activation density $w_E(\mathbf{x})$ is governed by temporal variations of the logarithmic intensity. Under brightness constancy, these variations are determined by the motion-aligned first-order increment $-\nabla L(\mathbf{x})\cdot\mathbf{v}$, leading to the activation model

$$w_E(\mathbf{x}) \propto \mathbf{1}(|-\nabla L(\mathbf{x})\cdot\mathbf{v}| > c)\,, \tag{S5}$$

where c denotes the fixed temporal contrast threshold of the event sensor. This defines an active spatial subset

$$\Omega' = \{\mathbf{x}\in\Omega : |-\nabla L(\mathbf{x})\cdot\mathbf{v}| > c\}\,, \qquad \Omega' \subseteq \Omega, \tag{S6}$$

and therefore $\#\Omega' \leq \#\Omega$ on the discrete sensor lattice.

Under the same global throughput constraint Equation (S1), asynchronous sensing reallocates measurements from uniform coverage of Ω to the activated subset Ω', which is enriched with motion-aligned gradient information. The temporal factor $w_E(t)$ corresponds to the event triggering rate and remains approximately stable under stationary texture statistics; a probabilistic justification is provided in a later supplementary note. This reallocation explains how, under a fixed throughput budget, asynchronous sensing concentrates measurements on informative first-order structures rather than uniformly sampling zeroth-order photometric content.

Figure S1 provides a schematic illustration of how a fixed spatio-temporal bandwidth budget constrains typical imaging system designs, positioning frame-based cameras along a trade-off between spatial sampling density and temporal throughput, and highlighting how asynchronous event sensing enables an alternative operating regime through hybrid fusion. Within this perspective, conventional frame cameras appear in two limiting forms: high-resolution multi-megapixel CMOS sensors that prioritize dense spatial sampling at the expense of temporal throughput, and single-pixel or few-pixel computational imaging systems based on compressive sampling, which trade spatial parallelism for ultrafast temporal acquisition[1].

Supplementary Note 2: Circuit-origin of event measurements and Physical operator

We briefly summarize the circuit-level origin of event measurements and their physical interpretation, which motivates the linear event generation operator used in the main text.

As illustrated in Figure S2, an event camera pixel converts incident photocurrent into a logarithmic voltage signal through a logarithmic $I-V$ circuit. Let $I(\mathbf{x},t)$ denote the incident optical intensity and $L(\mathbf{x},t) \triangleq \log I(\mathbf{x},t)$ the corresponding logarithmic intensity. A pixel continuously monitors the temporal change of $L(\mathbf{x},t)$ and emits an event whenever the accumulated increment exceeds a fixed contrast threshold c determined by the comparator circuit,

$$\Delta L(\mathbf{x}_k,t_k) = L(\mathbf{x}_k,t_k+\Delta t_k) - L(\mathbf{x}_k,t_k) = p_k\, c, \qquad p_k \in \{-1,+1\}, \tag{S7}$$

where p_k denotes the polarity of the log-intensity change and Δt_k is the inter-event interval at pixel $\mathbf{x}_k$.

The threshold c is a hardware-defined parameter governed by the comparator and reset circuitry, and is treated as constant throughout this work. Upon threshold crossing, the pixel voltage is reset, resulting in a sequence of polarity-coded events that encode the sign of the temporal log-intensity increment, as schematically shown in Figure S2b.

Under the photometric constancy assumption for rigid motion, the temporal derivative of the logarithmic intensity satisfies

$$\frac{\partial L(\mathbf{x},t)}{\partial t} = -\nabla L(\mathbf{x},t)\cdot \mathbf{v}(t), \tag{S8}$$

where $\mathbf{v}(t)$ denotes the instantaneous image-plane velocity. Over a sufficiently short temporal neighborhood, ∇L and $\mathbf{v}$ can be treated as locally constant, yielding a first-order finite-difference approximation

$$\Delta L(\mathbf{x};\Delta t) \approx -\nabla L(\mathbf{x})\cdot \mathbf{v}\Delta t. \tag{S9}$$

Combining the circuit-triggering condition in Equation (S7) with the motion-induced increment in Equation (S9), the event stream provides multiple measurements of a motion-aligned first-order feature of the logarithmic intensity field. After motion compensation to a reference time, these measurements can be aggregated to form a linear observation of the form

$$\mathbf{b}_E = \mathbf{A}_{\mathbf{v}} L(\mathbf{x}) + \mathbf{n}_E, \qquad \mathbf{A}_{\mathbf{v}} \triangleq -\mathbf{v}\cdot\nabla, \tag{S10}$$

where $L(\mathbf{x}) = \log I(\mathbf{x})$, $\mathbf{A}_{\mathbf{v}}$ denotes the event generation operator induced by motion, and $\mathbf{n}_E$ captures the effective noise after temporal aggregation.

This formulation establishes a direct link between the circuit-level event triggering mechanism and the linear feature-domain measurement operator used in the main text.

Supplementary Note 3: Probabilistic modeling of the event camera

This note formalizes how the motion-aligned first-order increment

$$\mu(\mathbf{x}) \triangleq -\nabla L(\mathbf{x}) \cdot \mathbf{v}, \tag{S11}$$

is converted by the event circuit into discrete stochastic events, and how temporal accumulation yields a continuous-valued statistic that is approximately linear in μ and can be used as a differentiable measurement.

