Eulerian Single-Photon Vision

Shantanu Gupta
sgupta@cs.wisc.edu

Mohit Gupta
mohitg@cs.wisc.edu

Department of Computer Sciences, University of Wisconsin-Madison

Abstract

Single-photon sensors measure light signals at the finest possible resolution — individual photons. These sensors introduce two major challenges in the form of strong Poisson noise and extremely large data acquisition rates, which are also inherited by downstream computer vision tasks. Previous work has largely focused on solving the image reconstruction problem first and then using off-the-shelf methods for downstream tasks, but the most general solutions that account for motion are costly and not scalable to large data volumes produced by single-photon sensors.

This work forgoes the image reconstruction problem. Instead, we demonstrate computationally light-weight phase-based algorithms for the tasks of edge detection and motion estimation. These methods directly process the raw single-photon data as a 3D volume with a bank of velocity-tuned filters, achieving speed-ups of more than two orders of magnitude compared to explicit reconstruction-based methods.

Project webpage: `https://wisionlab.com/project/eulerian-single-photon-vision/`

1. Introduction

The spatio-temporal resolution of digital image sensing has continually increased, culminating in *single-photon* or *quanta sensors* such as single-photon avalanche diodes (SPADs) and jots which can resolve individual photon arrivals [53, 71, 50, 46, 34]. These sensors enable an exciting array of applications, including photography in challenging conditions like low-light, fast-motion, and high dynamic range [24, 47, 18, 9], high-speed tracking [27], and 3D imaging [68]. While quanta sensors open up new opportunities by providing access to individual photons, the raw data from these sensors is heavily quantized (down to a single bit per pixel), and noisy from Poisson statistics. Cost is another problem – treating individual photons separately instead of aggregating them like conventional sensors means we fundamentally need more storage, computation, and communication (and ultimately power). These challenges are precluding large-scale adoption of this otherwise exciting technology.

SPAD arrays in particular capture binary frames (Fig. 1a) at high speeds of up to 97 kHz [71]. In this context, the most widely studied problem so far has been *image reconstruction*, with the idea being that recovering high-quality images is critical for any vision task. Reconstructing images from single binary frames is difficult, needing strong priors and computationally intensive algorithms [4, 63, 8]. A natural idea is to aggregate information over many frames [77], but this approach is prone to potentially severe motion blur – in Fig. 1a, the falling ball gets completely blurred when binary frames are naively averaged. Therefore, we need more sophisticated methods to handle motion [27, 9, 12] such as "explicit burst vision" [48], where the visual signal is reconstructed by aligning and robustly merging frames over time [47]. Motivated by the success of burst photography on smartphones [28], explicit burst vision yields high-quality results, but at heavy computational cost (Fig. 1b).

We propose a class of light-weight computer vision algorithms for SPAD arrays (or very high-speed video in general). They are motivated by the idea that many vision tasks ultimately do not need the full image [10], and are therefore not necessarily tied to the same cost-versus-quality tradeoff as image reconstruction. We propose *signal phase recovery* as a proxy problem (Sec. 3), which can be addressed without reconstructing the entire signal (image). Phase is an important feature both in visual perception [59] and in computer vision tasks [38, 51, 74, 20, 60]. Treating single-photon sensor data (video) as a 3D volume, the response of oriented and complex 3D filters applied to it encodes scene information such as motion and edge locations. Sec. 4 describes a method to accurately estimate the reliability of these filter responses, and Sec. 5 shows how we adapt classical phase-based low-level vision algorithms to extract the scene information. These methods can be run extremely fast due to involving only linear filtering and pixel-wise operations. We obtain speed-ups of more than two orders of magnitude compared to explicit burst vision, with comparable quality (Fig. 1c & Sec. 6).

The large difference in speed between explicit burst vi-

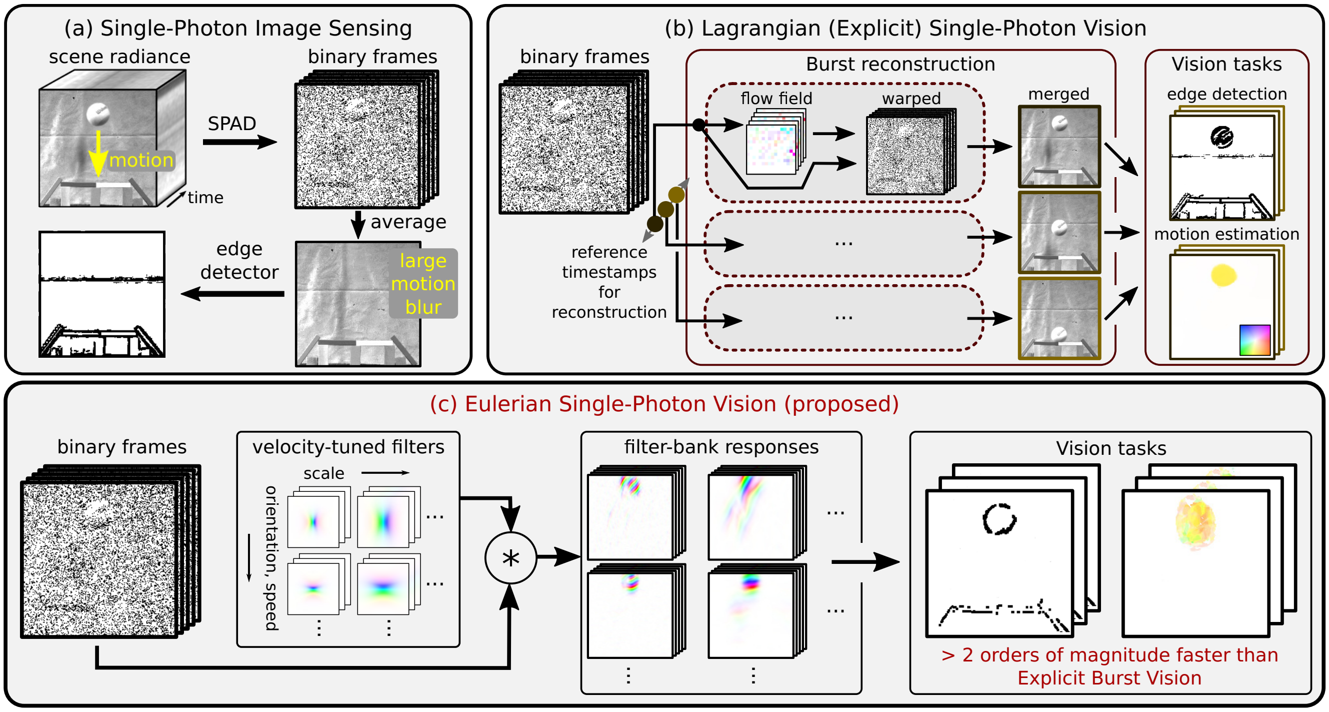

Figure 1: **Single-photon computer vision**. (a) A SPAD array captures a high-speed sequence of binary frames. A single frame is extremely noisy and quantized. Naively averaging frames over time increases the signal, but loses motion information. (b) A Lagrangian vision method based on frame-by-frame reconstruction with robust motion compensation [48, 47]. For each patch, similar patches are searched for over the rest of the sequence and noise is reduced by averaging. (c) Proposed Eulerian single-photon vision method. Single-photon data is processed in a single pass with velocity-tuned complex 3D filters (Sec. 4), and the phase of the responses is used to extract scene information such as edges and motion vectors (Secs. 5, 6), in a completely localized manner. The computation and data movement costs are both significantly lower.

sion and our approach follows from their different perspectives. Burst reconstruction invokes *search*: given a patch, the core task is to find similar patches across the other video frames. Searching over long sequences incurs a high cost, exacerbated when repeating the search for every patch. The general idea of tracking the trajectory of a patch through the exposure volume is similar to a *Lagrangian* specification in fluid mechanics, that describes the motion of individual particles in a flow field. In contrast, our approach is *Eulerian* in nature, where properties of the flow (such as rate) are described at each point in space and time, without the notion of a particle. This categorization was previously made for motion magnification [44, 76], where large speed-ups were also seen with Eulerian methods.

