Spatiotemporal Flux Probing for Single-Photon Videography

Jerry Yan[1*], Matteo Forlivesi[1*], Bowen Tan[1], Andrew Xie[2], Siddharth Somasundaram[3], and Sotiris Nousias[1]

[1] Dept. of Computer Science, Purdue University
[2] Dept. of Computer Science, University of Toronto
[3] MIT Media Lab, Massachusetts Institute of Technology
https://jerukan.github.io/spt-flux-probing

Abstract. We address the problem of recovering high-speed videos from dynamic scenes under extreme photon sparsity. Existing methods rely on aggregating photon detections in local spatiotemporal windows to improve signal-to-noise ratio; however, this local grouping discards global structure and fails in low-light regimes where photon detections are sparse in space and time. In this work, we show that the information needed to recover both motion and illumination is encoded in correlations over the full space-time pattern of photon arrivals. Building on this insight, we develop a spatiotemporal flux probing theory and an algorithm that estimates the Fourier coefficients of the underlying intensity directly from the photon stream. We demonstrate that our approach (1) recovers fast motion and temporal illumination dynamics with substantially fewer photons than prior methods, (2) enables velocity-selective videography that automatically refocuses video onto specific detected motions, and (3) generalizes across sensing modalities including single-photon, event, and spike cameras.

Keywords: High-Speed Videography · Single-Photon Imaging · Low-Light Imaging · Event Cameras

1 Introduction

High-speed videography is fundamentally limited by photon scarcity: as exposure windows shrink to capture rapid motion or illumination changes, photon detections become extremely sparse in both space and time, making reliable intensity estimation difficult. Single-photon cameras provide a different imaging paradigm by recording individual photon arrivals with precise timestamps, allowing photon detections to be flexibly aggregated after acquisition [30,33]. The key challenge is how to combine these sparse, asynchronous detections to recover a rapidly varying intensity.

Existing approaches for reconstructing high-speed video from single-photon data largely follow two strategies. Per-pixel temporal methods analyze each pixel

* Equal contribution: {yan569,mforlive}@purdue.edu

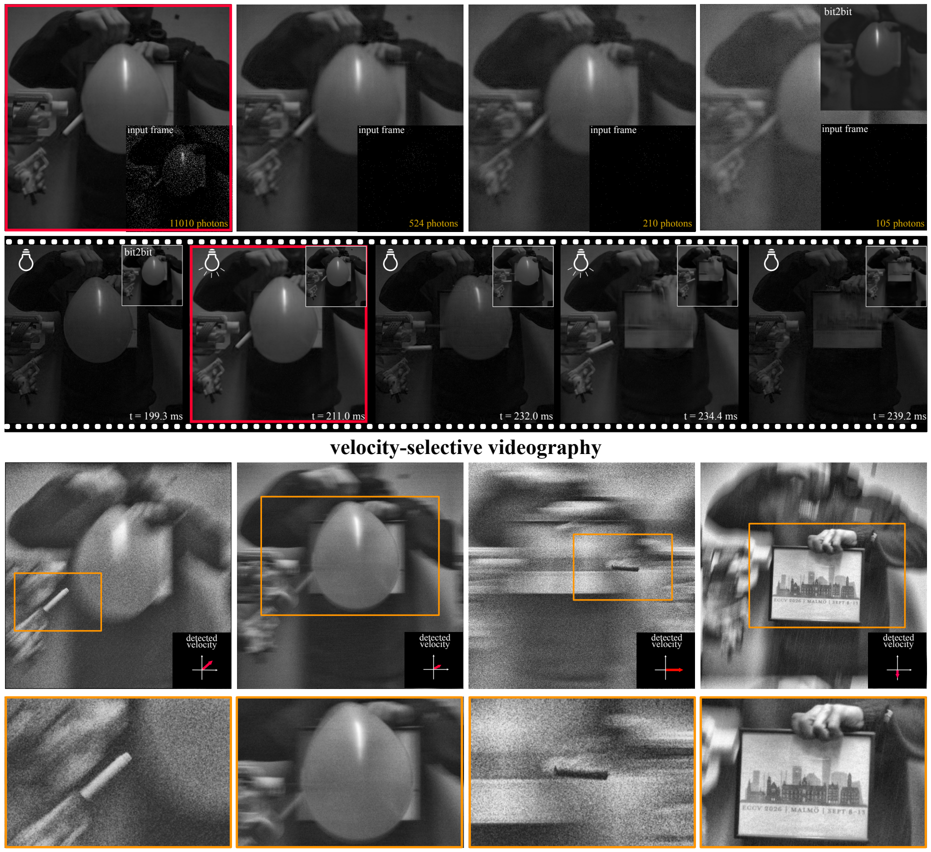


Fig. 1: Top: High-speed videography under extreme photon sparsity. Our method reconstructs high-quality 100 kfps video from binary photon detections or timestamp data, recovering both periodic illumination flicker (indicated by the lightbulb icons) and fast nonperiodic events (speeding bullets, balloon rupture) simultaneously. **Bottom:** Velocity-selective videography. Our method detects the velocities of moving objects directly from the photon stream (arrows indicate magnitude and direction) and reconstructs video focused on a chosen velocity, blurring all other motion.

independently, recovering time-varying intensity from photon arrival times [16, 38, 52]. Spatiotemporal pooling methods [25, 28, 30, 35, 48] improve photon efficiency by aggregating detections across local space-time neighborhoods; for example, collecting photons across motion trajectories of individual scene points [30]. However, both classes share the same fundamental premise: reconstruction begins by explicitly grouping photons locally in space, in time, or both. By confining estimation to local regions, existing methods discard long-range spatiotemporal correlations. Under extreme photon scarcity, these local windows may contain too few photons to reliably estimate intensity.

In this work, we recover high-speed video from sparse photon data in regimes where photon detections arrive far apart in both space and time and local reconstruction methods start to break down. Our central insight is that, even under extreme photon scarcity, the information needed to recover motion and illumination is encoded in correlations across the full space-time pattern of photon arrivals. The key challenge is how to extract these correlations from sparse, asynchronous photon streams. To tackle this challenge, we develop a spatiotemporal flux probing theory that establishes a direct relation between the photon detections in space-time and the Fourier coefficients of the underlying intensity. Instead of aggregating photons within predefined neighborhoods, our approach estimates global spatiotemporal structure jointly from all detections.

Our work brings together two previously distinct regimes: per-pixel ultra-wideband temporal recovery [52], and spatial pooling for photon efficiency [30, 35]. Rather than using local space-time neighborhoods to estimate intensity, we leverage photon detections in a global spatiotemporal volume. This enables coherent aggregation of photons across space and time, increasing the effective signal-to-noise ratio (SNR). As a result, fewer photons are required to achieve a given reconstruction quality, allowing us to operate in regimes that are particularly challenging for existing methods: scenes exhibiting simultaneous fast motion and rapid illumination modulation under extreme photon sparsity in both space and time. Overall, we make the following contributions:

- We develop a spatiotemporal flux probing theory that leverages photon detections across the full space-time observation volume, enabling joint recovery of fast motion and high-frequency illumination modulation directly from extremely sparse photon streams (Fig. 1, top).
- We introduce velocity-selective videography: by identifying motion-specific structure directly in the photon stream, we detect linear object velocities and reconstruct videos refocused at a chosen velocity while suppressing others (Fig. 1, bottom).
- We show that our theory extends across sensing modalities—including quanta sensors, asynchronous photon timestamp streams, spike cameras, and event cameras—suggesting a unified framework for reconstructing dynamic scenes from sparse asynchronous measurements.