S3.1 Instantaneous polarity as a thresholded random variable

Following Supplementary Note 2, the event circuit compares a noisy log-domain increment against a symmetric threshold $\pm c$. Specifically, within a sufficiently short interval δt, define the (pre-threshold) triggering variable

$$Z \sim \mathcal{N}\left(\mu, \sigma_Z^2\right), \tag{S12}$$

where μ is the motion-aligned log-intensity increment in Equation (S11) and σ_Z^2 summarizes the effective uncertainty induced by photon statistics and electronic readout (Supplementary Note 2). The ternary polarity outcome in that interval is

$$S = \begin{cases} +1, & Z > c, \\ -1, & Z < -c, \\ 0, & |Z| \le c. \end{cases} \tag{S13}$$

Therefore, the probabilities of positive and negative events are

$$p_+(\mu) = \Pr(S = +1) = 1 - \Phi\left(\frac{c-\mu}{\sigma_Z}\right), \qquad p_-(\mu) = \Pr(S = -1) = 1 - \Phi\left(\frac{c+\mu}{\sigma_Z}\right), \tag{S14}$$

and the expected polarity (the mean signed response) is

$$m(\mu) \triangleq \mathbb{E}[S] = p_+(\mu) - p_-(\mu) = \Phi\left(\frac{\mu+c}{\sigma_Z}\right) - \Phi\left(\frac{c-\mu}{\sigma_Z}\right), \tag{S15}$$

which is a symmetric sigmoidal function of μ (see Supplementary Fig. S3).

S3.2 Temporal accumulation and Skellam-to-Gaussian approximation

Consider a temporal window of duration Δt consisting of $N = \Delta t / \delta t$ independent trials of S. Let

$$N_+ = \#\{S_i = +1\}, \qquad N_- = \#\{S_i = -1\}, \qquad N_0 = \#\{S_i = 0\}, \tag{S16}$$

then

$$(N_+, N_-, N_0) \sim \text{Multinomial}(N;\, p_+(\mu),\, p_-(\mu),\, 1 - p_+(\mu) - p_-(\mu)). \tag{S17}$$

When δt is sufficiently small, the accumulated positive and negative counts are well approximated by two independent Poisson processes,

$$N_+ \sim \text{Poisson}(\Lambda_+), \qquad N_- \sim \text{Poisson}(\Lambda_-), \qquad \Lambda_\pm \triangleq r_\pm(\mu)\,\Delta t, \quad r_\pm(\mu) \triangleq \frac{p_\pm(\mu)}{\delta t}. \tag{S18}$$

Define the accumulated bipolar count (signed accumulation)

$$C \triangleq N_+ - N_-. \tag{S19}$$

Under Equation (S18), C follows a Skellam distribution with moments

$$\mathbb{E}[C] = \Lambda_+ - \Lambda_-, \qquad \mathrm{Var}(C) = \Lambda_+ + \Lambda_-. \tag{S20}$$

We then define the time-averaged polarity statistic

$$\hat{g} \triangleq \frac{C}{\Delta t}. \tag{S21}$$

It follows that

$$\mathbb{E}[\hat{g}] = r_+(\mu) - r_-(\mu), \qquad \mathrm{Var}(\hat{g}) = \frac{r_+(\mu) + r_-(\mu)}{\Delta t}. \tag{S22}$$

Moreover, for moderate-to-large $\Lambda_+ + \Lambda_-$, the Skellam distribution approaches a Gaussian by the central limit effect, yielding the approximation

$$\hat{g} \approx \mathcal{N}\left(r_+(\mu) - r_-(\mu),\ \frac{r_+(\mu) + r_-(\mu)}{\Delta t} \right). \tag{S23}$$

This provides a continuous-valued, approximately Gaussian measurement whose variance decreases as Δt increases, i.e., temporal accumulation suppresses stochastic fluctuations.

S3.3 Near-linear response around $\mu = 0$

In the small-increment regime relevant to sub-pixel motion, $m(\mu)$ in Equation (S15) admits a local linearization around $\mu = 0$,

$$m(\mu) \approx \eta\, \mu, \qquad \eta \triangleq m'(0) = \frac{2}{\sigma_Z}\, \phi\left(\frac{c}{\sigma_Z} \right), \tag{S24}$$

where $\phi(\cdot)$ is the standard Gaussian probability density function. Since $r_\pm(\mu) = p_\pm(\mu)/\delta t$, the mean of the time-averaged statistic satisfies

$$\mathbb{E}[\hat{g}] \approx \frac{\eta}{\delta t}\, \mu, \tag{S25}$$

and its variance obeys the scaling in Equation (S22). The validity of this local linear approximation is illustrated in Supplementary Fig. S3j, which highlights a near-linear response envelope around $\mu \approx 0$ under a locally constant variance assumption.

Supplementary Note 4: Cramér-Rao lower bound of frame and event camera

Building on the statistical measurement model analyzed in Supplementary Note 3, we examine the fundamental estimation limits imposed by frame and event sensing under the same throughput constraints.

The Cramér–Rao lower bound (CRLB) provides a lower bound on the variance of unbiased estimators[2–4]. Specifically, if $\hat{\Theta}$ is an unbiased estimator of a parameter Θ, then

$$\mathrm{Var}(\hat{\Theta}) \geq \frac{1}{\mathcal{I}(\Theta)}, \tag{S26}$$

where $\mathcal{I}(\Theta)$ denotes the Fisher information associated with the observation model.

Let $\mathbf{b}$ denote observable data generated according to the likelihood model $p(\mathbf{b} \mid \Theta)$. The Fisher information is defined as

$$\begin{aligned}\mathcal{I}(\Theta) &= \mathbb{E}\left[-\frac{\partial^2}{\partial \Theta^2}\log p(\mathbf{b} \mid \Theta)\right] \\ &= \mathbb{E}\left[\left(\frac{\partial}{\partial \Theta}\log p(\mathbf{b} \mid \Theta)\right)\left(\frac{\partial}{\partial \Theta}\log p(\mathbf{b} \mid \Theta)\right)^{\mathsf{T}}\right] \\ &= \mathbb{E}[J(\Theta)J(\Theta)^{\mathsf{T}}],\end{aligned} \tag{S27}$$

where

$$J(\Theta) = \frac{\partial}{\partial \Theta}\log p(\mathbf{b} \mid \Theta) \tag{S28}$$

is the score function. The expectation is taken with respect to the distribution of the observable data $\mathbf{b}$.

