Motion Adaptive Deblurring with Single-Photon Cameras

Trevor Seets[1] seets@wisc.edu
Atul Ingle[1] ingle@wisc.edu
Martin Laurenzis[2] martin.laurenzis@isl.eu
Andreas Velten[1] velten@wisc.edu

[1]University of Wisconsin-Madison
[2]French-German Research Institute of Saint-Louis

Abstract

Single-photon avalanche diodes (SPADs) are a rapidly developing image sensing technology with extreme low-light sensitivity and picosecond timing resolution. These unique capabilities have enabled SPADs to be used in applications like LiDAR, non-line-of-sight imaging and fluorescence microscopy that require imaging in photon-starved scenarios. In this work we harness these capabilities for dealing with motion blur in a passive imaging setting in low illumination conditions. Our key insight is that the data captured by a SPAD array camera can be represented as a 3D spatio-temporal tensor of photon detection events which can be integrated along arbitrary spatio-temporal trajectories with dynamically varying integration windows, depending on scene motion. We propose an algorithm that estimates pixel motion from photon timestamp data and dynamically adapts the integration windows to minimize motion blur. Our simulation results show the applicability of this algorithm to a variety of motion profiles including translation, rotation and local object motion. We also demonstrate the real-world feasibility of our method on data captured using a 32×32 *SPAD camera.*

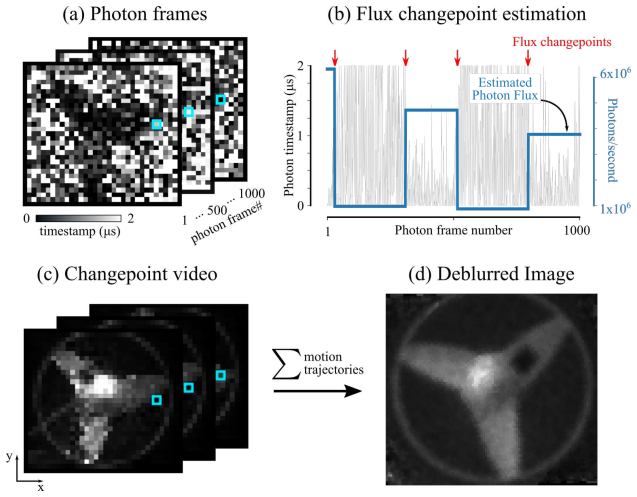


Figure 1. **Overview of our adaptive motion deblurring method.** (a) The photon frames captured from a rotating fan contain timestamps for the first detected photon in each frame. Note the smaller value of timestamps in brighter regions and vice versa. (b) Flux changepoints obtained after changepoint detection and flux estimation at pixel location highlighted in cyan in (a). Note the rapidly varying photon timestamps due to heavy-tailed nature of the raw timestamp data. (c) A changepoint video (CPV) maintains sharp edges while providing motion cues to estimate inter-frame motion trajectories; the temporal flux profile of the highlighted pixel in this CPV is the piecewise constant function shown in (b). (d) Integrating photons along estimated spatio-temporal motion trajectories generates a sharp image with high signal-to-noise ratio.

1. Introduction

When imaging dynamic scenes with a conventional camera, the finite exposure time of the camera sensor results in motion blur. This blur can be due to motion in the scene or motion of the camera. One solution to this problem is to simply lower the exposure time of the camera. However, this leads to noisy images, especially in low light conditions. In this paper we propose a technique to address the fundamental trade-off between noise and motion blur due to scene motion during image capture. We focus on the challenging scenario of capturing images in low light, with fast moving objects. Our method relies on the strengths of rapidly emerging single-photon sensors such as single-photon avalanche diodes (SPADs).

Light is fundamentally discrete and can be measured in terms of photons. Conventional camera pixels measure brightness by first converting the incident photon energy into an analog quantity (e.g. photocurrent, or charge) that is then measured and digitized. When imaging in low light levels, much of the information present in the incident photons is lost due to electronic noise inherent in the analog-to-digital conversion and readout process. Unlike conventional image sensor pixels that require 100's-1000's of photons to produce a meaningful signal, SPADs are sensitive down to individual photons. A SPAD pixel captures these photons at an instant in time, with a time resolution of hundreds of picoseconds. Each photon detection can therefore be seen as an instantaneous event, free from any motion blur. Recently, single photon sensors have been shown to be useful when imaging in low light, or equivalently, imaging at high frame rates where each image frame is photon-starved [22].