Implications and limitations As single-photon sensors are used more widely and specialized processor architectures are developed [3], the Eulerian approach's simplicity makes it a candidate for *on-chip implementation*, an important practical goal due to the cost of data movement. The proposed method provides a general strategy for designing lightweight algorithms for extremely fast vision tasks, directly from raw single-photon data. However, this paper should be seen just as a first step towards this stated goal: significant improvements are needed in both algorithm and implementation to be feasible on real hardware.

2. Related work

Single-photon sensors and imaging models The ability to resolve individual photons is the result of CMOS image sensors continually increasing in spatial resolution and quantum efficiency, culminating in "jot"-type sensors [70, 21, 46]. Their imaging model consists of 2D arrays of pixels independently detecting/counting photons.

SPAD arrays, on the other hand, provide the same ability through fine temporal resolution. In large pixel arrays, they can realize high-frame-rate videography with similar imaging model as jots [17, 71, 53]. SPADs have been used to realize alternate imaging models with different statistical properties, such as time-correlated single photon counting (TCSPC [26, 61]), inter-photon timing [31], and free-running SPADs [32]. We consider a passive imaging model (no controlled light source), where the SPAD captures a high-speed sequence of binary single-photon frames without needing potentially expensive timing information.

Computer vision on single-photon sensors Image reconstruction [77, 8, 24, 63] and inference [11, 25] on single-photon sensor data has been extensively studied over the past few years, but largely for static scenes and cameras. Many of these works borrow from the image denoising literature on exploiting non-local correlations or similarities [6, 14]. More recently, dynamic scenes have been addressed by motion compensation [47, 27] and convolutional networks [12, 9]. These methods reconstruct high-quality images from raw photon data as an intermediate representation and achieve high-accuracy on downstream tasks [48], albeit at extremely high computational and bandwidth costs. A few recent approaches have started incorporating and estimating motion from the temporal statistics of binary images [27, 65, 35], but for relatively simple global motion models restricted to scenes consisting of rigid objects. In contrast, our goal is to develop computationally lightweight and bandwidth-efficient vision algorithms that directly operate on single-photon data, for general motion models, including pixel-wise non-rigid motion.

Low-level vision under noise For edge detection, it has been shown that otherwise high-performing detectors have a steep drop-off when noisy images are presented as input [58, 57, 73]. Using video sequences to detect moving edges under noise has been considered for improving robustness [55, 64]. These works generally assume Gaussian noise in conventional cameras. In contrast, we consider quanta images captured by single-photon sensors, which suffer from strong Poisson noise and quantization.

3. Imaging model & a frequency-domain view

Consider a SPAD array observing a scene, capturing a sequence of frames over time. Suppose the average incident flux at a pixel is denoted by $f[\mathbf{p}]$ (in photons/second), where $\mathbf{p} := (i, j, n)$ represents the spatial location (i, j) and temporal frame index n of the pixel. The number of incident photons is modeled as a Poisson random variable, with mean $f[\mathbf{p}]$. During each frame exposure, a pixel detects at most one photon. Hence, the pixel measurements $B[\mathbf{p}]$ are binary-valued and follow a Bernoulli distribution [22]:

$$\begin{aligned} \Pr(B[\mathbf{p}] = 0) &= e^{-(\eta f[\mathbf{p}]+d)\tau} \\ \Pr(B[\mathbf{p}] = 1) &= 1 - e^{-(\eta f[\mathbf{p}]+d)\tau} \end{aligned} \tag{1}$$

where the exposure time of each frame is τ seconds, $\eta \in (0, 1]$ is the quantum efficiency, and d is the dark count rate (DCR) representing spurious detections unrelated to incident photons. We assume that distinct quanta samples $B[\mathbf{p}]$ and $B[\mathbf{p}']$ are statistically independent of each other. Under low flux ($<<$ 1 photon/pixel), Eq. 1 can be linearized effectively [77], but in general the response is non-linear, and can be described as a *soft saturation* [32, 21].

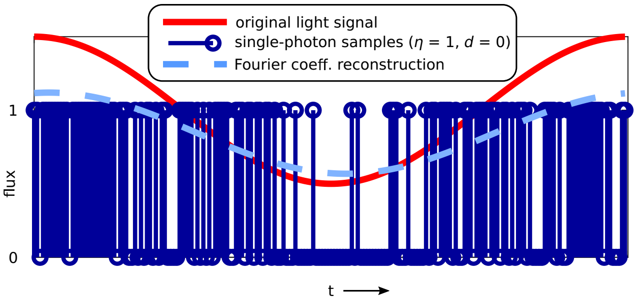

Figure 2: **Direct phase recovery from Fourier coefficients**. Simulated single-photon samples ($\# = 200$) of a sinusoid, and a direct reconstruction from the Fourier coefficient (offset adjusted manually).

We now view the imaging model in frequency-domain. An example with simulated data is shown in Fig. 2, where the original single-tone sinusoidal intensity signal of known frequency (in red) is compared with a reconstruction from the corresponding Fourier coefficient of the sampled binary data (dotted, light blue). We observe that the amplitude of the reconstructed wave is smaller, and that this deviation is largely due to bias from the soft saturation in the sensor and not the variance of photon noise. It is possible to correct for this bias using maximum-likelihood estimation [63] but the optimization algorithms (or related models learned from data) are typically expensive. In contrast to the amplitude, the phase of the reconstructed wave in Fig. 2 is quite close to the true value. It can be shown that for the single-tone case in particular, the Fourier coefficient phase is unbiased in most cases (see Appendix A), making it a simple closed-form estimator. Further, simulations suggest that the variance of the coefficient phase is also close to the Cramér-Rao lower bound. Please see the supplementary report for the details of this analysis. More work is needed to rigorously extend these observations from pure tones to general signals and localized band-pass filtering as done in practice.

Bias from soft saturation does not preclude the use of amplitude (also an Eulerian approach in terms of data-flow), as it is not always central to the task. For many problems, amplitude-based methods can be employed even with biased (but inexpensive) estimates, and we present their results for the case of edge detection in Sec. 6. But the reasons outlined above motivate us to focus our scope on designing phase-based approaches.

4. Encoding motion with velocity-tuned filters

How do we extract information from the *photon cube* captured by a SPAD array? To overcome the extreme noise and quantization of individual frames, it is necessary to aggregate information over the temporal dimension of the

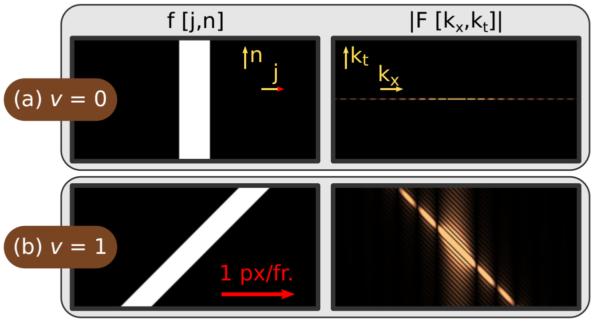

Figure 3: **Velocity-tuning principle**. A 1D box-shaped signal imaged over time (vertical) at two speeds: (a) $v = 0$, and (b) $v = 1$ pixel/frame. In both cases, the 2D (x-t) spectrum lies along a line given by $k_t = -v \cdot k_x$. This principle extends to moving 2D signals or video: the line in this case is given by $k_t = -v \cdot \sqrt{k_x^2 + k_y^2}$, with k_x and k_y representing spatial frequencies along the $x-$ and $y-$axes.

cube. Image reconstruction can aid this goal but has a cost-vs-quality trade-off (Sec. 1), where inexpensive methods such as summing frames over time can result in severe motion blur but burst reconstruction with motion compensation [47, 48] becomes infeasible for real-time processing.

Our approach extracts scene information *directly from the photon cube*, relying on classical analyses of motion as the spatio-temporal orientation of intensity or phase isosurfaces when viewing videos as 3D volumes [1, 20]. Motion information is extracted by 3D oriented band-pass filters, also termed *velocity-tuned* [20, 29] since they respond only to motion at a specific range of velocities (Fig. 3). With an appropriate design and analysis of their reliability under noise (Secs. 4.1 & 4.2), the responses from these filters can be readily used in classical phase-based low-level vision algorithms, discussed further in Sec. 5.