When multiple independent measurements of the parameter Θ are available, the additive structure of Fisher information becomes explicit. Let $\{\mathbf{b}_t\}_{t=1}^{T}$ denote T independent observations drawn from $p(\mathbf{b}_t \mid \Theta)$. Under the iid assumption, the joint likelihood factorizes as

$$p(\{\mathbf{b}_t\} \mid \Theta) = \prod_{t=1}^{T} p(\mathbf{b}_t \mid \Theta). \tag{S29}$$

The joint score function is therefore

$$\begin{aligned}J(\Theta) &= \frac{\partial}{\partial \Theta}\log p(\{\mathbf{b}_t\}_{t=1}^{T} \mid \Theta) \\ &= \sum_{t=1}^{T}\frac{\partial}{\partial \Theta}\log p(\mathbf{b}_t \mid \Theta) = \sum_{t=1}^{T} J_t(\Theta),\end{aligned} \tag{S30}$$

where $J_t(\Theta)$ denotes the score contribution from the t-th measurement. Taking expectation yields the additive Fisher information identity

$$\mathcal{I}(\Theta) = \sum_{t=1}^{T}\mathcal{I}_t(\Theta). \tag{S31}$$

Accordingly,

$$\mathrm{Var}(\hat{\Theta}) \geq \frac{1}{\sum_{t=1}^{T}\mathcal{I}_t(\Theta)}, \tag{S32}$$

indicating that repeated independent measurements reduce estimator variance through information accumulation.

S4.1 CRLB of frame measurements

In the frame modality, the observable data consist of photon counts $\{\mathbf{N}_t\}_{t=1}^{T}$ acquired independently at each exposure. The parameter of interest is the underlying scene intensity I, observed through the photon count $\lambda = I\,\delta t$ accumulated over an exposure duration δt. For completeness, we summarize standard CRLB expressions for frame-based photon-counting measurements, which serve as a reference scaling under the throughput constraint.

Poisson noise. Under photon-limited conditions,

$$\mathbf{N}_t \overset{\text{iid}}{\sim} \text{Poisson}(\lambda). \tag{S33}$$

The score function with respect to I is

$$J_t(I) = \frac{\partial}{\partial I} \log p(\mathbf{N}_t \mid I) = \delta t \left(\frac{\mathbf{N}_t}{\lambda} - 1 \right). \tag{S34}$$

Using $\mathbb{E}[\mathbf{N}_t] = \lambda$ and $\text{Var}(\mathbf{N}_t) = \lambda$, the per-frame Fisher information becomes

$$\mathcal{I}_t(I) = \frac{\delta t}{I}. \tag{S35}$$

By additivity,

$$\mathcal{I}(I) = \frac{T\,\delta t}{I}, \tag{S36}$$

leading to the CRLB

$$\text{Var}(\hat{I}) \geq \frac{I}{T\,\delta t}. \tag{S37}$$

Mixed photon and readout noise. When readout noise contributes additively, the photon-counting process can be approximated by

$$\mathbf{N}_t \overset{\text{iid}}{\sim} \mathcal{N}(\lambda,\, \lambda + \sigma_r^2). \tag{S38}$$

Treating the variance as locally constant with respect to I, the per-frame Fisher information is approximated as

$$\mathcal{I}_t(I) \approx \frac{(\delta t)^2}{\lambda + \sigma_r^2}. \tag{S39}$$

Summing over T frames yields

$$\mathcal{I}(I) \approx \frac{T(\delta t)^2}{I\,\delta t + \sigma_r^2}, \tag{S40}$$

and the corresponding CRLB

$$\text{Var}(\hat{I}) \gtrsim \frac{I\,\delta t + \sigma_r^2}{T(\delta t)^2}. \tag{S41}$$

These expressions provide a reference precision scaling for frame-based measurements under photon and readout noise.

S4.2 CRLB of event measurements

In the event modality, the observable data are accumulated bipolar event statistics $\{\hat{g}_t\}_{t=1}^{T}$ obtained by temporally aggregating motion-aligned events. The parameter of interest remains the scene intensity I, while the event statistics encode the motion-aligned log-intensity gradient

$$\mu \triangleq \mathbf{A}_{\mathbf{v}} \log I.$$

Gaussian approximation. As shown in Supplementary Note 3, accumulated bipolar events are well approximated by a Gaussian distribution,

$$\hat{g}_t \sim \mathcal{N}\left(m(\mu), \sigma_g^2(\mu)\right), \qquad \sigma_g^2(\mu) = \frac{p_+(\mu) + p_-(\mu)}{N}. \tag{S42}$$

Treating the variance as locally constant with respect to μ, the score function becomes

$$J(\mu) \approx \frac{\hat{g} - m(\mu)}{\sigma_g^2(\mu)} m'(\mu). \tag{S43}$$

The Fisher information with respect to μ is therefore

$$\mathcal{I}_\mu(\mu) \approx \frac{N\left(m'(\mu)\right)^2}{p_+(\mu) + p_-(\mu)}. \tag{S44}$$

In the near-linear regime where $m'(\mu) \approx \eta$, this yields the envelope

$$\mathcal{I}_\mu(\mu) \leq \frac{N\eta^2}{\alpha_0}, \qquad \alpha_0 \triangleq p_+(0) + p_-(0), \tag{S45}$$

corresponding to the maximum Fisher information attainable under the Gaussian approximation.

Gradient envelope. Since $\mu = \mathbf{A}_{\mathbf{v}} \log I$, the Jacobian with respect to the intensity field is

$$\frac{\partial \mu}{\partial I} = \mathbf{A}_{\mathbf{v}} \operatorname{diag}\left(\frac{1}{I}\right). \tag{S46}$$

The Fisher information for estimating I from event statistics therefore takes the quadratic form

$$\mathcal{I}(I) = \left(\frac{\partial \mu}{\partial I}\right)^{\mathsf{T}} \mathcal{I}_\mu(\mu) \left(\frac{\partial \mu}{\partial I}\right), \tag{S47}$$

which explicitly shows that $\mathbf{A}_{\mathbf{v}}$ acts as a motion-aligned gradient envelope governing information transfer from I to event measurements. The corresponding event-based CRLB follows as $\mathrm{Var}(\hat{I}) \gtrsim 1/\mathcal{I}(I)$ under the adopted approximation.