2 Single-Photon Videography

Spatiotemporal flux function. The goal of videography is to recover the spatiotemporal flux $\phi(x, y, t)$ [2] of light incident on the sensor. The flux is the instantaneous photon arrival rate per unit area and time. We assume that $\phi(x, y, t)$ has support over the spatiotemporal volume $(x, y, t) \in \Omega = [0, w] \times [0, h] \times [0, t_{\text{exp}}]$ defined by the bounded width w and height h of the sensor plane and the finite exposure time t_{exp}.

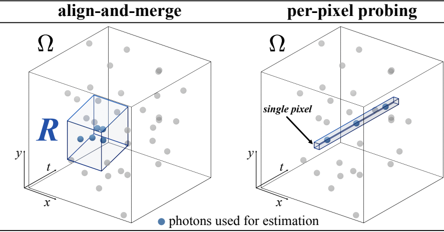
Stream of photon detections. Photon arrivals at the sensor are governed by a 3D inhomogeneous Poisson process with rate function $\phi(x, y, t)$ [10, 39]. This process generates a stream of photon detection events $\mathcal{P} = \{(x_i, y_i, t_i)\}_{i=1}^{N}$, from which we seek to recover $\phi(x, y, t)$. Due to recent advances in single-photon avalanche diode (SPAD) sensors, it is now feasible to obtain such event-based photon representations at up to picosecond temporal [18, 27, 53] and megapixel spatial resolution [33].

Flux constancy and local grouping. Conventional image sensors assume that $\phi(x, y, t)$ is locally constant within a bounded region of space and time—such as in a pixel area during the exposure. Under this assumption, nonoverlapping spatiotemporal volumes can be modeled as independent homogeneous Poisson processes. The flux in each region can then be estimated either by photon counting [30,43,46,47], where the number of detections scales linearly with the flux, or by inter-photon imaging [16,17,22,45], where the mean interval between photon arrivals is inversely proportional to the flux. Incorrectly assuming flux constancy in regions where the flux is changing results in motion blur or temporal aliasing.

Videography by motion compensation. In dynamic scenes, flux is not constant within a fixed pixel because scene points move over time. Motion compensation aggregates photons along motion-aligned space-time trajectories; in this warped coordinate system, detections along each trajectory approximate a homogeneous Poisson process, enabling flux estimation using a homogeneous Poisson model. Learning-based methods adopt similar strategies (either implicitly or through explicit alignment) [3, 4, 11, 20, 48, 54]. However, this is still a locally confined formulation—now along trajectories rather than pixels. In our regime of extreme sparsity with simultaneous fast motion and temporal modulation, trajectories may contain too few detections for reliable estimation, and appearance changes along the path make motion recovery unreliable.

Videography by denoising. Another common approach is to treat reconstruction as a denoising problem [24, 28]. However, these methods still rely on local flux smoothness. In practice, we observe that denoising priors prioritize spatial consistency, effectively smoothing temporal variation to stabilize reconstruction (Fig. 1, top). We provide further analysis in supplement Section E.

Temporal modulation from ultrafast illumination. In the presence of ultrafast illumination, such as pulsed lasers and fluorescent light bulbs, the scene radiance can experience temporal modulation at very high frequencies ranging from 10 kHz to 60 MHz [1, 12, 29, 50]. While this modulation can be useful [23, 35], recovery of such time-varying radiance is highly challenging because flux constancy breaks down at time scales much faster than the inter-photon arrival time. To tackle this problem, prior work recovers the time-varying flux of a pixel by estimating its Fourier coefficients from the photon timestamp realizations [52]. However, this method operates on each pixel independently, ignoring spatial correlations. As a result, these techniques fail when there are insufficient photons at each pixel, and spatial pooling is required to increase the signal-to-noise ratio [25, 35]. We refer to supplement Section F for a detailed analysis.



	align-and-merge	per-pixel probing	proposed
probing function	$\mathbb{1}_R(x, y, t)$	$\mathbb{1}_{\text{pixel}}(x, y) \cdot p(t)$	$p(x, y, t)$
spatial support	local (region ***R***)	local (single pixel)	global
temporal support	local (region ***R***)	global	global

Fig. 2: Spatiotemporal support of existing and proposed methods. Align-and-merge methods [30] aggregate photons within a local region R, and per-pixel probing methods [52] fix the spatial support to a single pixel while applying an arbitrary 1D temporal probe $p(t)$ to estimate its time-varying flux. Our framework admits a broader family of spatiotemporal probing functions $p(x, y, t)$ that act on every photon globally.

The gap: simultaneous ultrafast motion and illumination. We consider the imaging regime in which $\phi(x, y, t)$ varies rapidly due to both fast scene motion *and* high-frequency illumination modulation. In this setting, the two classes of methods described above each fail in complementary ways. Methods that rely on flux constancy cannot account for temporal intensity variations along a scene point's trajectory. Conversely, per-pixel temporal methods that can handle such modulation fail to account for scene motion that extends beyond a single pixel's footprint. As a scene point traverses a pixel, the observed temporal flux approximates a rect function with very high temporal bandwidth; in the low-photon-count regime, recovering such signals from few detections is severely ill-posed for pixel-wise methods [16, 43, 52]. Both failures trace back to the same root cause: *local spatiotemporal grouping of photons* (Fig. 2).

3 Spatiotemporal Flux Probing

Spatiotemporal flux variations arise from two primary sources: temporal modulation of external illumination [52] and natural scene motion [30]. Each produces a distinct signature in the underlying flux profile $\phi(x, y, t)$. Although we cannot directly observe $\phi(x, y, t)$, photon arrivals implicitly encode this structure. Here, we describe our method to extract this structure directly from photons, requiring only that $\phi(x, y, t)$ is well-behaved—*i.e.*, nonnegative, bounded, and measurable (see supplement Section A). We consider the low-light regime, where photon detections are modeled by Poisson statistics.

Spatiotemporal flux probing. To recover different scene phenomena, we leverage spatiotemporal probing functions. A *probing function* $p(x, y, t)$ is a deterministic function whose inner product with $\phi(x, y, t)$ informs the presence or absence of particular physical phenomena. However, this inner product—the *flux probing integral*—cannot be computed directly because $\phi(x, y, t)$ is unknown. We prove in supplement Section B.2 that this projection can be estimated from the set of photon detections $\mathcal{P}$ through the probing measurements

$$\underbrace{\sum_{(x_i,y_i,t_i)\in\mathcal{P}} p(x_i, y_i, t_i)}_{\text{probing measurements}} = \underbrace{\int_0^{t_{\text{exp}}} \int_0^{h} \int_0^{w} p(x, y, t)\, \phi(x, y, t)\, dx\, dy\, dt}_{\text{flux probing integral } \langle p,\phi\rangle} + \underbrace{M_p(t_{\text{exp}})}_{\text{martingale noise}} . \tag{1}$$

Intuitively, the probing measurements indicate how correlated the photon arrivals are with the probing function.