4.1. Filter-bank design & implementation

We use *space-time separable* log-Gabor filters [19, 37] and apply them in frequency-domain. The spatial filters are polar-separable and similar to complex steerable pyramids [23, 62, 75, 72] – we construct the filters over three scales at six orientations each. The temporal filters are designed separately for each scale, with the center frequencies obtained through the velocity-tuning relation (Fig. 3) for a fixed set of velocities – we use three speeds $\{0, v_1, v_2\}$, with v_1 and v_2 set depending on the scene. The supplementary material provides more specific details on the design.

We extract all filter responses at the original video resolution. With this choice we use a large amount of memory: the filter-bank is over-complete by a factor of $2 \times \#\text{scales} \times \#\text{orientations} \times \#\text{speeds} = 2 \times 3 \times 6 \times 3 = 108$. If memory is limited (*e.g.* on GPU), savings can be made by sub-sampling the responses of filters at coarse scales (both spatially and temporally). Sparse array representations may also be useful: since real video signals are highly structured, every filter responds strongly at only relatively few pixels, especially for the non-zero velocity tunings.

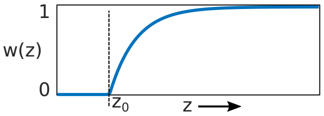

Figure 4: Weight function for filter z-scores (Sec. 4.2).

4.2. Reliability analysis of filter responses

It is critical for the filter-bank to extract relevant details from the binary and noisy SPAD samples *while rejecting spurious responses* which dominate the data. Therefore, it is crucial to robustly estimate uncertainty in filter responses.

Consider a filter $h_{\mathbf{k}}$ tuned around the 3D frequency $\mathbf{k} := (k_x, k_y, k_t)$. Its response to the input video stream $B[\mathbf{p}]$ is denoted by $R_{\mathbf{k}}[\mathbf{p}] := \sum_{\mathbf{q} \in \text{Support}(h_{\mathbf{k}})} h_{\mathbf{k}}[\mathbf{q}] B[\mathbf{p}-\mathbf{q}]$. From the central limit theorem, we expect $R_{\mathbf{k}}[\mathbf{p}]$ to be approximately (complex) normally-distributed, with a variance

$$\text{Var}\left[R_{\mathbf{k}}[\mathbf{p}]\right] = \textstyle\sum_{\mathbf{q}} |h_{\mathbf{k}}[\mathbf{q}]|^2 \cdot \text{Var}(B[\mathbf{p}-\mathbf{q}]) \,. \qquad (2)$$

Assuming an ideal sensor with quantum efficiency $\eta = 1$ and dark counts $d = 0$ (from Eq. 1), we can approximate this variance as $\text{Var}(B[\mathbf{p}]) \approxeq V(f[\mathbf{p}])$ [21], where

$$V(x) := e^{-x}(1 - e^{-x}) \,.$$

From a rough estimate $\hat{c}[\mathbf{p}]$ of the local flux (*e.g.* through a blur kernel on $B[\mathbf{p}]$), we can approximate Eq. 2 further:

$$\text{V}_{\mathbf{k}}[\mathbf{p}] := V(\hat{c}[\mathbf{p}]) \textstyle\sum_{\mathbf{q}} |h_{\mathbf{k}}[\mathbf{q}]|^2 \approx \text{Var}(R_{\mathbf{k}}[\mathbf{p}]) \,. \qquad (3)$$

The sum $\sum_{\mathbf{q}} |h_{\mathbf{k}}[\mathbf{q}]|^2$ is known. At run-time, $R_{\mathbf{k}}[\mathbf{p}]$ is normalized to a standard- or z-score [39]:

$$z_{\mathbf{k}}[\mathbf{p}] := {|R_{\mathbf{k}}[\mathbf{p}]|}/{\sqrt{V_{\mathbf{k}}[\mathbf{p}]}} \qquad (4)$$

For later algorithms, we further map the z-score to a weight $w \in [0, 1]$ as $w(z) := 1 - \exp\left(-\max(0, z - z_0)\right)$, plotted in Fig. 4. The parameter z_0 is set between 2 and 6, to ensure weak responses do not contribute.

Implications for filter design From the Gabor uncertainty relation, smaller band-width corresponds to larger spatio-temporal support (*e.g.* for coarse scales, or for elongated filters with small angular sensitivity). Typically such filters have lower variance in Eq. 2, but also poor localization, resulting in a well-studied trade-off [7, 15]. Sec. 6 discusses some related examples from real videos captured with a SPAD sensor.

Since the noise level changes with light levels, a reliable filter in strong light can become unreliable in low light. The filter design and downstream algorithms need to adapt to this variation, motivating our use of multi-scale filter-banks. z-scores further enable a principled approach.

5. Low-level vision algorithms

We adapt classical phase-based algorithms from the image and video processing literature to the single-photon setting, for the tasks of edge detection and motion estimation. These methods can be parallelized easily due to having largely pixel-wise computations.

5.1. Edge detection: temporal phase congruency

Phase congruency [54, 37] is the observation that features like edges are discontinuities where the phase of all frequency components aligns. It also applies to video, as a moving edge traces a plane in 3D. In this case a multi-scale bank of velocity-tuned filters plays the role of the frequency and *temporal phase congruency* (TPC [55]) is detected.

For a filter tuned to frequency $\mathbf{k} := s\hat{\mathbf{k}}$, where s denotes the scale and $\hat{\mathbf{k}}$ the unit vector along its spatio-temporal orientation, the phase congruency PC along $\hat{\mathbf{k}}$ is given as

$$\mathrm{PC}_{\hat{\mathbf{k}}}\left[\mathbf{p}\right] := \frac{\left|\sum_s R_{s\hat{\mathbf{k}}}[\mathbf{p}]\right|}{\sum_s \left|R_{s\hat{\mathbf{k}}}[\mathbf{p}]\right|} \tag{5}$$

which is 1 if the responses at all scales have the same phase. [1] This definition yields a normalized quantity invariant to light level or signal contrast, and avoids the phase wrapping problem. Once $\mathrm{PC}_{\hat{\mathbf{k}}}$ is computed for all orientations, edge strength and orientation are estimated using principal component analysis [66, 36]. The second eigenvalue here (when significant) yields space-time "corners" [40, 43, 38], but its use is outside this paper's scope.

In our implementation, the right-hand side of Eq. 5 is multiplied by the weight term of Fig. 4, to exclude orientations with weak responses (we set z_0 to 2).

Normal velocities from 3D edge orientation The edge direction represents the normal to the spatio-temporal plane traced out by a moving edge over time, and therefore directly yields normal velocity estimates at edge locations. Some related results are presented in Sec. 6.2.

The basic principle also underlies equivalent techniques in event vision [2, 5] — due to their fundamental similarity we may expect similar-quality results from them in regions with motion, after accounting for differences from quantum efficiency and the sensing threshold of the event camera. Directly sensing intensity with SPAD has the advantage that we automatically recover static edges at the same time.

5.2. 2D motion estimation: local frequency method

A frequency-domain approach to motion estimation is formulated through the *phase constancy* constraint [20], which is structurally similar to the brightness constancy relation behind intensity-based optical flow. In this case, the constraint is applied separately for each filter, yielding a *component velocity estimate*. For the filter tuned to frequency $\mathbf{k}$, the phase constancy relation is given by

$$\phi_{\mathbf{k}}(x, y, t) = (\text{constant}), \tag{6}$$

where $\phi_{\mathbf{k}} := \arg(R_{\mathbf{k}})$ represents the local phase of the response $R_{\mathbf{k}}$. Differentiating with respect to t, we get

$$\nabla\phi_{\mathbf{k}} \cdot (v_x, v_y, 1) = 0 \tag{7}$$

where $\nabla\phi_{\mathbf{k}}$ represents the (3D) *local frequency* or phase gradient. Fleet and Jepson [20] provide a method to extract local frequency directly from the responses without phase unwrapping. With component velocity equations formed as above from all reliable responses, we solve a *weighted* least-squares problem to obtain the full 2D velocity or optical flow (v_x, v_y). The weights for each equation are taken from Sec. 4.2, with the threshold z_0 set to 6. Please see the supplementary material for more implementation details.

Scale The method is applied independently at each scale; some related results are discussed in Sec. 6.2.