The data captured by a SPAD camera is thus quite dif-

ferent than a conventional camera: in addition to the two spatial dimensions, we also capture data on a high resolution time axis resulting in a 3D spatio-temporal tensor of photon detection events. We exploit this novel data format to deal with the noise-blur trade-off. Our key observation is that the photon timestamps can be combined dynamically, across space and time, when estimating scene brightness. If the scene motion is known *a priori*, photons can be accumulated along corresponding spatio-temporal trajectories to create an image with no motion blur. A conventional camera does not have this property because at capture time high frequency components are lost [24]. So even with known scene motion, image deblurring is an ill-posed problem.

Our method relies on dynamically changing exposure times, where each pixel can have its own set of exposure times. An overview of our approach is shown in Fig. 1. We propose using statistical changepoint detection to infer points in time where the photon flux at a given pixel changes from one steady state rate to another. This allows us to choose exposure times that adapt to scene motion. Changepoints allow us to track high contrast edges in the scene and align "high frequency" motion details across multiple image frames. We show that for the case of global motion (e.g., rotation) we can track the varying motion speeds of different scene pixels and combine photons spatially to create deblurred images. We also show that for the case of local scene motion, our method can provide robust estimates of flux changepoints around the edges of the moving objects that improves deblurring results obtained from downstream motion alignment and integration.

The locations of changepoints can be thought of as a spike-train generated by a neuromorphic event camera [9], but with three key differences: First, unlike an event camera, our method preserves original intensity information for each pixel. Second, the numbers and locations of events, chosen by our algorithm, adapt to each pixel, without the need for a hard-coded change threshold. Third, direct photon measurements are inherently more sensitive, less noisy, and have a higher time resolution than conventional cameras. Current cameras create an analog approximation of the quantized photon stream (usually a charge level in a capacitor) and then measure and digitize that analog quantity. This process introduces noise and does not take advantage of the inherent quantized properties of the photons. We therefore expect photon counting methods to have higher sensitivity, accuracy, and temporal resolution than analog cameras, when imaging in low light scenarios.

Single-Photon Sensitive Cameras for Machine Vision? Currently, single-photon sensitive image sensors are quite limited in their spatial resolution compared to conventional complementary metal–oxide–semiconductor (CMOS) and charge-coupled device (CCD) image sensors. However, single-photon image sensing is a rapidly developing field; recent work has demonstrated the feasibility of making megapixel resolution single-photon sensitive camera arrays [23, 12]. Moreover, silicon SPAD arrays are amenable to manufacturing at scale using the same photolithographic fabrication techniques as conventional CMOS image sensors. This means that many of the foundries producing our cellphone camera sensors today could also make SPAD array sensors at similar cost. Other current limitations of SPAD sensors are the small fraction of chip area sensitive to light (fill factor) resulting from the large amount of additional circuitry that is required in each pixel. Emerging 3D stacking technologies could alleviate this problem by placing the circuitry behind the pixel.

Limitations Our experimental demonstration uses a first generation commercial SPAD array that is limited to a 32×32 pixel spatial resolution. More recently 256×256 pixel commercial arrays have become available [5]. Although current SPAD cameras cannot compete with the image quality of commercial CMOS and CCD cameras, with the rapid development of SPAD array technology, we envision our techniques could be applied to future arrays with spatial resolution similar to that of existing CMOS cameras.

Contributions:

- We introduce the notion of *flux changepoints*; these can be estimated using an off-the-shelf statistical changepoint detection algorithm.
- We show that flux changepoints enable inter-frame motion estimation while preserving edge details when imaging in low light and at high speed.
- We show experimental demonstration using data acquired from a commercially available SPAD camera.