Noise model. Approximation of the flux probing integral through probing is a noisy process due to randomness in photon arrivals [10]. Equation (1) decomposes the probing measurements into a deterministic integral and martingale noise [5], which is a continuous stochastic process [7] caused by photon shot noise. We show in supplement Section B.3 that the probing measurements are approximately normal[4] with mean equal to $\langle p, \phi\rangle$ and variance $\langle p^2, \phi\rangle$. As a result, probing measurements are an unbiased estimate of the flux probing integral. Characterization of the noise model enables statistical detection of the phenomena represented by different probing functions. We also derive from first principles the decomposition in Eq. (1) for spatiotemporal flux functions, extending prior work that considered temporal-only flux variations [52].

Spatiotemporal flux probing as a unifying lens. As illustrated in Fig. 2, existing nonlearning methods can be thought of as special cases of spatiotemporal flux probing, differing only in their probing function and its support. Rather than fixing a local support or a restricted function class, our framework admits any bounded 3D probing function over the full spatiotemporal volume, so that every detected photon contributes jointly to estimation.

3.1 Spatiotemporal Fourier Basis Functions

While our underlying probing theory is general, we consider the special case of 3D Fourier basis functions due to their inherent ability to simultaneously encode scene motion and illumination flicker. Specifically, a periodically flickering light source induces a frequency comb along the temporal frequency domain [35, 52], whereas local linear motion of a bounded object generates a planar structure in the spatiotemporal frequency domain, with the plane's slope encoding object velocity [49]. Detecting these physical phenomena corresponds to detecting the spatiotemporal frequencies $\mathbf{f}$ associated with those phenomena.

Probing with Fourier basis. A 3D Fourier basis function has the form $p_{\mathbf{f}}(\mathbf{x}) = e^{-j2\pi\mathbf{f}^\top\mathbf{x}}$, where $\mathbf{x} = [x, y, t]^\top$ is a point in space-time and $\mathbf{f} = [f_x, f_y, f_t]^\top$ is a

[4] This approximation holds with as few as 20 photons (supplement Section I.1).

spatiotemporal frequency. When probing with a single frequency $\mathbf{f}$ of the Fourier basis, the resulting probing measurements

$$\mathcal{E}_{\mathbf{f}} \triangleq \sum_{\mathbf{x}\in\mathcal{P}} e^{-j2\pi\mathbf{f}^\top\mathbf{x}} \tag{2}$$

correspond to a phasor vector summation. If $\phi(x, y, t)$ contains frequency $\mathbf{f}$, the phasors add coherently and $|\mathcal{E}_{\mathbf{f}}|$ is large; if $\mathbf{f}$ is absent, photon arrivals are uncorrelated with the complex sinusoid, the phasors cancel out, and $|\mathcal{E}_{\mathbf{f}}|$ is small. However, due to the stochastic nature of photon arrivals, $|\mathcal{E}_{\mathbf{f}}|$ has nonzero fluctuations even if $\mathbf{f}$ is absent. To detect a frequency $\mathbf{f}$, the probing energy $|\mathcal{E}_{\mathbf{f}}|$ must be larger than the energy predicted by these noisy fluctuations.

Distribution of probing energy. For 3D Fourier basis functions, we show in supplement Section B.4 that the probing measurements $\mathcal{E}_{\mathbf{f}}$ approximately follow a complex normal distribution with mean and covariance

$$\boldsymbol{\mu} = \begin{bmatrix} \langle\cos(2\pi\mathbf{f}^\top\mathbf{x}), \phi(\mathbf{x})\rangle \\ \langle-\sin(2\pi\mathbf{f}^\top\mathbf{x}), \phi(\mathbf{x})\rangle \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \langle\cos^2(2\pi\mathbf{f}^\top\mathbf{x}), \phi(\mathbf{x})\rangle & 0 \\ 0 & \langle\sin^2(2\pi\mathbf{f}^\top\mathbf{x}), \phi(\mathbf{x})\rangle \end{bmatrix}. \tag{3}$$

The normalized probing energy

$$|\widetilde{\mathcal{E}}_{\mathbf{f}}|^2 \triangleq \mathrm{Re}\left(\frac{\mathcal{E}_{\mathbf{f}}}{\sqrt{\boldsymbol{\Sigma}_{1,1}}}\right)^2 + \mathrm{Im}\left(\frac{\mathcal{E}_{\mathbf{f}}}{\sqrt{\boldsymbol{\Sigma}_{2,2}}}\right)^2 \tag{4}$$

follows a noncentral χ^2 distribution with two degrees of freedom, with noncentrality parameters determined by $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

Frequency detection. We use normalized Fourier probing functions $p_{\mathbf{f}}(\mathbf{x}) = \frac{1}{\sqrt{v}} e^{-j2\pi\mathbf{f}^\top\mathbf{x}}$ such that $p_{\mathbf{f}}$ is orthonormal over the spatiotemporal volume Ω, where $v = |\Omega| = hwt_{\text{exp}}$ represents the observation volume of photons. If frequency $\mathbf{f}$ is absent from the flux, the probing measurement $\mathcal{E}_{\mathbf{f}}$ has zero mean. We can express this as a constant false alarm rate (CFAR) detector [40], where a frequency $\mathbf{f}$ is detected if $|\widetilde{\mathcal{E}}_{\mathbf{f}}|^2 \geq \mathrm{CDF}^{-1}_{\chi^2_2}(1-\alpha)$. Here, α denotes the significance value and false alarm probability. We show in supplement Section B.5 that this relation can be written in terms of $|\mathcal{E}_{\mathbf{f}}|^2$ and critical value c_α as follows:

$$|\mathcal{E}_{\mathbf{f}}|^2 \geq \underbrace{\mathrm{CDF}^{-1}_{\chi^2_2}(1-\alpha)\frac{|\mathcal{P}|}{2v}}_{c_\alpha}. \tag{5}$$

3.2 Velocity Detection

Because linear motion induces planar structure in the spatiotemporal spectrum, we can use the probed spatiotemporal frequencies for motion analysis [49]. Consider a local patch whose intensity I translates with approximately constant velocity $\mathbf{v} = [v_x, v_y]^\top$, such that $\phi(x, y, t) = I(x - v_x t, y - v_y t)$. Here, the spectral energy of $\phi(x, y, t)$ due to the patch motion is confined to the plane

$$v_x f_x + v_y f_y + f_t = 0 \tag{6}$$

in the spatiotemporal frequency domain [49]. Thus, estimating linear motion reduces to identifying planar concentrations of energy in 3D frequency space.