Supplementary Note 5: Working with conventional computational imaging solvers

Conventional computational imaging solvers typically describe the measurement process through a unified forward operator that maps a latent scene representation to observable measurements. This note explains how the NeuroSR formulation can be instantiated within existing computational imaging solvers. Our goal is not to redefine task-specific optical forward models, which are often well established in frame-based or other synchronous imaging modalities, but to clarify how the asynchronous event sensing likelihood can be *plugged into* conventional Maximum a posteriori (MAP)-based solvers in a compatible and modular manner. Throughout this note, the forward image formation model is assumed to be given, as specified in the main text. Here we focus on how this model interfaces with event-based constraints at the optimization level.

S5.1 Maximum a posteriori formulation

We consider a general MAP formulation in which the unknown quantity represents a latent scene variable, such as an intensity image or a task-dependent object/field. Let $I(\mathbf{x})$ denote a latent sharp intensity at a reference time (when applicable), and let θ denote additional parameters to be jointly optimized. Given measurements $\mathbf{b}$, MAP estimation takes the form

$$\begin{aligned} I^*(\mathbf{x}), \theta^* &= \arg\max_{I,\theta} p(I,\theta|\mathbf{b}) = \frac{p(\mathbf{b}|I,\theta)p(I)p(\theta)}{p(\mathbf{b})}, \\ &\propto \arg\min_{I,\theta} \underbrace{-\log p(\mathbf{b}|I,\theta)}_{\text{Log-Likelihood}} + \underbrace{\big(-\log p(I) - \log p(\theta)\big)}_{\text{Prior}} + \underbrace{\log p(\mathbf{b})}_{\text{Constant}}, \end{aligned} \tag{S48}$$

where the first term encodes the measurement likelihood and the remaining terms impose priors or regularization. Different sensing modalities correspond to different likelihood terms, while sharing the same latent variables.

S5.2 Event-based MAP estimation

A characteristic property of event measurements is that motion and intensity can be efficiently decoupled at the likelihood level. We therefore first consider a motion-resolved ($\theta = \mathbf{v}$) MAP formulation based on the event channel alone. Assuming independent priors on I and $\mathbf{v}$, MAP estimation from event measurements $\mathbf{b}_E$ yields

$$\begin{aligned} I^*(\mathbf{x}), \mathbf{v}^* &= \arg\max_{I,\mathbf{v}} p(I,\mathbf{v} \mid \mathbf{b}_E) \\ &\propto \arg\min_{I,\mathbf{v}} -\log p(\mathbf{b}_E \mid I,\mathbf{v}) - \log p(I) - \log p(\mathbf{v}). \end{aligned} \tag{S49}$$

This formulation admits a natural two-stage instantiation, which is commonly adopted in event-based inference and is fully compatible with the NeuroSR framework.

1. **Motion retrieval from events.** The motion field $\mathbf{v}$ can be estimated independently by leveraging the spatio-temporal contrast encoded in the event measurement $\mathbf{b}_E$. When the latent intensity I is unknown, $\mathbf{b}_E$ alone serves as a motion-sensitive proxy constructed from warped events. In this setting, Equation (S49) can be interpreted probabilistically as selecting the flow that maximizes a decoupled likelihood term, with the intensity-dependent component replaced by a surrogate objective.

In practice, it is well established that correctly motion-compensated events exhibit strong spatial concentration[5], while incorrect motion hypotheses lead to temporally smeared and spatially diffuse event distributions. Using contrast maximization as an efficient surrogate, motion estimation reduces to

$$\mathbf{v}^* = \arg\min_{\mathbf{v}} \underbrace{-\mathrm{Var}\big(\mathbf{b}_E(\mathbf{v})\big)}_{\text{Surrogate fidelity}} + \underbrace{\beta_{\mathbf{v}}\mathcal{R}(\mathbf{v})}_{\text{Flow regularizer}} \quad , \tag{S50}$$

where $\mathcal{R}(\mathbf{v})$ encodes spatial or temporal regularity of the flow field.

Importantly, the objective in Equation (S50) is independent of I and enables an efficient event-only pre-optimization of motion. This step provides a reliable initialization and constraint for subsequent joint intensity–motion refinement.

2. **Intensity retrieval.** Given the estimated motion $\mathbf{v}^*$, the event-channel likelihood simplifies considerably. As analyzed in Supplementary Note 3, temporal accumulation of motion-aligned events yields an approximately Gaussian statistic whose mean is proportional to $\mathbf{A}_{\mathbf{v}^*}\log I$. Under this Gaussian approximation, the event likelihood can be written as

$$p(\mathbf{b}_E \mid I, \mathbf{v}^*) \;\propto\; \exp\Big(-\mathcal{D}\{\mathbf{b}_E(\mathbf{v}^*), \mathbf{A}_{\mathbf{v}^*}\log I\}\Big),$$

leading to the conditional MAP problem

$$I^*(\mathbf{x}) \propto \arg\min_{I} \underbrace{\mathcal{D}\{\mathbf{b}_E(\mathbf{v}^*), \mathbf{A}_{\mathbf{v}^*}\log I\}}_{\text{Event data fidelity}} + \underbrace{\beta_I\mathcal{R}(I)}_{\text{Image regularizer}} \quad , \tag{S51}$$

where the first term is also known as the data fidelity term, since it should provide a large proportion of the back-propagated gradient flow during optimization/training.