2. Related Work

Motion Deblurring Motion deblurring for existing cameras can be performed using blind deconvolution [21]. Adding a fast shutter ("flutter shutter") sequence can aid this deconvolution task [24]. We push the idea of a fluttered shutter to the extreme limit of individual photons: our image frames consist of individual photon timestamps allowing dynamic adaptation of sub-exposure times for the shutter function. Our deblurring method is inspired by burst photography pipelines used for conventional CMOS cameras. Burst photography relies on combining frames captured with short exposure times [16], resulting in large amounts of data that suffer from added readout noise. Moreover, conventional motion deblurring methods give optimal performance when the exposure time is matched to the true motion speed which is not known a priori.

Event-based Vision Sensors Event cameras directly capture temporal changes in intensity instead of capturing scene brightness [9]. Although it is possible to create intensity images from event data in post-processing [3], our

method natively captures scene intensities at single-photon resolution: the "events" in our sensing modality are individual photons. The notion of using photon detections as "spiking events" has also been explored in the context of biologically inspired vision sensors [30, 1]. We derive flux changepoints from the high-resolution photon timestamp data. Due to the single-photon sensitivity, our method enjoys lower noise in low light conditions, and pixel-level adaptivity for flux changepoint estimation.

Deblurring Methods for Quanta Image Sensors There are two main single-photon detection technologies for passive imaging: SPADs [20, 17] and quanta image sensors (QIS) [8]. Although our proof-of-concept uses a SPAD camera, our idea of adaptively varying exposure times can be applied to QIS data as well. Existing motion deblurring algorithms for QIS [13, 14, 22] rely on a fixed integration window to sum the binary photon frames. However, the initial step of picking the size of this window requires *a priori* knowledge about the motion speed and scene brightness. Our technique is therefore complementary to existing motion deblurring algorithms. For example, our method can be considered as a generalization of the method in [18] which uses two different window sizes. Although we use a classical correlation-based method for motion alignment, the sequence of flux changepoints generated using our method can be used with state-of-the-art align-and-merge algorithms [22] instead.

3. SPAD Image Formation Model

SPADs are most commonly used in synchronization with an active light source such as a pulsed laser for applications including LiDAR and fluorescence microscopy. In contrast, here we operate the SPAD passively and only collect ambient photons from the scene. In this section, we describe the imaging model for a single SPAD pixel collecting light from a fixed scene point whose brightness may vary as a function of time due to camera or scene motion.

3.1. Pixelwise Photon Flux Estimator

Our imaging model assumes a frame-based readout mechanism: each SPAD pixel in a *photon frame* stores at most one timestamp of the first captured ambient photon. This is depicted in Fig. 1(a). Photon frames are read out synchronously from the entire SPAD array; the data can be read out quite rapidly allowing photon frame rates of 100s of kHz (frame times on the order of a few microseconds).

Let N_{pf} denote the number of photon frames, and T_{pf} be the frame period. So the total exposure time is given by $T = N_{\text{pf}} T_{\text{pf}}$. We now focus on a specific pixel in the SPAD array. In the i^{th} photon frame ($1 \leq i \leq N_{\text{pf}}$), the output of this pixel is tagged with a photon arrival timestamp t_i relative to the start of that frame.* If no photons are detected during a photon frame, we assume $t_i = T_{\text{pf}}$.

Photon arrivals at a SPAD pixel can be modeled as a Poisson process [15]. It is possible to estimate the intensity of this process (i.e. the perceived brightness at the pixel) from the sequence of photon arrival times [20, 17]. The maximum likelihood brightness estimator $\widehat{\Phi}$ for the true photon flux Φ is given by [20]:

$$\widehat{\Phi} = \frac{\sum_{i=1}^{N_{\text{pf}}} \mathbf{1}(t_i \neq T_{\text{pf}})}{q \sum_{i=1}^{N_{\text{pf}}} t_i} \tag{1}$$

where q is the sensor's photon detection efficiency and $\mathbf{1}$ denotes a binary indicator variable.