We formulate linear velocity detection as a plane-consistency problem. First, we scan a discrete set of M velocity hypotheses $\{\mathbf{v}_i\}_{i=1}^M$. For each hypothesis, we measure the support of the corresponding velocity plane by accumulating the energy of the detected frequencies lying near that plane

$$E_i \triangleq \sum_{\mathbf{f}\,:\,|\mathbf{v}_i'^\top \mathbf{f}| \le \epsilon} |\mathcal{E}_\mathbf{f}|^2, \tag{7}$$

where $\mathbf{v}' = [v_x, v_y, 1]^\top$ and ϵ controls the tolerance to deviations from planarity. Then, velocity planes are detected based on their accumulated energies. Unlike the probing measurement, the plane scores do not follow a simple parametric distribution because (1) the cumulative energy aggregates the energy across many different frequencies, and (2) neighboring velocity hypotheses have overlapping frequencies, making their energies correlated. As a result, we use a nonparametric rank-CFAR test [13] to detect velocity planes. For each $\mathbf{v}_i$, we compute its rank $r_i \triangleq |\{j \in \mathcal{N}_i : E_j < E_i\}|$ among the $|\mathcal{N}_i|$ neighboring hypotheses, and detect $\mathbf{v}_i$ only if E_i falls in the top α_{vel} fraction,

$$r_i \ \ge\ \lceil (1 - \alpha_{\text{vel}})(|\mathcal{N}_i| + 1) \rceil, \tag{8}$$

which targets a false alarm rate α_{vel}. See supplement Section C for derivations.

3.3 Extension to Other Asynchronous Sensors

Our probing theory can be extended beyond single-photon sensors to other asynchronous sensors whose events stochastically encode a continuous intensity signal. The only requirement is that the image formation model and noise properties of the sensor must be known. We discuss two cases here—spike and event cameras—and defer derivations to supplement Section D.

Spike cameras. Spike cameras output an event when the accumulated flux at a pixel $\int \phi(t)dt$ crosses a fixed threshold [15]. Therefore, the instantaneous spike rate is proportional to the flux $\phi(x, y, t)$ and spatiotemporal flux probing with the spike events yields a noisy estimate of the spatiotemporal Fourier coefficients of $\phi(x, y, t)$, up to a scale factor.

Event cameras. Event cameras output an event with polarity $\sigma = \pm 1$ when the change in log-intensity $\log \phi(x, y, t)$ at a pixel exceeds a contrast threshold [26]. Fourier probing on these events $(x_i, y_i, t_i) \in \mathcal{P}$ is done via the following sum

$$\mathcal{E}_\mathbf{f} = \sum_{(x_i, y_i, t_i) \in \mathcal{P}} \sigma_i p_\mathbf{f}(x_i, y_i, t_i). \tag{9}$$

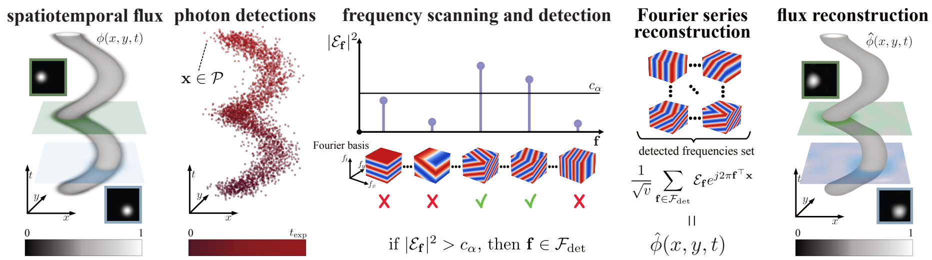


Fig. 3: Overview of spatiotemporal flux probing. *Spatiotemporal flux:* The goal of videography is to recover the spatiotemporal flux function $\phi(x, y, t)$ (insets visualize two time slices). *Photon detections:* The observed photon detections $\mathcal{P}$ are a stochastic realization of an inhomogeneous Poisson process with rate $\phi(x, y, t)$. *Frequency scanning and detection:* We compute 3D Fourier probing measurements on $\mathcal{P}$ via Eq. (2) and declare $\mathbf{f}$ detected (in $\mathcal{F}_{\text{det}}$) if its probing energy $|\mathcal{E}_\mathbf{f}|^2$ exceeds c_α (set for a target false alarm rate α). *Fourier series reconstruction:* We obtain $\hat{\phi}(x, y, t)$ by performing an inverse Fourier transform on the set of detected frequencies $\mathcal{F}_{\text{det}}$.

This yields a noisy estimate of the Fourier coefficients of $\partial_t \log \phi(x, y, t)$ up to the scale of contrast threshold C.

4 Videography by Spatiotemporal Flux Probing

Our theory enables videography across asynchronous modalities such as single-photon, spike, and event cameras. We discuss the algorithms for each modality, with full implementation details in supplement Section G.

Algorithm 1 Flux reconstruction

procedure FluxRecon($\mathcal{P}, v, \alpha$)
 // **Frequency scanning**
 $\mathcal{F} \leftarrow \{\mathbf{f}_i\}_{i=1}^K$ up to Nyquist ▷ Supp. G.1.1
 for $\mathbf{f} \in \mathcal{F}$ **do**
  $\mathcal{E}_\mathbf{f} \leftarrow \frac{1}{\sqrt{v}} \sum_{\mathbf{x}\in\mathcal{P}} e^{-j2\pi\mathbf{f}^\top\mathbf{x}}$ ▷ Eq. (2)
 end for
 // **CFAR Detection**
 $c_\alpha \leftarrow \mathrm{CDF}^{-1}_{\chi^2_2}(1-\alpha) \cdot \frac{|\mathcal{P}|}{2v}$ ▷ Eq. (5)
 $\mathcal{F}_{\text{det}} \leftarrow \{\mathbf{f} \mid |\mathcal{E}_\mathbf{f}|^2 \geq c_\alpha\}$
 // **Reconstruction**
 $\hat{\phi}(\mathbf{x}) \leftarrow \frac{1}{\sqrt{v}} \sum_{\mathbf{f}\in\mathcal{F}_{\text{det}}} \mathcal{E}_\mathbf{f}\, e^{j2\pi\mathbf{f}^\top\mathbf{x}}$
 return $\hat{\phi}$
end procedure

Single-photon videography. Using our probing theory from Section 3.1, we can recover the spatiotemporal flux $\phi(x, y, t)$ directly from a set of photon detections $\mathcal{P}$ in three steps, summarized in Algorithm 1. First, we estimate $\mathcal{E}_\mathbf{f}$ from Eq. (2) for all frequencies in a discrete spatiotemporal Fourier basis $\mathcal{F} = \{\mathbf{f}_1, \ldots, \mathbf{f}_K\}$— restricted to those frequencies recoverable under the given spatial and temporal sampling rates—and obtain the corresponding probing coefficients $\{\mathcal{E}_{\mathbf{f}_1}, \ldots, \mathcal{E}_{\mathbf{f}_K}\}$. Second, we apply the CFAR detection rule in Eq. (5) to identify the set of detected frequencies $\mathcal{F}_{\text{det}}$. Finally, we recover the spatiotemporal flux by Fourier series reconstruction over the detected frequencies. The reconstructed flux can then be resampled at arbitrary temporal resolutions. An illustration of the pipeline is shown in Fig. 3.