In principle, the data-fidelity term in the MAP formulation should be chosen to match the underlying event generation process. Assuming that each motion-compensated event e'_k is conditionally independent given the latent intensity $I(\mathbf{x})$ and the estimated flow $\mathbf{v}^*$, the likelihood of the aligned event cluster $\mathcal{E}' = \{e'_k\}_{k=1}^{N_e}$ factorizes as $\prod_k p(e'_k \mid I)$. Accordingly, the time-averaged event measurement $\mathbf{b}_E = \Sigma_{k=1}^{N_e} e'_k/\Delta t$ admits a Gaussian approximation under temporal accumulation, *i.e.*, $p(\mathbf{b}_E \mid I) \sim \mathcal{N}(\mu(e'_k), \mathrm{Var}(e'_k)/N_e)$, where the mean response satisfies $\mu \propto \mathbf{A}_{\mathbf{v}^*}\log I$, as derived in Supplementary Note 3. This Gaussian approximation provides a statistically consistent and solver-compatible likelihood for event measurements, despite the underlying event generation being nonlinear and asymmetric. An illustrative joint intensity–motion reconstruction example is shown in Supplementary Fig. Figure S4.

S5.3 Joint event-frame MAP and practical instantiation

NeuroSR combines the complementary constraints imposed by frame and event measurements by jointly optimizing their likelihoods under a shared latent intensity and motion field. Conditioning on the pre-estimated motion $\mathbf{v}^*$ from Equation (S50), the resulting joint objective reduces to the event–frame MAP formulation presented in the main text, in which frame and event channels impose complementary likelihood terms on the same latent encoded intensity. This structure is directly compatible with existing frame-based solvers and

enables a plug-and-play integration of the event likelihood without modifying the underlying solver architecture. In practice, the motion field $\mathbf{v}(t)$ is parameterized as a piecewise-linear function and estimated from events using contrast maximization. Motion-compensated event aggregation specifies how the event measurement $\mathbf{b}_E(\mathbf{x};\mathbf{v})$ is constructed in practice, with subpixel support achieved through pixel reassignment in IWE rendering (see Figure S5 and Supplementary Fig. Figure S4). These implementation details determine how the event likelihood enters the joint MAP formulation, while leaving the optimization structure unchanged.

We next summarize representative instantiations of this formulation across three computational imaging tasks: (i) large-FOV industrial inspection, (ii) biological microscopy screening, and (iii) snapshot pixel-SR lensless holography. The table below lists typical frame-domain operators, task-dependent optical forward models, and corresponding regularization choices used in these settings. These examples illustrate how the same joint MAP structure accommodates different imaging configurations.

Table S1. Examples of forward models and corresponding regularizers in typical application scenes. Here, $\downarrow$ denotes the down-sampling operation, and $\int$ denotes the temporal accumulation operation (long-exposure).

Setup	Sensor Model	f_Θ	$\mathcal{R}$
Industrial Inspection	$\downarrow$	Unity / Magnified Relay	2D Image Regularizer (TV)
Microscopic Screening	$\downarrow + \int$	Objective with Occlusion	2D Image Regularizer (TV, L1)
Pixel-SR Holography	$\downarrow + \int$	Free-space Propagation	Complex Field Regularizer (TV, L1)
More			

To accommodate different statistical assumptions or simplifications about the measurement noise, the likelihood terms in the joint MAP formulation can be instantiated with various negative log-likelihood losses, *i.e.*, the distance function $\mathcal{D}\{\cdot\}$. We summarize several commonly used choices in computational imaging below:

Table S2. Examples of negative log-likelihood losses under different statistical assumptions. Here, $\mathbf{a} = \{a_i\}$ denotes the prediction, and $\mathbf{b} = \{b_i\}$ the observation.

$\mathcal{L}(\mathbf{b},\mathbf{a})$	Loss Name	Noise Model
$\|\mathbf{b}-\mathbf{a}\|_2^2$	L_2 loss	Gaussian distribution noise
$\|\mathbf{b}-\mathbf{a}\|_1$	L_1 loss	Laplace distribution
$\sum_i (a_i - b_i \log a_i)$	Poisson loss	Poisson distribution
More		

References

1. Duarte, M. F. *et al.* Single-pixel imaging via compressive sampling. *IEEE signal processing magazine* **25**, 83–91 (2008).

2. Kay, S. M. *Fundamentals of statistical signal processing: estimation theory* (Prentice-Hall, Inc., 1993).

3. Goodman, J. W. *Statistical optics* (John Wiley & Sons, 2015).

4. Fournier, C., Denis, L. & Fournel, T. On the single point resolution of on-axis digital holography. *J. Opt. Soc. Am. A* **27**, 1856–1862 (2010).