This equation assumes that the pixel intensity does not change for the exposure time T. The assumption is violated in case of scene motion. If we can determine the temporal locations of intensity changes, we can use still use Eq. (1) to estimate a time-varying intensity profile for each pixel. In the next section we introduce the idea of *flux changepoints* and methods to locate them with photon timestamps.

3.2. Flux Changepoints

Photon flux at a given pixel may change over time in case of scene motion. This makes it challenging to choose pixel exposure times *a priori*: ideally, for pixels with rapidly varying brightness, we should use a shorter exposure time and vice versa. We propose an algorithm that dynamically adapts to brightness variations and chooses time-varying exposure times on a per pixel basis. Our method relies on locating temporal change locations where the pixel's photon flux has a large change: we call these *flux changepoints*.

In general, each pixel in the array can have different numbers and locations of flux changepoints. For pixels that maintain constant brightness over the entire capture period (e.g. pixels in a static background), there will be no flux changepoints detected and we can integrate photons over the entire capture time T. For pixels with motion, we assume that the intensity between flux changepoints is constant, and we call these regions *virtual exposures*. Photons in each virtual exposure are aggregated to create a piecewise constant flux estimate. The length of each virtual exposure will depend on how quickly the local pixel brightness varies over time, which in turn, depends on the true motion speed.

An example is shown in Fig. 1(b). Note that the photon timestamps are rapidly varying and extremely noisy due to a heavy-tailed exponential distribution. Our changepoint detection algorithm detects flux changepoints (red arrows) and estimates a piecewise constant flux waveform for the pixel (blue plot). In the example shown, five different flux levels are detected.

*In practice, due to random timing jitter and finite resolution of timing electronics, this timestamp is stored as a discrete fixed-point value. The SPAD camera used in our experiments has a 250 picosecond discretization.

3.2.1 Changepoint Detection

Detecting changepoints is a well studied problem in the statistics literature [4, 26]. The goal is to split a time-series of measurements into regions with similar statistical properties. Here, our time series is a sequence of photon timestamps at a pixel location, and we would like to find regions where the timestamps have the same mean arrival rate, i.e., the photon flux during this time is roughly constant.

Using the sequence of photon arrival times $\{t_i\}_{i=1}^{N_{\text{pf}}}$, we wish to find a subset $\{t_{l_1}, \ldots, t_{l_L}\}$ representing the flux changepoints. For convenience, we let the first and the last flux changepoint be the first and the last photons captured by the pixel ($l_1 := 1$ and $l_L = N_{\text{pf}}$).

Offline changepoint detection is non-causal, i.e., it uses the full sequence of photon timestamps for a pixel to estimate flux changepoints at that pixel. In Suppl. Note S.1 we describe an online changepoint detection method that can be used for real-time applications.

For a sequence of photon timestamps, we solve the following optimization problem [29] (see Suppl. Note S.1):

$$(l_i^*)_{i=1}^{L} = \underset{l_1,..l_L}{\arg\min} \sum_{i=1}^{L-1} \left[-\log(\widehat{\Phi}_i) \sum_{j=l_i}^{l_{i+1}} \mathbf{1}(t_j \neq T_{\text{pf}}) \right] + \lambda L \tag{2}$$

where $\widehat{\Phi}_i$ is the photon flux estimate given by Eq. (1) using only the subset of photons between times t_{l_i} and $t_{l_{i+1}}$. Here λ is the penalty term that prevents overfitting by penalizing the number of flux changepoints. For larger test images we use the BOTTOMUP algorithm [28] which is approximate but has a faster run time. For lower resolution images we use the Pruned Exact Linear Time (PELT) algorithm [19] which gives an exact solution but runs slower. See Suppl. Section S.4 for results using the BOTTOMUP algorithm.

Flux Changepoints for QIS Data The cost function in Eq. (2) applies to photon timestamp data, but the same idea of adaptive flux changepoint detection can be used with QIS photon count data as well. A modified cost function that uses photon counts (0 or 1) is derived in Suppl. Note S.1.2.