Velocity-selective videography. Because we have an explicit mechanism for velocity detection (Section 3.2), we can reconstruct the video to focus on a particular linear velocity while blurring away others. This "velocity focusing" is performed directly in the spatiotemporal frequency domain. Given a detected velocity $\mathbf{v}_{\text{det}}$, we focus on that linear motion by shearing the 3D Fourier spectrum via the change of variables $f_t \leftarrow f_t + \mathbf{v}_{\text{det}}^\top \mathbf{f}_{x,y}$, with $\mathbf{f}_{x,y} = [f_x, f_y]^\top$. This shear maps the corresponding velocity plane to the temporal DC component, making linear motion at velocity $\mathbf{v}_{\text{det}}$ appear static in the reconstructed volume. Reconstruction is then performed in a sliding temporal window over the photon stream to generate video frames. Although related ideas have been explored in conventional videography [49] and single-photon imaging [46], our method performs velocity focusing at ultra-low light levels and without requiring manual specification of the motion direction or magnitude as in previous work [46].

Spike camera videography. We compute the probing measurements from the spike stream and estimate the Fourier coefficients of the flux. However, in regions of constant flux, spike streams generate highly periodic spike trains, which introduce strong artificial high-frequency peaks. To suppress these spurious components, we apply a low-pass filter in addition to CFAR detection.

Event camera videography. The probing measurements of an event stream yield the Fourier coefficients of the temporal derivative of the log flux, $\partial_t \log \phi(x, y, t)$. To recover the log-intensity, we integrate this estimate of the derivative in the frequency domain by dividing the spectrum by $j2\pi f_t$ for $f_t \neq 0$. In practice, low-frequency temporal components are generally dominated by bias and drift [42], so a high-pass filter is applied in addition to CFAR detection. The resulting reconstruction is determined up to an additive constant in log-intensity. This constant and lost low-frequency information can be obtained from intensity frames recorded alongside the events, which we fuse together using a Kalman filter [21].

5 Results

We validate our theory on real captures from multiple sensing modalities. Our experiments demonstrate four capabilities: (1) high-speed video recovery of dynamic scenes with simultaneous temporal flicker under extreme photon sparsity, a regime in which existing methods struggle, (2) velocity-selective videography, where detected velocity planes enable reconstruction focused on individual moving objects, (3) ultra-wideband videography in the GHz regime, achieved at 30× lower photon budgets than previous approaches, and (4) extension of our spatiotemporal probing framework to event and spike cameras. We provide additional results and videos in the supplementary materials. Supplement Sections H and I cover further experimental details and simulated-data results, respectively.

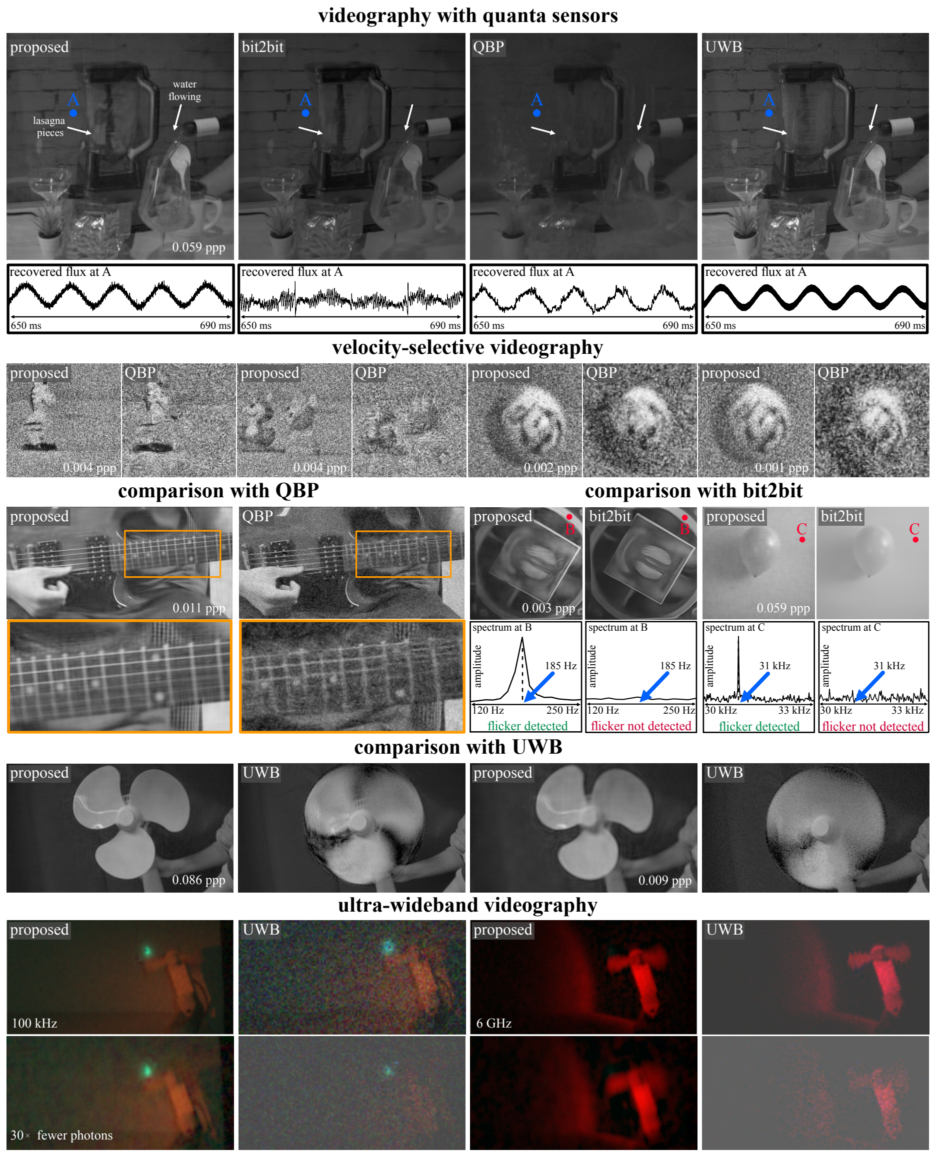


Fig. 4: Spatiotemporal flux probing results. **Row 1**: Videography with quanta sensors. We reconstruct 100 kfps video of a scene with complex dynamics under 120 Hz illumination flicker. Prior methods fail to recover motion and flicker simultaneously; ours captures both. **Row 2:** Velocity-selective videography. Our method detects the velocities of multiple objects directly from the photon stream and produces higher-quality reconstructions than QBP [30]. **Row 3:** Comparison with QBP (left) and bit2bit [28] (right) on their respective datasets. Our method produces higher-quality reconstructions while simultaneously recovering the illumination flicker. **Row 4:** Comparison with UWB. Even in periodic motion scenarios favorable to UWB [52], our approach achieves higher-quality reconstructions at two different light levels. **Row 5:** Ultra-wideband videography. Our approach can recover the fan blade and the laser wavefront even with 30× fewer photons than previous work [52].

Sensor hardware. We use multiple sensing modalities—both synchronous and asynchronous—to validate our theory. For synchronous capture, we use a 512×512-pixel SPAD512 camera [37] operating at 100 kHz, which outputs binary frames where a value of 1 indicates a photon detection. For this modality, we report light levels in photons per pixel (ppp) [30]. For asynchronous sensing, we use the dataset from Wei et al. [52], acquired with a single-pixel SPAD providing 68 ps timing resolution and outputting a stream of photon timestamps. We additionally evaluate our framework on data from a DAVIS240C event camera and a spike camera using publicly available datasets [34, 55].