5. Gallego, G., Rebecq, H. & Scaramuzza, D. A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3867–3876 (2018).

S1 Supplementary Figures

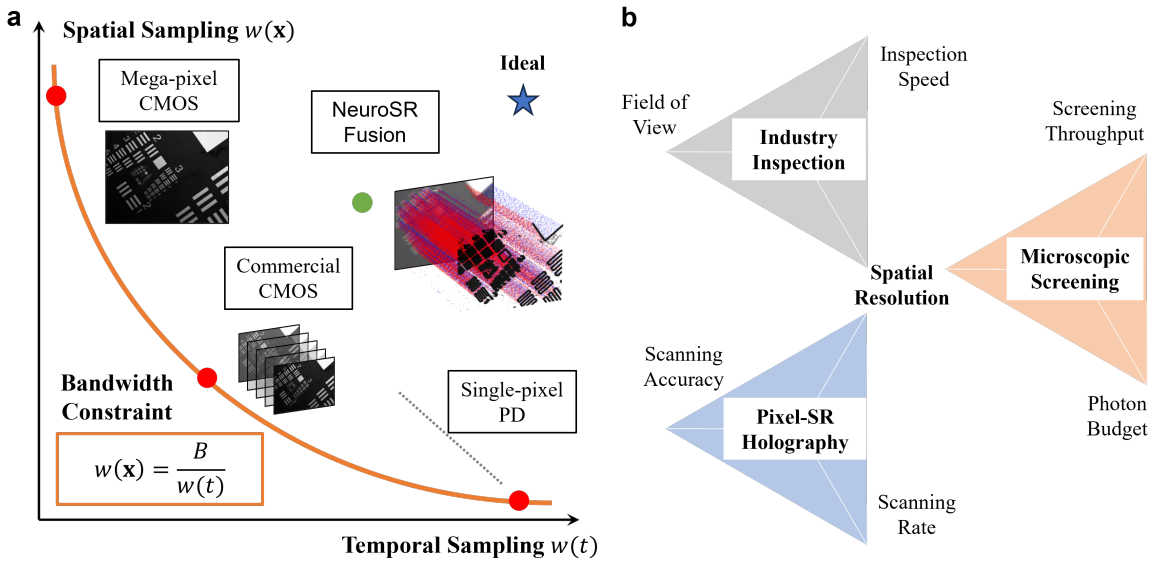

Fig. S1. Throughput-constrained sampling and application regimes. a, Illustration of the spatiotemporal throughput constraint in sensing, where spatial sampling density $w(\mathbf{x})$ and temporal sampling density $w(t)$ are coupled by a fixed budget B. Conventional frame-based imaging operates along the trade-off curve $w(\mathbf{x})w(t) = B$, linking spatial resolution to frame rate. NeuroSR leverages asynchronous, motion-induced sampling to redistribute measurements within this constraint. **b,** Projection of representative high-throughput imaging tasks into the resolution–throughput space, including industrial inspection, microscopic screening, and pixel super-resolution holography, highlighting their distinct operating regimes under shared throughput limitations.

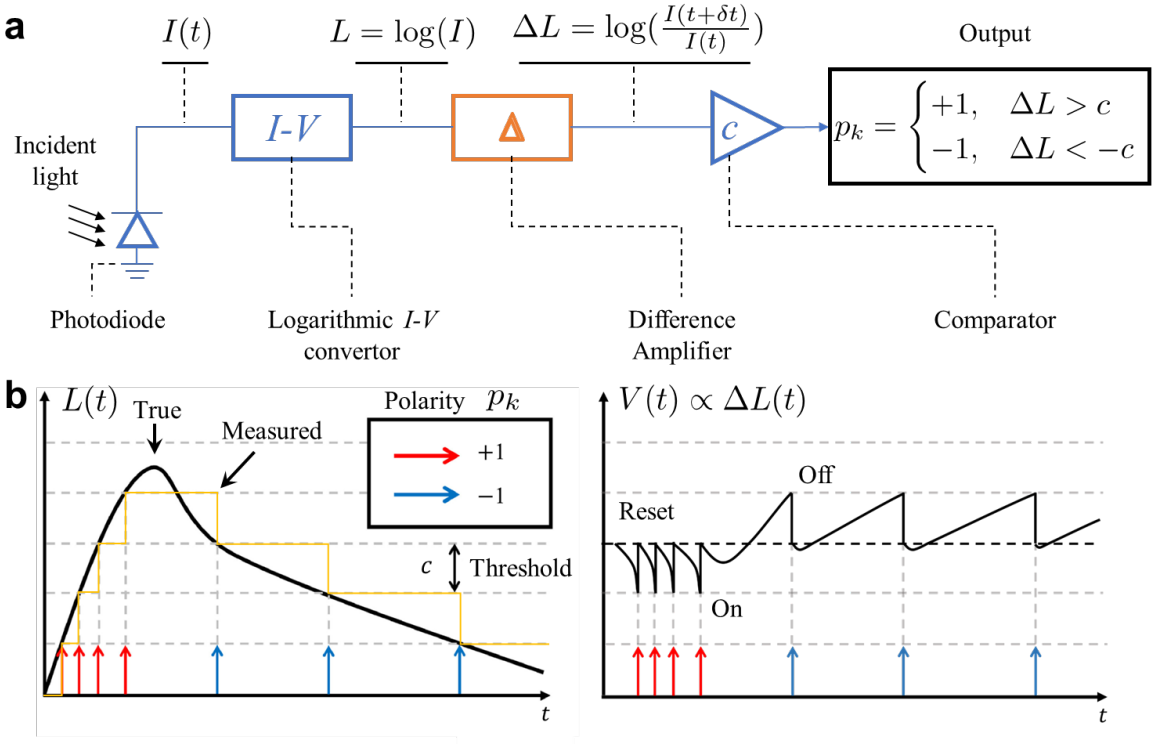

Fig. S2. Asynchronous triggering mechanism of an event camera. a, Circuit-level principle of event generation: a pixel reports an event when the logarithmic intensity change exceeds a fixed contrast threshold c. **b,** Upon threshold crossing, the pixel voltage is reset and a polarity-coded event is emitted, encoding the sign of the log-intensity increment.