6. Results

We demonstrate the proposed techniques on real binary frame sequences captured with the SwissSPAD2 sensor [71], which has a 256×512 resolution and frame rate up to 97,700 FPS. Flux levels are reported as *photons-per-pixel*, abbreviated as *ppp*.

Pre-processing Sequences with gradual motion (where the flow $<< 1$ pixel per frame, common with high frame rates) are temporally low-passed and sub-sampled, approximately equivalent to sampling with a multi-bit sensor [21]. The set of tuned velocities is adapted accordingly.

The SPAD prototype is a research-grade device with several "hot pixels" with high dark count rate. These pixels are detected offline with a dark frame, and interpolated.

Implementation Filtering is done in frequency-domain due to the large support of the filters. For fair comparisons, all our algorithms and the methods compared to (BM3D [56] & burst reconstruction [47]) are implemented in MATLAB [49] and run on a single CPU core.

Video clips The results are best visualized as videos, available at the project URL: `https://wisionlab.com/project/eulerian-single-photon-vision/`.

6.1. Edge detection

We compare the TPC detector applied directly to the video sequence (after temporal low-pass filtering) to reconstruction-based approaches, comprising of taking the sum of all frames, single-image denoising with BM3D [4, 56], and the Lagrangian approach of Fig. 1b [47, 48].

[1] In practice, the right-hand side of Eq. 5 is weighted separately to exclude blurred features. Further, to better localize features we ultimately compute $1 - \cos^{-1}(\mathrm{PC})$ as the edge strength measure – please see [37] for a more detailed discussion.

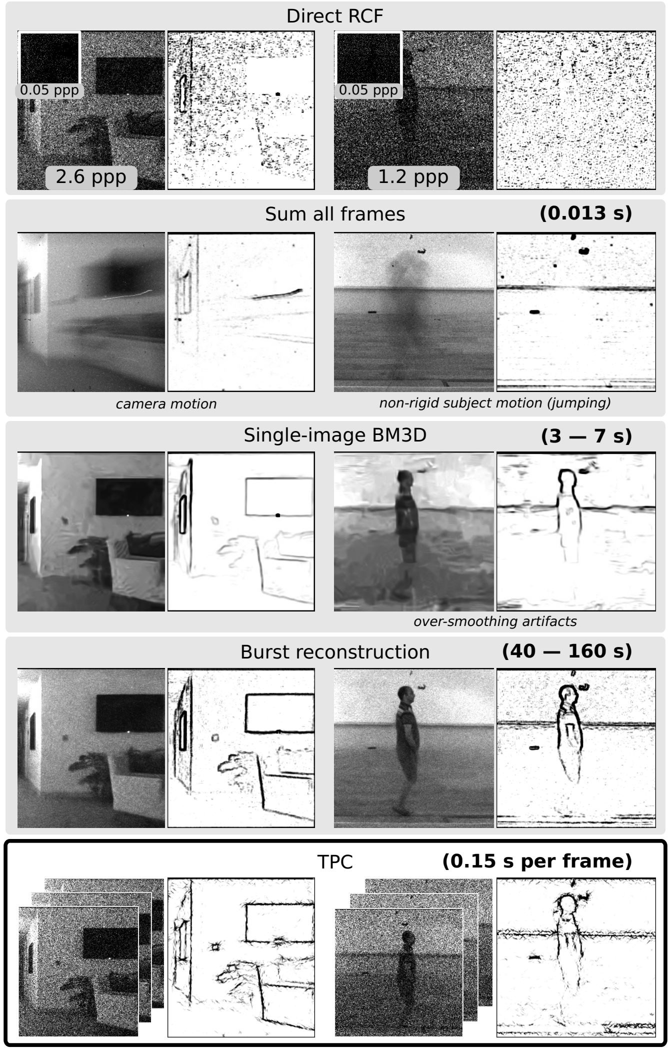

Figure 5: **Edge detection on real SPAD video** (Sec. 6.1). From top: single frames (after temporal low-pass filtering) with approximate flux levels indicated. The original binary frames from SwissSPAD2 are shown inset. Next: results of the Richer Convolutional Features detector (RCF [45]) applied directly, followed by those after various reconstruction-based approaches: directly summing frames, BM3D [4, 56], and burst photography [47]. Bottom row shows edges from the Eulerian approach with the TPC algorithm (Sec. 5.1). Run-times are reported for the original 256×512-sized frames from which these images are cropped, and the time taken by RCF is *not* included.

We use the Richer Convolutional Features network (RCF [45]) as the reference detector for all reconstructed images.

Fig. 5 shows results on videos captured with the SwissSPAD2 sensor, lowpass-filtered to sequences of 120 frames. Non-maximal suppression is *not* performed, to enable direct comparison of the localization of the underlying detector. The proposed Eulerian approach (especially TPC) achieves similar-quality results as the Lagrangian method, but more than two orders of magnitude faster. It is also faster than the tested BM3D implementation [56] by an order of magnitude, with the same hardware and software environment. Further, we note here that the single-image denoising approach is prone to artifacts from over-smoothing in extremely low-flux conditions (*e.g.* in the right column of Fig. 5, the straight edge in the background gets bent).

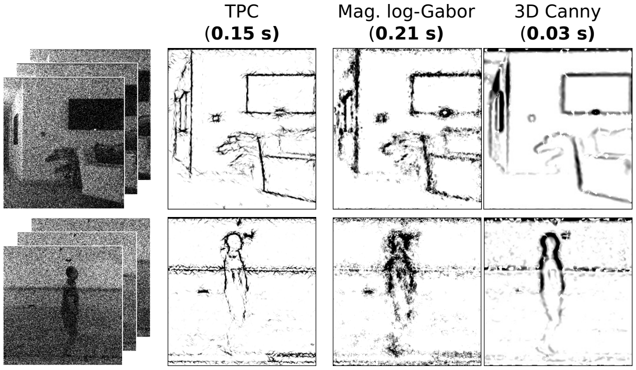

Figure 6: **Comparing phase-based detection with amplitude-based methods**. From left: first two columns show edges extracted with temporal phase congruency (same as in Fig. 5) and the magnitude of the responses, respectively, from the same log-Gabor filter-bank. Right column shows edges extracted by a 3D version of the Canny detector [64, 52]. While 3D Canny performs significantly better than the magnitude of the log-Gabor responses (and much faster), TPC yields better-localized (sharper) edges.

Comparing phase-based and magnitude-based detectors We implement a simple magnitude-based detector run separately at each scale of the log-Gabor filter-bank, where principal components analysis is performed similarly to TPC to obtain edge strengths and orientations. Feature information is aggregated across scales by averaging edge strengths, following previous works [16, 45]. We also consider a 3D version of the classic Canny edge detector [7, 52, 64], making it multi-scale by setting different values for the σ parameter of its underlying Gaussian kernel, and averaging as above.

Fig. 6 shows the results of magnitude-based detectors on the same scenes as Fig. 5. We find that the edges from the raw log-Gabor magnitude have much poorer localization than TPC, although resistance to noise or SNR is similar as they are based on the same filter responses. The 3D Canny detector performs much better than the log-Gabor magnitude, and is much faster as it uses fewer filters (only three smoothed gradients). While its performance may be adequate in many settings, the edges are typically not as sharp as TPC, suggesting inferior localization.

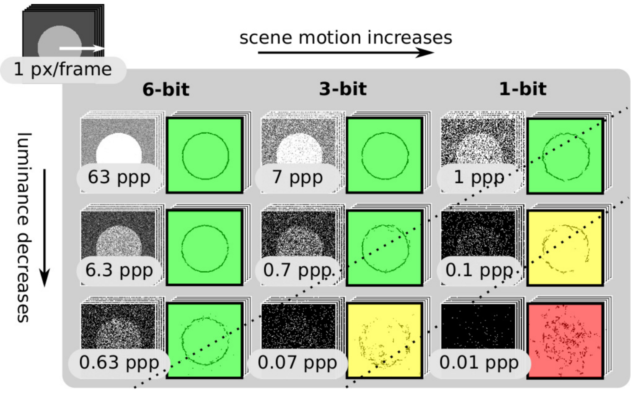

Figure 7: **Influence of flux on edge detection (simulations)**. A SPAD video with 51 frames of size 128×128, simulated with a fixed observed motion of 1 pixel per frame and at varying flux levels (measured here in photons per pixel, ppp) and precision. Edges are detected by temporal phase congruency. Dotted lines represent approximate contour lines of effective per-frame flux or SNR, which correspond closely to edge map quality. For sufficiently well-lit scenes (flux around 1 ppp), edges can be detected even from very fast binary video. See Sec. 6.1 for more discussion.