Baselines. We compare against several state-of-the-art techniques in high-speed single-photon imaging. *Ultra-wideband imaging* (UWB) recovers temporal modulations at a pixel through flux probing in the time dimension only [52]. *Quanta burst photography (QBP)* uses an "align-and-merge" strategy to aggregate photons along motion trajectories of scene points [30]. *bit2bit* is a self-supervised denoising technique that leverages the inductive smoothness bias of neural networks [28]. Of these techniques, only UWB and our method are able to handle both asynchronous and synchronous photon streams. QBP and bit2bit can only handle synchronous frame data. Furthermore, these techniques produce qualitatively inferior results compared to our method, due to their limited ability to leverage global spatiotemporal photon correlations.

High-speed videography under extreme photon sparsity. We demonstrate 100 kfps videography of two foam bullets striking and rupturing a balloon captured with a SPAD512 camera (Fig. 1). The scene is illuminated by a 120 Hz ceiling light and a 31 kHz modulated LED. Our method reconstructs the projectile trajectories, the balloon rupture, and both illumination flickers simultaneously. A visualization of the 31 kHz flicker is provided in supplement Section H.1.1. Even under extreme photon sparsity, our approach continues to recover the white bullet trajectory and the ambient illumination flicker, demonstrating robust high-speed videography in photon-starved conditions.

We further reconstruct a 100 kfps video of a scene in which lasagna is blended in the background while water is poured in the foreground under 120 Hz AC illumination. Despite photon-starved conditions, our method recovers fine spatial details of the flowing water and blender contents while faithfully capturing the global 120 Hz flicker (Fig. 4, row 1). In contrast, methods based on local neighborhoods either smooth temporal modulation into spatial structure (*e.g.*, bit2bit) or produce unstable modulation (*e.g.*, QBP), while per-pixel approaches (*e.g.*, UWB) fail to capture the dynamic motion.

Velocity-selective videography. We apply our velocity detection (Section 3.2) to the scene shown in Fig. 1. Our method detects the velocities of multiple objects in the scene and reconstructs separate videos focused on each (Fig. 1, bottom), even for the gray high-speed foam bullet. To our knowledge, this is the first demonstration of post-capture velocity-selective video reconstruction from a single-photon stream with automatic velocity detection.

We evaluate velocity detection on the toys dataset from [30]. We synthetically thin the timestamps to simulate extreme low-light conditions. Even under severe thinning, our method reliably detects object velocities and produces higher-quality reconstructions than QBP (Fig. 4, row 2, left). In this regime, local photon accumulation becomes unreliable, whereas velocity-aware aggregation provides a strong structural prior. We also captured a scene of falling balls under AC illumination flicker, introducing simultaneous motion and high-frequency modulation. Our method accurately detects the ball's velocity and reconstructs fine spatial detail despite extreme sparsity and flicker. In contrast, QBP blurs temporally varying intensity (Fig. 4, row 2, right).

Comparison with baselines. We evaluate our method against QBP on the guitar dataset [30] (Fig. 4, row 3, left). Our method is able to recover the fine features of the guitar strings at 0.011 ppp, whereas QBP produces noisy reconstructions. At this level of photon sparsity, local temporal aggregation cannot improve intensity estimates since there are not enough photons to accumulate.

We compare with bit2bit on the Mandrill dataset [28]. Our approach resolves finer spatial details (Fig. 4, row 3, right) and faithfully recovers the faint CPU fan flicker present in the data. In contrast, bit2bit oversmooths both spatial texture and temporal variation, failing to capture the subtle flicker. We additionally captured a scene of a static balloon illuminated by a localized 31 kHz sinusoidally modulated LED. Our method successfully recovers the high-frequency modulation, whereas bit2bit suppresses it as noise.

Finally, using the SPAD512 camera, we captured a scene of a fan rotating at approximately 15 Hz and compare against UWB. Although the motion is periodic—typically a favorable setting for UWB's periodic priors—the rotating blades induce high temporal bandwidth in the flux at each pixel. Under extreme photon sparsity, UWB fails to recover this bandwidth due to the low per-pixel SNR and cannot reconstruct the blade structure (Fig. 4, row 4). In contrast, our method accurately recovers both the fan blades and their specular reflections, even with $10\times$ fewer detected photons, where UWB completely breaks down. This experiment demonstrates the advantage of estimating global spatiotemporal structure rather than treating each pixel's temporal signal independently.

Ultra-wideband videography. We evaluate our method on the same dataset used by Wei et al. [52], where a diffused laser pulse illuminates a scene containing a rotating fan. Our approach successfully reconstructs both the laser wavefront and the spinning fan blades (Fig. 4, row 5, top). In contrast, UWB blurs the fan motion at lower light levels due to the limited number of photons per pixel. Notably, our method yields high-quality reconstructions even with $30\times$ less light (Fig. 4, row 5, bottom), demonstrating the advantage of exploiting global spatiotemporal correlations for reconstructing ultra-wideband videos.

Videography with asynchronous sensors. We further demonstrate videography results on event and spike cameras. For event cameras, we evaluate on the dataset from [34] captured with a DAVIS240C sensor. Direct integration of events alone

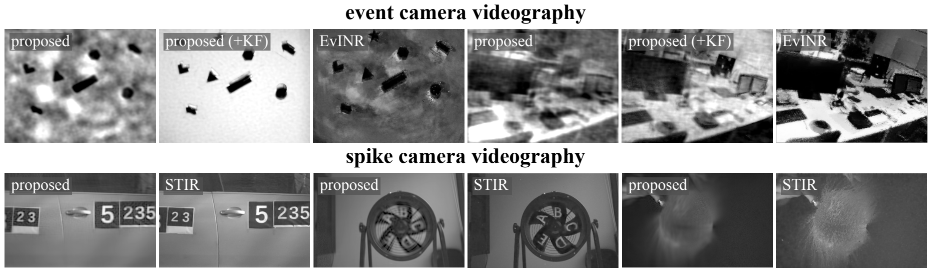


Fig. 5: Videography with asynchronous sensors. **Top:** Our method reconstructs 1 kfps video from raw event streams, achieving competitive results with the learning-based method EvINR [51]. **Bottom:** We reconstruct 10 kfps video from binary spike streams with quality on par with STIR [8].

produces blurred reconstructions due to accumulated drift and the lack of absolute intensity information. By combining our frequency-domain estimation of $\partial_t \log \phi(x, y, t)$ with an intensity image via Kalman filtering [21], we obtain reconstructions that preserve both spatial detail and temporal dynamics (Fig. 5, top). The resulting quality is comparable to EvINR [51], a state-of-the-art learning-based method for event-based videography. We also evaluate our method on the spike camera dataset from [55], covering scenes with fast (moving car), periodic (fan spinning), and highly nonperiodic motion (water balloon rupture) (Fig. 5, bottom). Our reconstructions remain visually consistent and detailed across these varied dynamics. For reference, we compare against the state-of-the-art learning-based method STIR [8].

Importantly, the core probing formulation transfers across quanta sensors, SPAD timestamp streams, event cameras, and spike cameras, while the image-formation models and reconstruction steps remain sensor-specific. To our knowledge, no existing method applies across all these sensing modalities within a single unified framework.