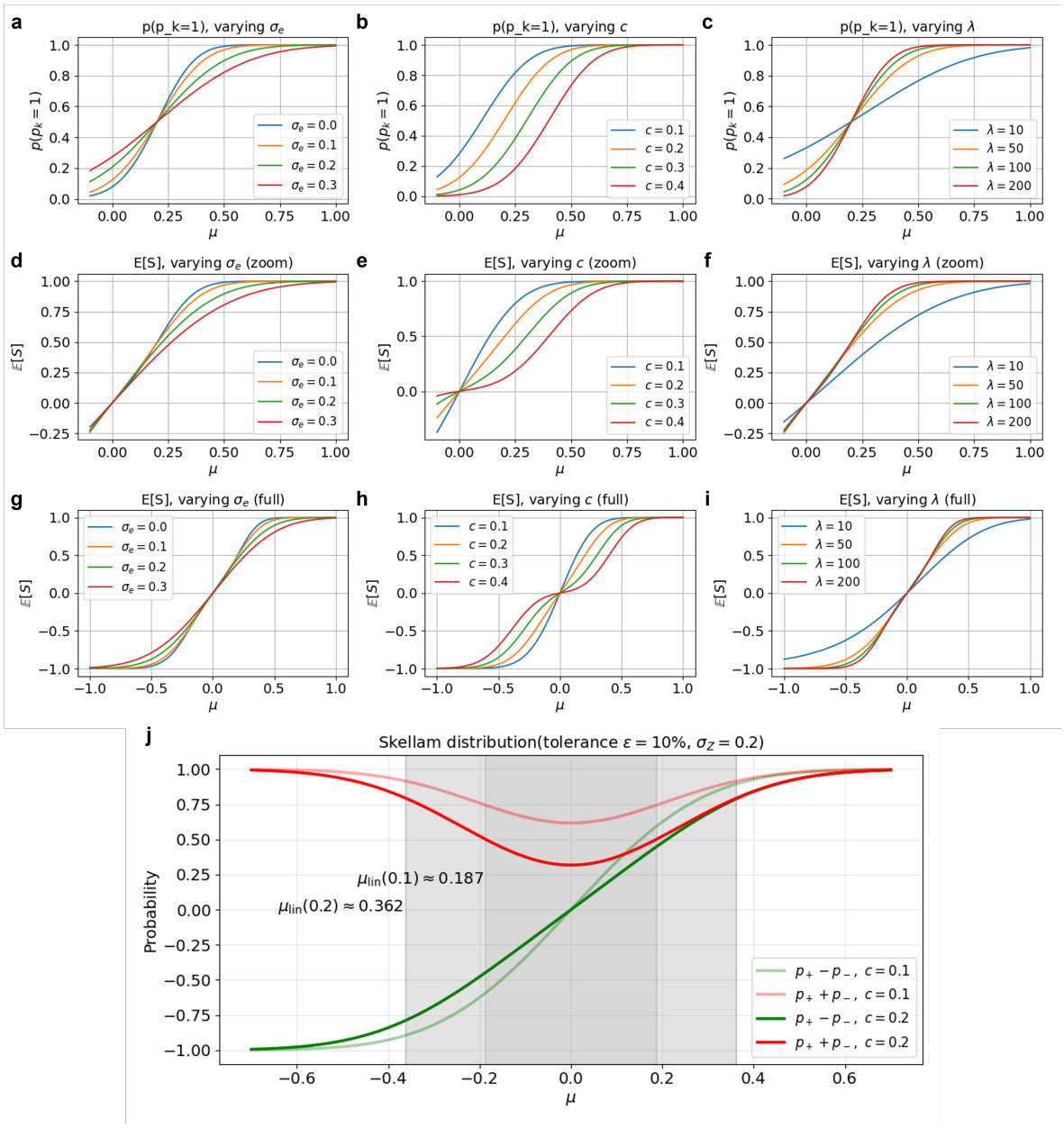

Fig. S3. Event triggering statistics and near-linear bipolar response. a-c, Single-polarity probability $p(p_k = 1)$ as a function of the motion-aligned increment μ under variations of readout noise σ_e, threshold c, and photon level λ (zoomed range). **d-f,** Expected bipolar polarity $\mathbb{E}[S]$ in the same zoomed range, showing reduced sensitivity near $\mu \approx 0$. **g-i,** Full-range $\mathbb{E}[S]$ curves, exhibiting stable sigmoidal transitions. **j,** Linearization envelope under a locally constant variance approximation (σ_Z fixed), where the shaded interval indicates the near-linear regime $|\mu| \leq \mu_{\text{lin}}(c)$ defined by a relative linearization error tolerance.