Impact of light level The performance of the edge detector depends on the light level in the scene as well as the amount of motion. We can standardize the motion between frames by appropriate low-pass filtering: for slow-moving scenes, this effectively yields higher-precision data. Fig. 7 shows this variation for the TPC detector with a fixed filter-bank, with a simulated synthetic scene. We assume an ideal sensor with no dark counts and full quantum efficiency. Even under extremely challenging conditions (1-bit samples and motion), TPC can successfully recover edges. The recovery ultimately depends on the *total number of incident photons*, with flux levels as low as ~ 1 photon-per-pixel being sufficient for reasonable quality.

Role of filter-bank design A similar example as Fig. 7, but with real data, is presented in Fig. 8. In this case, the same scene (a person juggling two footballs) is captured twice under moderate and low light, respectively. TPC is run with filters of two different scale ranges: one spanning spatial wavelengths from 3 to 13 pixels ("fine scales"), and another spanning 6 to 28 pixels ("coarse scales"). The fine-scales filter-bank yields sharp edges under more light, but suffers in low light. In contrast, the coarse-scale filters give thicker (less resolved) edges, but are more reliable under low light. The SNR can be improved further by reducing the angular bandwidth, which helps with long edges but is prone to over-shooting around curved edges – this is another form of de-localization or loss of resolution, and the trade-off has been studied in previous works [7, 15, 33].

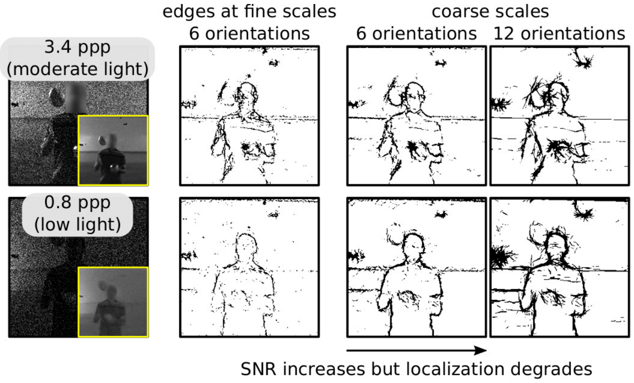

Figure 8: **Edge detection under varying filter configurations**. Scene with non-rigid motion (person juggling two footballs). Top row: input frame after temporal LPF (a tone-mapped burst reconstruction [47] is inset, only for reference), and edges detected by temporal phase congruency (see Sec. 6.1 for details). Bottom row: same scene recorded again under less light. The fine-scale edges worsen significantly, but the corresponding coarse-scale result (middle column) retains quality. For the last column the angular bandwidth of the filters is reduced. Long edges are now recovered more reliably, but the detector overshoots around curved edges such as the football, the head, and the elbows.

Quantitative evaluation Unlike the Lagrangian approach with burst reconstruction, single-image denoising methods are localized (in time) by design, and may be comparable in cost depending on implementation and the hardware platform. Therefore, we compare its result quality with our approach in more detail. Detailed results from a numerical evaluation on simulated data are presented in the supplementary report. In summary, we find that TPC is relatively resilient down to flux levels of 1 ppp (as seen in Fig. 7). The tested BM3D algorithm [4] was found to yield reasonable results if the flux level was above 3 ppp, but lower-flux settings result in a severe break-down in performance.

6.2. Motion estimation

We demonstrate estimates of the normal velocity at edges and the 2D velocity in general, as described in Sec. 5. Only TPC edges are considered for normal velocity estimation as the amplitude-based edges were of inferior quality. The scale for 2D velocity estimation is chosen manually. For comparison we reconstruct pairs of images using the same methods as the edge detection experiments, and use *RAFT-it* [69, 67] to estimate optical flow.

Results on SwissSPAD2 video sequences [2] are shown in Fig. 9. The proposed method is significantly better than directly applying RAFT-it on noisy frames, in that it can

[2] While the scenes shown here are captured with a static camera, the algorithm of Sec. 5.2 can also be used with camera motion.

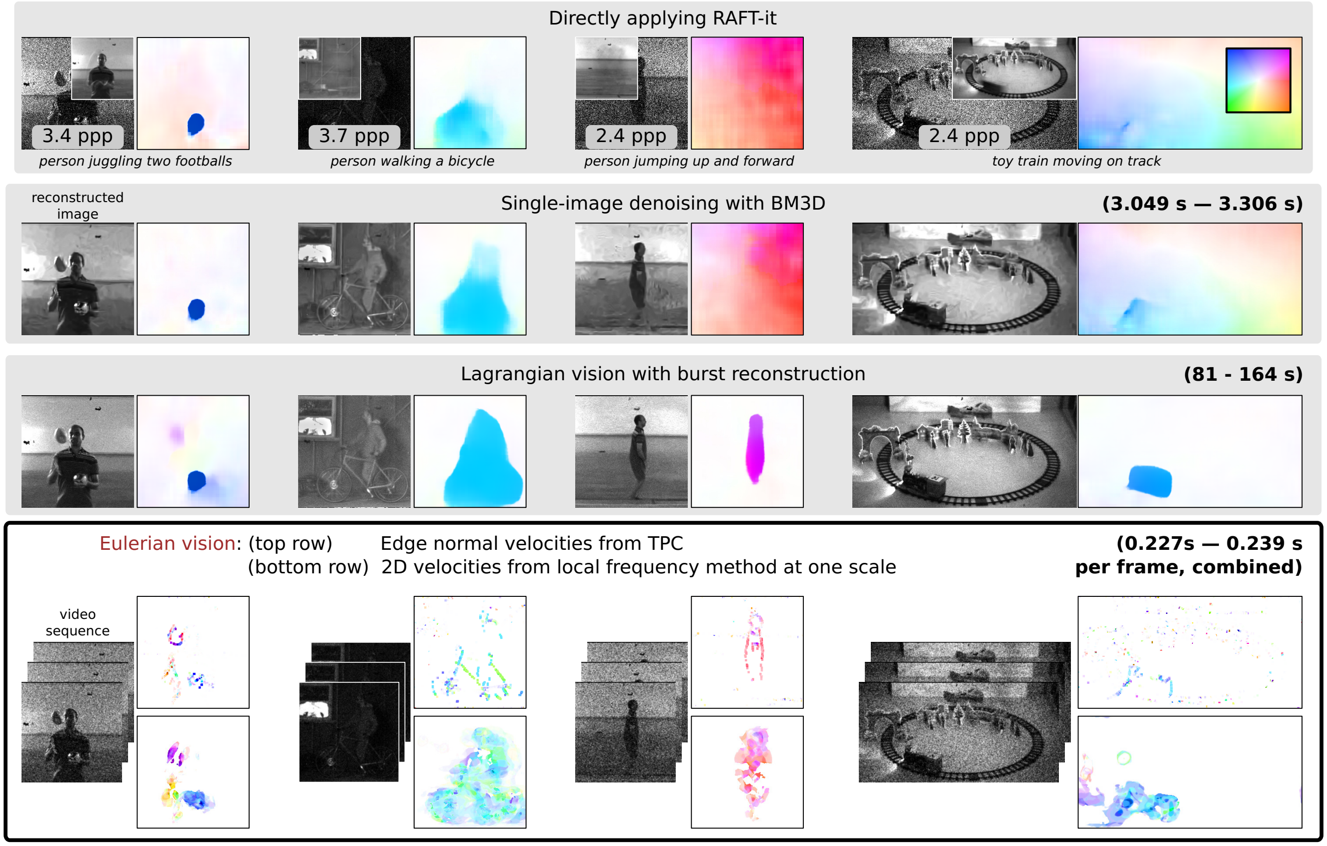
Directly applying RAFT-it
3.4 ppp
person juggling two footballs
3.7 ppp
person walking a bicycle
2.4 ppp
person jumping up and forward
2.4 ppp
toy train moving on track
reconstructed image
Single-image denoising with BM3D
(3.049 s — 3.306 s)
Lagrangian vision with burst reconstruction
(81 - 164 s)
Eulerian vision: (top row) Edge normal velocities from TPC
(bottom row) 2D velocities from local frequency method at one scale
(0.227s — 0.239 s per frame, combined)
video sequence

Figure 9: **Motion estimation from SPAD video**. Top row: input frames after temporally low-pass filtering the binary SwissSPAD2 sequence, and optical flow estimated with RAFT-it [69, 67]. Inset shows the average of the entire sequence to help visualize the true motion. Second row: frame-by-frame denoising with BM3D [4]. Third: Lagrangian/explicit burst vision [47, 48]. Bottom row: normal velocities from edges extracted by Temporal Phase Congruency (Sec. 5.1), and 2D velocities from the method of Sec. 5.2. TPC edges are thickened for visualization. The phase-based methods reliably isolate object motion, unlike two-frame estimation directly or after BM3D. The Lagrangian method provides good quality and reliability, but at much higher cost. Similar to Fig. 5, reported run-times do not include the time taken by RAFT-it.

reliably separate the moving object from the static background. It also achieves considerably better performance than single-image denoising, due to the temporal incoherence of denoising artifacts. The Lagrangian method yields high-quality results, but also at significantly higher cost.