6 Conclusion

High-speed videography remains fundamentally limited by photon scarcity; our work does not change that physical constraint. What it changes is how the available photons are used. Existing methods confine estimation to local spatiotemporal supports, discarding information encoded in the broader pattern of photon arrivals. We have shown that the global structure of *where and when* photons arrive carries sufficient information to jointly recover fast motion and rapid illumination dynamics, even in regimes where prior methods degrade. We hope that viewing the full spatiotemporal photon stream as a coherent signal, rather than a collection of isolated detections, will inspire new approaches to imaging under extreme photon sparsity.

6.1 Limitations

Our current implementation has two main limitations. First, because we use global Fourier probing functions, sharp spatial edges and temporal discontinuities can spread energy across Fourier coefficients and produce Gibbs-like ringing artifacts in the reconstruction [9, 14]. This limitation arises from the specific choice of probing basis rather than from our spatiotemporal probing theory itself. Since our theory only requires linear projections of the photon stream, alternative bases with better spatial or temporal localization—such as wavelets [31]—could reduce these artifacts without changing the underlying formulation.

Second, our velocity-selective videography assumes approximately constant linear motion over the reconstruction window. When motion is highly nonlinear, accelerating, or chaotic—for example in a balloon rupture or fluid splash—that concentration breaks down, leading to reduced selectivity and reconstruction blur. Extending the framework to richer motion models, such as higher-order kinematics using polynomial-phase [36] or chirplet [32] probing functions, is an important future direction.

6.2 Future Directions

There are several promising directions for future work. First, our formulation suggests a unified perspective on videography from asynchronous sensors. Whether events correspond to first-photon arrivals, intensity changes, or thresholded flux accumulation, the probing measurements reduce to sums over event timestamps. This provides a common mathematical framework for comparing sensing modalities in terms of the spatiotemporal information they preserve and the reconstruction quality they support under bandwidth constraints [46, 47].

Second, our velocity-selective videography currently recovers image-plane motion, but these estimates could also support geometric inference. Under known camera motion, stereo, or multi-view capture, the recovered spectral shear may provide a route to depth estimation [41] and structure-from-motion [19] directly from sparse photon streams, linking our framework to core problems in dynamic 3D scene understanding in extreme low-light conditions.

Third, our framework may be useful for active and scientific imaging. In particular, the ability to handle fast motion together with high-frequency illumination changes could be valuable for single-photon lidar and Doppler imaging [23]. More broadly, because our method analyzes structure jointly in space and time, it may also be useful for vibration sensing [6, 44], where subtle periodic motions must be recovered from extremely faint intensity signals.

Acknowledgements

We are grateful to Kyros Kutulakos and David Lindell for providing the SPAD512 camera. Kyros Kutulakos acknowledges the support of NSERC under RTI-2023-00153. We also thank the Marine Physical Laboratory at the Scripps Institution of Oceanography for providing access to their compute resources.

References

1. ams OSRAM Group: TMF8820/21/28 Time-of-Flight Sensor Datasheet (2023), https://look.ams-osram.com/m/52236c476132a095/original/TMF8820-21-28-Multizone-Time-of-Flight-Sensor.pdf, Accessed: 2026-06-30
2. Boyd, R.W.: Radiometry and the detection of optical radiation. John Wiley & Sons (1983)
3. Chennuri, P., Chi, Y., Jiang, E., Godaliyadda, G.M.D., Gnanasambandam, A., Sheikh, H.R., Gyongy, I., Chan, S.H.: Quanta Video Restoration. In: Proc. Eur. Conf. on Computer Vision (ECCV). pp. 152–171. Springer (2024)
4. Chennuri, P., Fu, D., Chan, S.H.: Quanta diffusion. In: Proc. IEEE Int. Conf. on Image Processing (ICIP). pp. 229–234 (2025)
5. Daley, D.J., Vere-Jones, D.: An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods. Springer (2003)
6. Davis, A., Rubinstein, M., Wadhwa, N., Mysore, G., Durand, F., Freeman, W.T.: The Visual Microphone: Passive Recovery of Sound from Video. ACM Trans. Graph. **33**(4), 79:1–79:10 (2014)
7. Doob, J.L.: Stochastic Processes. John Wiley & Sons (1991)
8. Fan, B., Yin, J., Dai, Y., Xu, C., Huang, T., Shi, B.: Spatio-Temporal Interactive Learning for Efficient Image Reconstruction of Spiking Cameras. In: Adv. Neural Inf. Process. Syst. (NeurIPS). vol. 37, pp. 21401–21427 (2024)
9. Gonzalez, R.C., Woods, R.E.: Digital Image Processing. Pearson, 4th edn. (2018)
10. Goodman, J.W.: Statistical Optics. John Wiley & Sons (1985)
11. Gupta, S., Gupta, M.: Eulerian single-photon vision. In: Proc. IEEE/CVF Int. Conf. on Computer Vision (ICCV). pp. 10431–10442 (2023)
12. Hampf, D., Schafer, E., Sproll, F., Otsubo, T., Wagner, P., Riede, W.: Satellite laser ranging at 100 kHz pulse repetition rate. CEAS Space J. **11**, 363–370 (2019)
13. Hansen, V.G., Olsen, B.A.: Nonparametric radar extraction using a generalized sign test. IEEE Trans. Aerosp. Electron. Syst. (TAES) **AES-7**(5), 942–950 (1971)
14. Harris, F.J.: On the use of windows for harmonic analysis with the discrete Fourier transform. Proc. IEEE **66**(1), 51–83 (1978)
15. Huang, T., Zheng, Y., Yu, Z., Chen, R., Li, Y., Xiong, R., Ma, L., Zhao, J., Dong, S., Zhu, L., Li, J., Jia, S., Fu, Y., Shi, B., Wu, S., Tian, Y.: 1000$\times$ Faster Camera and Machine Vision with Ordinary Devices. Engineering **25**, 110–119 (2023)
16. Ingle, A., Seets, T., Buttafava, M., Gupta, S., Tosi, A., Gupta, M., Velten, A.: Passive inter-photon imaging. In: Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 8581–8591 (2021)
17. Ingle, A., Velten, A., Gupta, M.: High flux passive imaging with single-photon sensors. In: Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 6753–6762 (2019)
18. Itzler, M.A., Ben-Michael, R., Hsu, C.F., Slomkowski, K., Tosi, A., Cova, S., Zappa, F., Ispasoiu, R.: Single photon avalanche diodes (SPADs) for 1.5 μm photon counting applications. J. Mod. Opt. **54**(2), 283–304 (2007)
19. Jungerman, S., Garg, A., Gupta, M.: Radiance Fields from Photons. ACM Trans. Graph. **45**(1), 11:1–11:16 (2025)
20. Jungerman, S., Ingle, A., Gupta, M.: Panoramas from photons. In: Proc. IEEE/CVF Int. Conf. on Computer Vision (ICCV). pp. 10592–10602 (2023)
21. Kalman, R.E.: A New Approach to Linear Filtering and Prediction Problems. J. Basic Eng. **82**(1), 35–45 (1960)