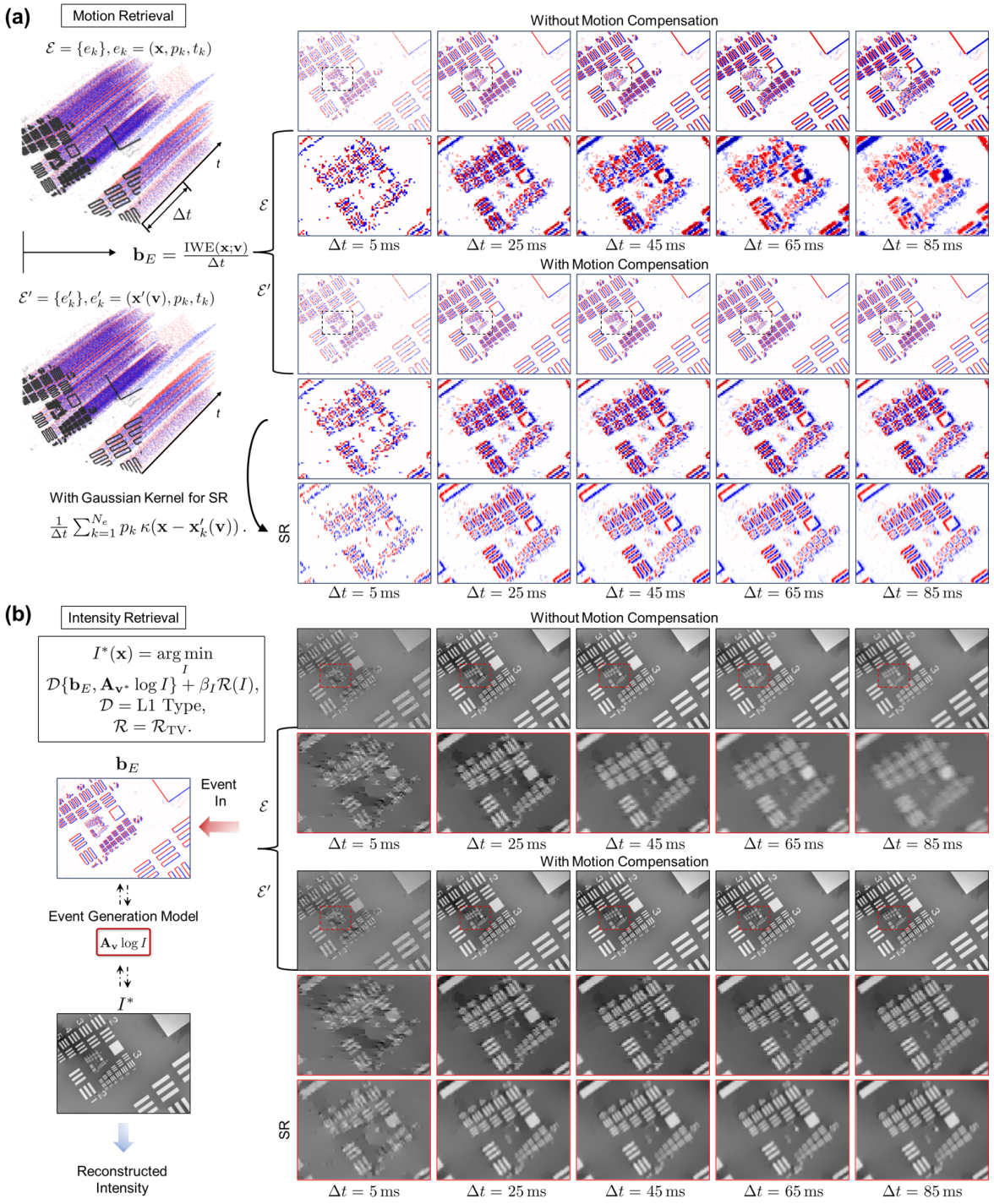

Fig. S4. Asynchronous feature encoding. a, Warping events along the estimated flow $\mathbf{v}^*$ reassigns temporally separated event samples to consistent sub-pixel spatial locations at a common reference time, enabling coherent accumulation of sub-pixel information over an extended temporal window. **b,** Under correct motion compensation, increasing the accumulation duration Δt improves statistical stability. Apparent smearing arises only from motion mismatch or model violation. The observed reconstruction sharpening reflects the recovery of sub-pixel structure. Bottom row (Zoom-in).

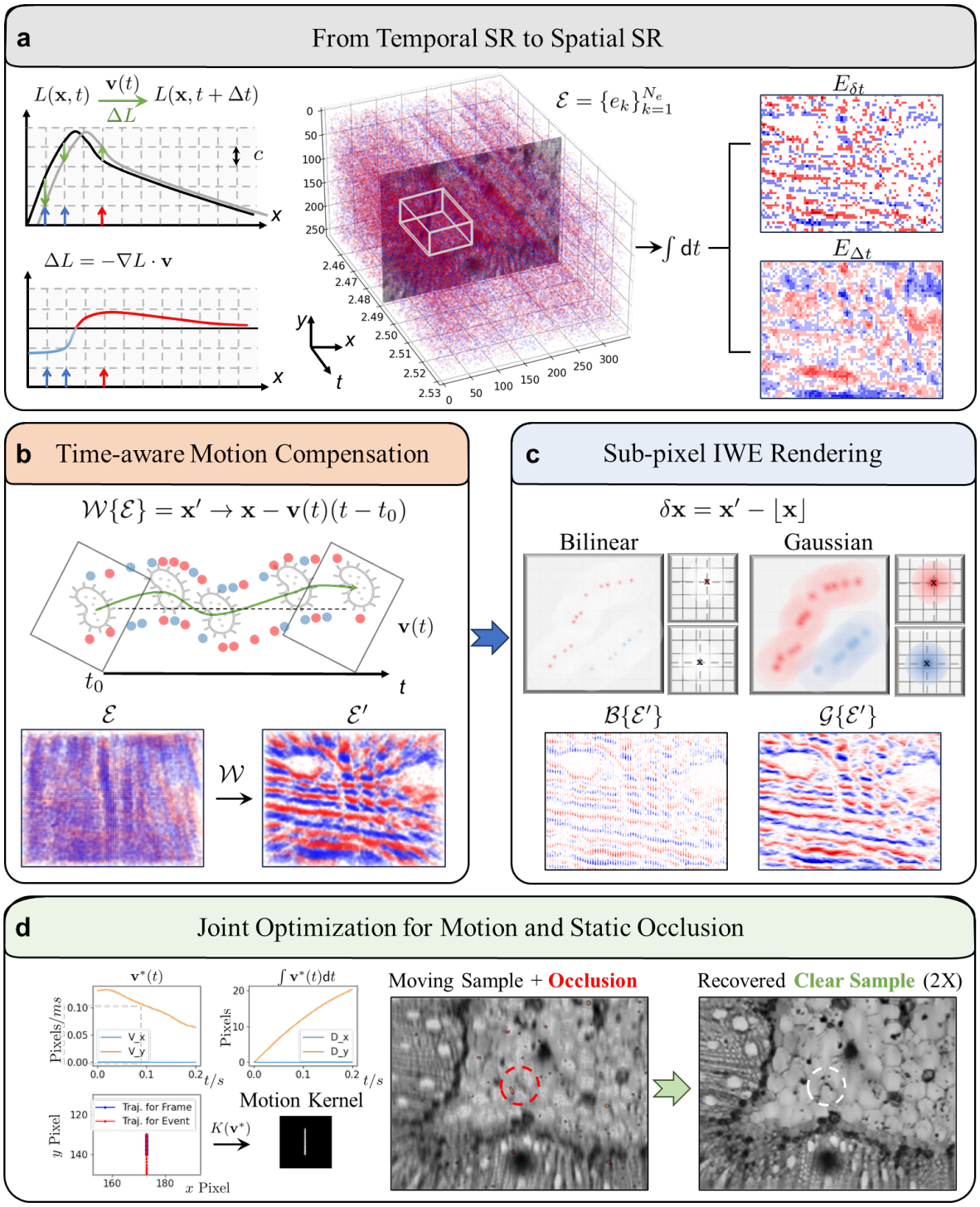

Fig. S5. Event-based pixel reassignment for sub-pixel-resolved IWE rendering. a, Conceptual illustration of how motion-induced temporal variations encoded by events are converted into spatially resolved information through integration. **b,** Motion-compensated warping of events to a common reference time using a time-varying flow model. **c,** Sub-pixel IWE construction, comparing discrete interpolation and continuous kernel-based rendering. **d,** Joint refinement of motion, occlusion and intensity enables deblurring and defect separation.