3.2.2 Single-Pixel Simulations

There is a trade-off between contrast and motion speed when trying to detect changepoints. For example, it is harder to detect the flux change with a fast moving object with a lower contrast with respect to its background. To evaluate this trade-off, we simulate a single SPAD pixel for 800 photon frames with a time varying flux signal with a randomly placed pulse wave. We use the pulse width as a proxy for motion speed, and vary the contrast by varying the ratio of the pulse height. We measure the absolute difference between the number of changepoints detected by our algorithm and the true number of flux changes (which is exactly 2 for a single pulse). We call this "annotation error."

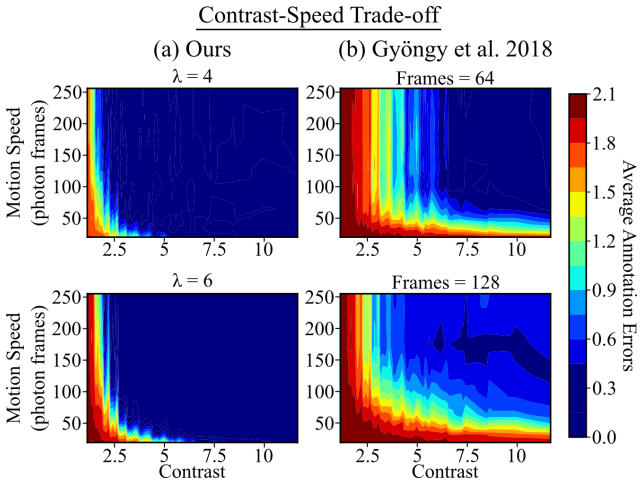


Figure 2. **Contrast-Speed Trade-off.** We simulate a single SPAD pixel measuring a pulse-shaped photon flux signal to analyze the contrast-speed trade-off when detection motion. Our method (a) based on PELT changepoint detection adapts to a wider range of contrast and speed combinations than the comparison method (b) from Gyöngy et al. [14].

Fig. 2 shows the annotation errors for different values of contrast and motion speeds for two different algorithms. For each set of contrasts and motion speeds we display the average number of annotation errors over 120 simulation runs. We use the PELT algorithm to detect flux changepoints. For comparison, we also show the fixed windowing approach used by Gyöngy et al. [14, 2]. The changepoint detection algorithm is able to adapt to a wider range of contrasts and speeds.

4. Pixel-Adaptive Deblurring

4.1. Changepoint Video

Using methods described in the previous section we locate flux changepoints for each pixel in the image. The changepoints for each pixel represent when a new virtual exposure starts and stops; photons within each virtual exposure can be used to estimate photon flux using Eq. (1). We call this collection of piecewise constant functions over all pixels in the array the *changepoint video* (CPV).

The CPV does not have an inherent frame rate; since each pixel has a continuous time piecewise constant function, it can be sampled at arbitrarily spaced time instants in $[0, T]$ to obtain any desired frame rate. We sample the CPV at non-uniform time intervals using the following criterion. Starting with the initial frame sampled at $t = 0$, we sample the subsequent frames at instants when at least 1% of the pixel values have switched to a new photon flux. This leads to a variable frame rate CPV that adapts to changes in scene