Multiple cues The edge normal velocity maps and the 2D maps of Fig. 9, have distinct characteristics. The former is well-localized, but only available sparsely (at edges). In contrast, the 2D estimates (obtained at a very coarse scale), cover more area but the estimates often over-shoot the subject's boundaries. Choosing to estimate 2D velocity at finer scales doesn't always fix the problem, as seen in Fig. 10 — the filter-bank responses may not be reliable enough at most pixels to obtain any estimates, leading again to very sparse flow maps. Fine-scale estimates are also more likely to suffer from the aperture problem. Integrating these different flow cues (edge normal velocities and 2D estimates at each scale) could yield more informative results, and has been considered in the event vision literature [2].

7. Discussion & future outlook

While single-photon sensors provide the prospect of recording visual details at the resolution of individual photons, they also introduce challenges: a very noisy and quantized imaging model, and extremely large volumes of data generated, resulting in prohibitive compute and bandwidth requirements. In this work, we demonstrated light-weight vision algorithms based on linear filtering and local phase-based processing of raw single-photon data, bypassing the expensive intermediate step of image reconstruction.

On-chip implementation As hardware architectures are developed for single-photon sensors that can perform complex calculations at the photon-level [3], the proposed approach may enable completely on-chip *real-time photon-*

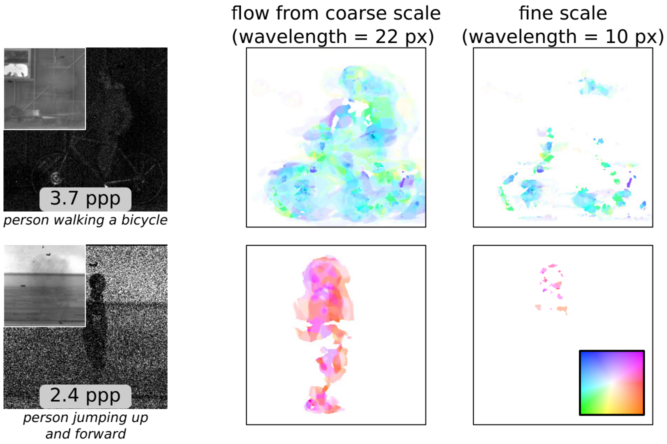

Figure 10: **Motion estimates at two scales**. For the middle two scenes in Fig. 9, 2D velocity estimates are extracted from two distinct scales of the filter-bank using the algorithm of Sec. 5.2. Coarse-scale estimates are reliably obtained at more pixels, but can over-shoot object boundaries. Fine-scale estimates are better-localized, but suffer from the aperture problem and are not reliably obtained at as many pixels, due to more noise in filter responses.

processing. However, our methods, as implemented currently, have large memory requirements due to frequency-domain filtering. An important next step is to filter in the primal ("spatial") domain, and *recursively* in time, such that memory of past frames is not required [42, 13]. Such on-chip vision systems could spur wider deployment of single-photon imaging in real-world computer vision applications including SLAM, scientific fields like bio-mechanics, and in consumer domains like sports videography.

Sensor parameters Our techniques can be applied with other frame-based image sensors including jots, after adapting the filter design and analysis of Sec. 4 to the frame rate (*e.g.* the velocity tunings of the filter-bank), and after including read noise, fixed-pattern noise, *etc.* A possible source of error is motion aliasing at lower frame rates [41]. We may need to restrict the filters to only low spatial frequencies in such cases since the aliasing is stronger for higher-frequency textures.

Optimal filter design Better filter-banks may be obtained through learning-based or even classical methods [15]. In addition to target metrics such as SNR, we may have other relevant constraints such as causality and resource cost.

Acknowledgments

This research was supported in part by NSF CAREER award #1943149. We also thank the anonymous reviewers for their valuable feedback.

A. Single-tone DFT phase is nearly unbiased

Consider a non-negative 1D signal $f[n]$ imaged by an ideal sensor (quantum efficiency $\eta = 1$ and dark counts $d = 0$) as $B[n]$, with a unit exposure time. From the imaging model of Sec. 3, we have

$$\mathrm{E}\left[B[n]\right] = 1 - e^{-f[n]}, \tag{8}$$

and from the power series representation of the exponential function, we can represent $\mathrm{E}\left[B[n]\right]$ as

$$\mathrm{E}\left[B[n]\right] = f[n] - \frac{1}{2!} f[n]^2 + \frac{1}{3!} f[n]^3 - \ldots + \ldots \tag{9}$$

Assume that a total of N samples are acquired, and that $f[n]$ is a single-tone sinusoid:

$$f[n] = c + a \cos\left(\frac{2\pi}{N} k_0 n + \phi\right) \tag{10}$$

where k_0 denotes the signal frequency (assumed integer), c the constant offset, a the amplitude, and ϕ the initial phase. We use the discrete Fourier transform, denoted by $\mathcal{F}_p$ for the p-th power of $f[n]$ and by $\mathcal{B}$ for $B[n]$:

$$\mathcal{F}_p[k] := \textstyle\sum_{n=0}^{N-1} f[n]^p \, e^{-i \frac{2\pi}{N} kn} \tag{11}$$

$$\mathcal{B}[k] := \textstyle\sum_{n=0}^{N-1} B[n] e^{-i \frac{2\pi}{N} kn}. \tag{12}$$

Then from Eq. 9 and the linearity of the Fourier transform:

$$\mathrm{E}\left[\mathcal{B}[k]\right] = \mathcal{F}_1[k] - \frac{1}{2!}\mathcal{F}_2[k] + \frac{1}{3!}\mathcal{F}_3[k] - \ldots . \tag{13}$$

We further take the expectation of the phase to be approximately equal to the phase of the expectation, *i.e.* $\mathrm{E}\left[\arg\left(\mathcal{B}[k]\right)\right] \approxeq \arg\left(\mathrm{E}\left[\mathcal{B}[k]\right]\right)$, which is justified when its SNR is high, or equivalently, the bulk of the distribution of $\mathcal{B}[k]$ is far from the origin of the complex plane. Now, to obtain $\mathcal{F}_2$, we use Eq. 10 as follows:

$$\begin{aligned} f[n]^2 = c^2 &+ 2ca \cos\left(\frac{2\pi}{N} k_0 n + \phi\right) \\ &+ \frac{a^2}{2}\left(1 + \cos\left(2\left(\frac{2\pi}{N} k_0 n + \phi\right)\right)\right), \end{aligned} \tag{14}$$

from which the Fourier coefficient is readily extracted. A similar process is repeated for higher powers $\mathcal{F}_3, \mathcal{F}_4$, and so on. The key is that *the term corresponding to the frequency* k_0 *always has phase* ϕ *for all powers*, and the phase of the expected Fourier coefficient $\arg\left(\mathrm{E}\left[\mathcal{B}[k_0]\right]\right)$ should therefore be ϕ as well. The only case where this does not happen is when any of the higher harmonic components from the higher-power terms *alias* onto k_0 after sampling. A possible case is when $k_0 = {}^{N}\!/_{3}$, where the second harmonic aliases onto the component at $-k_0$: higher-order harmonics can alias in a similar way (though the impact reduces sharply with order). Similar findings have been reported previously in the communications theory literature [30].