22. Kirmani, A., Venkatraman, D., Shin, D., Colaço, A., Wong, F.N., Shapiro, J.H., Goyal, V.K.: First-photon imaging. Science **343**(6166), 58–61 (2014)
23. Kitichotkul, R., Rapp, J., Ma, Y., Mansour, H.: Simultaneous range and velocity measurement with Doppler single-photon lidar. Optica **12**(5), 604–613 (2025)
24. Krull, A., Basevi, H., Salmon, B., Zeug, A., Muller, F., Tonks, S., Muppala, L., Leonardis, A.: Image denoising and the generative accumulation of photons. In: Proc. IEEE/CVF Winter Conf. on Applications of Computer Vision (WACV). pp. 1517–1526 (2024)
25. Lee, J., Ingle, A., Chacko, J.V., Eliceiri, K.W., Gupta, M.: CASPI: collaborative photon processing for active single-photon imaging. Nat. Commun. **14**(1), 3158 (2023)
26. Lichtsteiner, P., Posch, C., Delbruck, T.: A 128 × 128 120 dB 15 μs latency asynchronous temporal contrast vision sensor. IEEE J. Solid-State Circuits **43**(2), 566–576 (2008)
27. Lin, Y., Mos, P., Ardelean, A., Bruschini, C., Charbon, E.: Coupling a recurrent neural network to SPAD TCSPC systems for real-time fluorescence lifetime imaging. Sci. Rep. **14**(1), 3286 (2024)
28. Liu, Y., Krull, A., Basevi, H., Leonardis, A., Jenkins, M.: bit2bit: 1-bit quanta video reconstruction via self-supervised photon prediction. In: Adv. Neural Inf. Process. Syst. (NeurIPS). vol. 37, pp. 88443–88485 (2024)
29. Lumentum: High-Power Q-Switched Diode-Pumped UV and Green Laser (2024), `https://www.lumentum.com/en/products/laser-solid-state-q-switched-355-532-q-series`, Accessed: 2026-06-30
30. Ma, S., Gupta, S., Ulku, A.C., Bruschini, C., Charbon, E., Gupta, M.: Quanta burst photography. ACM Trans. Graph. **39**(4), 79:1–79:16 (2020)
31. Mallat, S.: A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans. Pattern Anal. Mach. Intell. **11**(7), 674–693 (1989)
32. Mann, S., Haykin, S.: The chirplet transform: physical considerations. IEEE Trans. Signal Process. **43**(11), 2745–2761 (1995)
33. Morimoto, K., Ardelean, A., Wu, M.L., Ulku, A.C., Antolović, I.M., Bruschini, C., Charbon, E.: Megapixel time-gated SPAD image sensor for 2D and 3D imaging applications. Optica **7**(4), 346–354 (2020)
34. Mueggler, E., Rebecq, H., Gallego, G., Delbruck, T., Scaramuzza, D.: The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and SLAM. Int. J. Robot. Res. **36**(2), 142–149 (2017)
35. Nousias, S., Wei, M., Xiao, H., Wu, M., Athar, S., Wang, K.J., Malik, A., Barmherzig, D.A., Lindell, D.B., Kutulakos, K.N.: Opportunistic single-photon time of flight. In: Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 15852–15862 (2025)
36. Peleg, S., Friedlander, B.: Multicomponent signal analysis using the polynomial-phase transform. IEEE Trans. Aerosp. Electron. Syst. (TAES) **32**(1), 378–387 (1996)
37. Pi Imaging: SPAD 512. Pi Imaging Technology SA (2021), `https://piimaging.com/spad-512/`, Accessed: 2026-06-25
38. Rapp, J., Goyal, V.: A Few Photons Among Many: Unmixing Signal and Noise for Photon-Efficient Active Imaging. IEEE Trans. Comput. Imaging **3**(3), 445–459 (2017)
39. Ross, S.M.: Stochastic Processes. John Wiley & Sons (1995)
40. Scharf, L.L., Demeure, C.: Statistical Signal Processing: Detection, Estimation, and Time Series Analysis. Addison-Wesley Pub. Co. (1991)

41. Scharstein, D., Szeliski, R.: A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. Int. J. Comput. Vis. **47**(1), 7–42 (2002)
42. Scheerlinck, C., Barnes, N., Mahony, R.: Continuous-time Intensity Estimation Using Event Cameras. In: Proc. Asian Conf. on Computer Vision (ACCV). vol. 11365, pp. 308–324. Springer (2018)
43. Seets, T., Ingle, A., Laurenzis, M., Velten, A.: Motion adaptive deblurring with single-photon cameras. In: Proc. IEEE/CVF Winter Conf. on Applications of Computer Vision (WACV). pp. 1944–1953 (2021)
44. Sheinin, M., Chan, D., O'Toole, M., Narasimhan, S.G.: Dual-Shutter Optical Vibration Sensing. In: Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 16303–16312 (2022)
45. Shin, D., Xu, F., Venkatraman, D., Lussana, R., Villa, F., Zappa, F., Goyal, V.K., Wong, F.N., Shapiro, J.H.: Photon-efficient imaging with a single-photon camera. Nat. Commun. **7**(1), 12046 (2016)
46. Sundar, V., Ardelean, A., Swedish, T., Bruschini, C., Charbon, E., Gupta, M.: Sodacam: Software-defined cameras via single-photon imaging. In: Proc. IEEE/CVF Int. Conf. on Computer Vision (ICCV). pp. 8165–8176 (2023)
47. Sundar, V., Dutson, M., Ardelean, A., Bruschini, C., Charbon, E., Gupta, M.: Generalized event cameras. In: Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 25007–25017 (2024)
48. Sundar, V., Zhang, T., Jungerman, S., Gupta, M.: Quanta neural networks: From photons to perception. In: Proc. IEEE/CVF Int. Conf. on Computer Vision (ICCV). pp. 5091–5101 (2025)
49. Vernon, D.: Fourier Vision: Segmentation and Velocity Measurement using the Fourier Transform. Springer (2001)
50. Wang, P., Steindorfer, M.A., Koidl, F., Kirchner, G., Leitgeb, E.: Megahertz repetition rate satellite laser ranging demonstration at Graz observatory. Opt. Lett. **46**(5), 937–940 (2021)
51. Wang, Z., Lu, Y., Wang, L.: Revisit Event Generation Model: Self-supervised Learning of Event-to-Video Reconstruction with Implicit Neural Representations. In: Proc. Eur. Conf. on Computer Vision (ECCV). pp. 321–339. Springer (2024)
52. Wei, M., Nousias, S., Gulve, R., Lindell, D.B., Kutulakos, K.N.: Passive ultra-wideband single-photon imaging. In: Proc. IEEE/CVF Int. Conf. on Computer Vision (ICCV). pp. 8135–8146 (2023)
53. Zappa, F., Tisa, S., Tosi, A., Cova, S.: Principles and features of single-photon avalanche diode arrays. Sens. Actuators A: Phys. **140**(1), 103–112 (2007)
54. Zhang, T., Dutson, M., Boominathan, V., Gupta, M., Veeraraghavan, A.: Streaming quanta sensors for online, high-performance imaging and vision. IEEE Trans. Pattern Anal. Mach. Intell. **47**(3), 1564–1577 (2025)
55. Zhu, L., Dong, S., Li, J., Huang, T., Tian, Y.: Retina-like Visual Image Reconstruction via Spiking Neural Model. In: Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 1435–1443 (2020)