The above expressions also illustrate the difficulty of directly estimating the amplitude a from the single-photon samples, as the magnitude of $\mathrm{E}\left[\mathcal{B}[k_0]\right]$ is related non-linearly to a.

References

[1] Edward H. Adelson and James R. Bergen. Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, 2(2):284, Feb. 1985.

[2] Himanshu Akolkar, Sio Hoi Ieng, and Ryad Benosman. Real-time high speed motion prediction using fast aperture-robust event-driven visual flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[3] Andrei Ardelean. *Computational Imaging SPAD Cameras*. PhD thesis, EPFL, 2023.

[4] Lucio Azzari and Alessandro Foi. Variance Stabilization for Noisy+Estimate Combination in Iterative Poisson Denoising. *IEEE Signal Processing Letters*, 23(8):1086–1090, Aug. 2016.

[5] Francisco Barranco, Cornelia Fermuller, and Yiannis Aloimonos. Bio-inspired Motion Estimation with Event-Driven Sensors. In Ignacio Rojas, Gonzalo Joya, and Andreu Catala, editors, *Advances in Computational Intelligence*, volume 9094, pages 309–321. Springer International Publishing, Cham, 2015.

[6] A. Buades, B. Coll, and J. M. Morel. A Review of Image Denoising Algorithms, with a New One. *Multiscale Modeling & Simulation*, 4(2):490–530, Jan. 2005.

[7] John Canny. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6), Nov. 1986.

[8] Stanley Chan, Omar Elgendy, and Xiran Wang. Images from Bits: Non-Iterative Image Reconstruction for Quanta Image Sensors. *Sensors*, 16(11):1961, Nov. 2016.

[9] Paramanand Chandramouli, Samuel Burri, Claudio Bruschini, Edoardo Charbon, and Andreas Kolb. A Bit Too Much? High Speed Imaging from Sparse Photon Counts. In *2019 IEEE International Conference on Computational Photography*, May 2019.

[10] Bo Chen and Pietro Perona. Vision without the Image. *Sensors*, 16(4):484, Apr. 2016.

[11] Bo Chen and Pietro Perona. Seeing into Darkness: Scotopic Visual Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[12] Yiheng Chi, Abhiram Gnanasambandam, Vladlen Koltun, and Stanley H Chan. Dynamic Low-light Imaging with Quanta Image Sensors. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision – ECCV 2020*, pages 122–138, 2020.

[13] C.W.G. Clifford, K. Langley, and D.J. Fleet. Centre-Frequency Adaptive IIR Temporal Filters for Phase-Based Image Velocity Estimation. In *IEE International Conference on Image Processing and Applications*, pages 173–178, Edinburgh, United Kingdom, July 1995.

[14] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3D transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8), Aug. 2007.

[15] D. Demigny. On optimal linear filtering for edge detection. *IEEE Transactions on Image Processing*, 11(7):728–737, July 2002.

[16] Piotr Dollar and C. Lawrence Zitnick. Fast Edge Detection Using Structured Forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(8):1558–1570, Aug. 2015.

[17] Neale A. W. Dutton, Istvan Gyongy, Luca Parmesan, Salvatore Gnecchi, Neil Calder, Bruce R. Rae, Sara Pellegrini, Lindsay A. Grant, and Robert K. Henderson. A SPAD-Based QVGA Image Sensor for Single-Photon Counting and Quanta Imaging. *IEEE Transactions on Electron Devices*, 63(1):189–196, Jan. 2016.

[18] Omar A. Elgendy, Abhiram Gnanasambandam, Stanley H. Chan, and Jiaju Ma. Low-Light Demosaicking and Denoising for Small Pixels Using Learned Frequency Selection. *IEEE Transactions on Computational Imaging*, 7:137–150, 2021.

[19] David J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, 4(12):2379, Dec. 1987.

[20] David J. Fleet and Allan D. Jepson. Computation of component image velocity from local phase information. *International Journal of Computer Vision*, 5(1):77–104, Aug. 1990.

[21] Eric R. Fossum. Modeling the performance of single-bit and multi-bit quanta image sensors. *IEEE Journal of the Electron Devices Society*, 1(9):166–174, 2013.

[22] Eric R. Fossum, Jiaju Ma, and Saleh Masoodian. Quanta image sensor: Concepts and progress. In Mark A. Itzler and Joe C. Campbell, editors, *SPIE Commercial + Scientific Sensing and Imaging*, page 985805, Baltimore, Maryland, United States, May 2016.

[23] William T. Freeman and Edward H. Adelson. The Design and Use of Steerable Filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(9), Sept. 1991.

[24] Abhiram Gnanasambandam and Stanley H. Chan. HDR Imaging with Quanta Image Sensors: Theoretical Limits and Optimal Reconstruction. *IEEE Transactions on Computational Imaging*, 6:1571–1585, Nov. 2020.

[25] Bhavya Goyal and Mohit Gupta. Photon-Starved Scene Inference Using Single Photon Cameras. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2512–2521, 2021.

[26] Anant Gupta, Atul Ingle, Andreas Velten, and Mohit Gupta. Photon-Flooded Single-Photon 3D Cameras. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6763–6772, Long Beach, CA, USA, June 2019. IEEE.

[27] Istvan Gyongy, Neale Dutton, and Robert Henderson. Single-Photon Tracking for High-Speed Vision. *Sensors*, 18(2):323, Jan. 2018.

[28] Samuel W. Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T. Barron, Florian Kainz, Jiawen Chen, and Marc Levoy. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Transactions on Graphics*, 35(6), Nov. 2016.

[29] David J. Heeger. Model for the extraction of image flow. *Journal of the Optical Society of America A*, 4(8):1455, Aug. 1987.

[30] A. Host-Madsen and P. Handel. Effects of sampling and quantization on single-tone frequency estimation. *IEEE*

Transactions on Signal Processing, 48(3):650–662, Mar. 2000.

[31] Atul Ingle, Trevor Seets, Mauro Buttafava, Shantanu Gupta, Alberto Tosi, Mohit Gupta, and Andreas Velten. Passive Inter-Photon Imaging. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8581–8591, Nashville, TN, USA, June 2021. IEEE.

[32] Atul Ingle, Andreas Velten, and Mohit Gupta. High Flux Passive Imaging with Single-Photon Sensors. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6753–6762, 2019.

[33] M. Jacob and M. Unser. Design of steerable filters for feature detection using canny-like criteria. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(8):1007–1019, Aug. 2004.

[34] Jiaju Ma, Donald Hondongwa, and Eric R. Fossum. Jot devices and the Quanta Image Sensor. In *2014 IEEE International Electron Devices Meeting*, pages 10.1.1–10.1.4, San Francisco, CA, USA, Dec. 2014. IEEE.

[35] Kiyotaka Iwabuchi and Yusuke Kameda and Takayuki Hamamoto. Image quality improvements based on motion-based deblurring for single-photon imaging. *IEEE Access*, 9, 2021.

[36] Joachim Kopp. Efficient Numerical Diagonalization of Hermitian 3 × 3 Matrices. *International Journal of Modern Physics C*, 19(03):523–548, Mar. 2008.

[37] Peter Kovesi. Image Features from Phase Congruency. *Videre: Journal of Computer Vision Research*, 1(3), 1999.

[38] Peter Kovesi. Phase Congruency Detects Corners and Edges. In *DICTA 2003*, pages 309–318, Sydney, Australia, 2003.

[39] Erwin Kreyszig, Herbert Kreyszig, and E. J. Norminton. *Advanced Engineering Mathematics*. Wiley, Hoboken, NJ, tenth edition, 2011.

[40] Ivan Laptev. On Space-Time Interest Points. *International Journal of Computer Vision*, 64(2-3):107–123, Sept. 2005.

[41] S. Lim, J.G. Apostolopoulos, and A.E. Gamal. Optical flow estimation using temporally oversampled video. *IEEE Transactions on Image Processing*, 14(8):1074–1087, Aug. 2005.

[42] Tony Lindeberg. Time-Causal and Time-Recursive Spatio-Temporal Receptive Fields. *Journal of Mathematical Imaging and Vision*, 55(1):50–88, May 2016.

[43] Tony Lindeberg. Spatio-Temporal Scale Selection in Video Data. *Journal of Mathematical Imaging and Vision*, 60(4):525–562, May 2018.

[44] Ce Liu, Antonio Torralba, William T. Freeman, Frédo Durand, and Edward H. Adelson. Motion Magnification. *ACM Transactions on Graphics (TOG)*, 24(3):519–526, 2005.

[45] Yun Liu, Ming-Ming Cheng, Xiaowei Hu, Kai Wang, and Xiang Bai. Richer Convolutional Features for Edge Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3000–3009, 2017.

[46] Jiaju Ma, Dexue Zhang, Dakota Robledo, Leo Anzagira, and Saleh Masoodian. Ultra-high-resolution quanta image sensor with reliable photon-number-resolving and high dynamic range capabilities. *Scientific Reports*, 12(1):13869, Aug. 2022.

[47] Sizhuo Ma, Shantanu Gupta, Arin C. Ulku, Claudio Bruschini, Edoardo Charbon, and Mohit Gupta. Quanta burst photography. *ACM Transactions on Graphics*, 39(4), July 2020.

[48] Sizhuo Ma, Paul Mos, Edoardo Charbon, and Mohit Gupta. Burst Vision Using Single-Photon Cameras. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2023.

[49] MATLAB. *version 9.11.0 (R2021b)*. The MathWorks Inc., Natick, Massachusetts, 2021.

[50] Francescopaolo Mattioli Della Rocca, Tarek Al Abbas, Neale A. W. Dutton, and Robert K. Henderson. A high dynamic range SPAD pixel for time of flight imaging. In *2017 IEEE SENSORS*, Glasgow, Oct. 2017. IEEE.

[51] Simone Meyer, Oliver Wang, Henning Zimmer, Max Grosse, and Alexander Sorkine-Hornung. Phase-based frame interpolation for video. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1410–1418, Boston, MA, USA, June 2015. IEEE.

[52] Oliver Monga, Rachid Deriche, Grégoire Malandain, and Jean Pierre Cocquerez. Recursive filtering and edge tracking: Two primary tools for 3D edge detection. *Image and Vision Computing*, 9(4):203–214, Aug. 1991.

[53] K Morimoto, J Iwata, M Shinohara, H Sekine, A Abdelghafar, H Tsuchiya, Y Kuroda, K Tojima, W Endo, Y Maehashi, Y Ota, T Sasago, S Maekawa, S Hikosaka, T Kanou, A Kato, T Tezuka, S Yoshizaki, T Ogawa, K Uehira, A Ehara, F Inui, Y Matsuno, K Sakurai, and T Ichikawa. 3.2 Megapixel 3D-Stacked Charge Focusing SPAD for Low-Light Imaging and Depth Sensing. In *67th Annual IEEE International Electron Devices Meeting*, Dec. 2021.

[54] M.C. Morrone and R.A. Owens. Feature detection from local energy. *Pattern Recognition Letters*, 6(5):303–313, Dec. 1987.

[55] P.J. Myerscough and M.S. Nixon. Temporal phase congruency. In *6th IEEE Southwest Symposium on Image Analysis and Interpretation, 2004.*, pages 76–79, Lake Tahoe, NV, USA, 2004. IEEE.

[56] Tampere University of Technology. BM3D software. `https://webpages.tuni.fi/foi/GCF-BM3D/index.html`.

[57] Nati Ofir, Meirav Galun, Sharon Alpert, Achi Brandt, Boaz Nadler, and Ronen Basri. On Detection of Faint Edges in Noisy Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(4):894–908, Apr. 2020.

[58] Nati Ofir, Meirav Galun, Boaz Nadler, and Ronen Basri. Fast Detection of Curved Edges at Low SNR. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 213–221, Las Vegas, NV, USA, June 2016. IEEE.

[59] A.V. Oppenheim and J.S. Lim. The importance of phase in signals. *Proceedings of the IEEE*, 69(5):529–541, 1981.

[60] Karl Pauwels and Marc M. Van Hulle. Realtime phase-based optical flow on the GPU. In *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Anchorage, AK, USA, June 2008. IEEE.

[61] Adithya K. Pediredla, Aswin C. Sankaranarayanan, Mauro Buttafava, Alberto Tosi, and Ashok Veeraraghavan. Signal

Processing Based Pile-up Compensation for Gated Single-Photon Avalanche Diodes, June 2018.

[62] Javier Portilla and Eero P Simoncelli. A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients. *International Journal of Computer Vision*, 40(1):49–71, 2000.

[63] Tal Remez, Or Litany, and Alex Bronstein. A picture is worth a billion bits: Real-time image reconstruction from dense binary threshold pixels. In *2016 IEEE International Conference on Computational Photography (ICCP)*, Evanston, IL, USA, May 2016. IEEE.

[64] David A. Schug, Glenn R. Easley, and Dianne P. O'Leary. Precise State Tracking Using Three-Dimensional Edge Detection. In Radu Balan, John J. Benedetto, Wojciech Czaja, Matthew Dellatorre, and Kasso A. Okoudjou, editors, *Excursions in Harmonic Analysis, Volume 5*, pages 89–111. Springer International Publishing, Cham, 2017. Series Title: Applied and Numerical Harmonic Analysis.

[65] Trevor Seets, Atul Ingle, Martin Laurenzis, and Andreas Velten. Motion adaptive deblurring with single-photon cameras. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2021.

[66] Oliver K. Smith. Eigenvalues of a symmetric 3×3 matrix. *Communications of the ACM*, 4(4):168, Apr. 1961.

[67] Deqing Sun, Michael Rubinstein, Charles Herrmann, David J Fleet, Fitsum Reda, and William T Freeman. Disentangling Architecture and Training for Optical Flow. In *ECCV 2022*, 2022.

[68] Varun Sundar, Sizhuo Ma, Aswin C. Sankaranarayanan, and Mohit Gupta. Single-Photon Structured Light. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17844–17854, New Orleans, LA, USA, June 2022. IEEE.

[69] Zachary Teed and Jia Deng. RAFT: Recurrent All-Pairs Field Transforms for Optical Flow. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision – ECCV 2020*, volume 12347, pages 402–419. Springer International Publishing, Cham, 2020.

[70] Nobukazu Teranishi. Required Conditions for Photon-Counting Image Sensors. *IEEE Transactions on Electron Devices*, 59(8):2199–2205, Aug. 2012.

[71] Arin Can Ulku, Claudio Bruschini, Ivan Michel Antolovic, Yung Kuo, Rinat Ankri, Shimon Weiss, Xavier Michalet, and Edoardo Charbon. A 512×512 SPAD Image Sensor With Integrated Gating for Widefield FLIM. *IEEE Journal of Selected Topics in Quantum Electronics*, 25(1):1–12, Jan. 2019.

[72] M. Unser, D. Sage, and D. Van De Ville. Multiresolution Monogenic Signal Analysis Using the Riesz–Laplace Wavelet Transform. *IEEE Transactions on Image Processing*, 18(11):2402–2418, Nov. 2009.

[73] Dor Verbin and Todd Zickler. Field of Junctions: Extracting Boundary Structure at Low SNR. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6849–6858, Montreal, QC, Canada, Oct. 2021. IEEE.

[74] Neal Wadhwa, Michael Rubinstein, Frédo Durand, and William T. Freeman. Phase-based video motion processing. *ACM Transactions on Graphics*, 32(4), July 2013.

[75] Neal Wadhwa, Michael Rubinstein, Fredo Durand, and William T. Freeman. Riesz pyramids for fast phase-based video magnification. In *2014 IEEE International Conference on Computational Photography (ICCP)*, Santa Clara, CA, USA, May 2014. IEEE.

[76] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William T Freeman. Eulerian Video Magnification for Revealing Subtle Changes in the World. *ACM Transactions on Graphics (Proc. SIGGRAPH 2012)*, 31(4), 2012.

[77] Feng Yang, Yue M. Lu, Luciano Sbaiz, and Martin Vetterli. Bits from Photons: Oversampled Image Acquisition Using Binary Poisson Statistics. *IEEE Transactions on Image Processing*, 21(4):1421–1436, Apr. 2012.