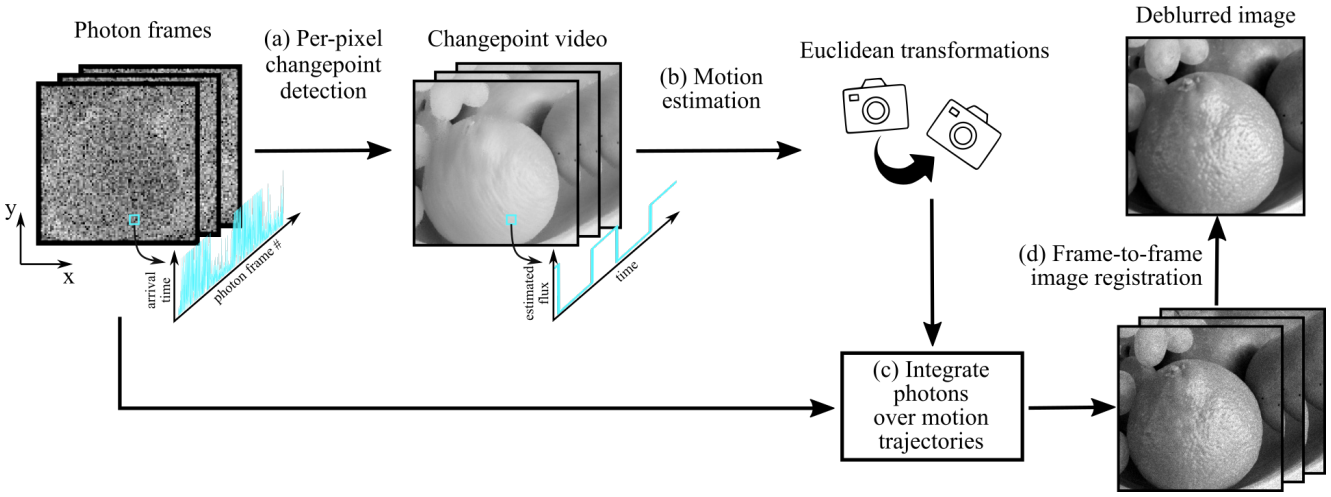


Figure 3. **Changepoint video deblurring**. Photon timestamps for each pixel location in the 3D timestamp tensor are analyzed for changepoints (a) to generate a changepoint video. Pixels may have different numbers of changepoints, and hence, dynamically changing frame-rate over time. We estimate motion between successive frames of the changepoint video (b) and estimate motion parameters. Finally, using this motion estimate, we sum photons (c) from the photon frames along motion trajectories resulting in a deblurred video. Frames in the deblurred video are registered and summed (d) to obtain the final deblurred image. (Original images from FreeImages.com)

velocity: scenes with fast moving objects will have a higher average CPV frame rate.

The CPV preserves large brightness changes in a scene. For example, the edges of a bright object moving across a dark background remain sharp. However, finer texture details within an object may appear blurred.

4.2. Spatio-Temporal Motion Integration

We use global motion cues from the CPV and then combine the motion estimates with the photon frames to create a deblurred video. The overall flowchart is shown in Fig. 3.

We use a correlation-based image registration algorithm to find the motion between consecutive frames in the CPV. For each location $\mathbf{p} = [x, y]$ in the i^{th} CPV frame, we assume it maps to a location $\mathbf{p}'$ in $(i+1)^{\text{st}}$ CPV frame. We assume that the mapping is a linear transformation:

$$A\mathbf{p} = \mathbf{p}'. \tag{3}$$

We constrain A to represent planar Euclidean motion (rotation and translation). Let θ be the amount of rotation centered around a point $[r_x, r_y]$ and let $[\tau_x, \tau_y]$ be the translation vector. Then A has the form:

$$A = \begin{bmatrix} \cos\theta & \sin\theta & r_x(1-\cos\theta)+r_y\sin\theta+\tau_x \\ -\sin\theta & \cos\theta & r_y(1-\cos\theta)-r_x\sin\theta+\tau_y \\ 0 & 0 & 1 \end{bmatrix}. \tag{4}$$

We apply the enhanced correlation coefficient maximization algorithm [7] to estimate the transformation matrix A for consecutive pairs of CPV frames $i \rightarrow i+1$. A sequence of frame-to-frame linear transformations generates arbitrarily shaped global motion trajectories. We aggregate the original photon frame data along these estimated spatio-temporal motion trajectories.

We assume that the rotation center, $[r_x, r_y]$, is the middle of the image and a change in rotation center can be modeled as a translation. We solve for θ, τ_x, and τ_y which we linearly interpolate. Then using the interpolated motion parameters and Eq. (3), we align all photon frames corresponding to the time interval between CPV frames $i \rightarrow i+1$ and sum these frames to get a photon flux image by using Eq. (1) at each pixel. This generates a motion deblurred video with the same frame rate as the CPV, but with finer textures preserved as shown in Fig. 3.

If the final goal is to obtain a single deblurred image, we repeat the steps described above on consecutive frames in the deblurred video, each time decreasing the frame rate by a factor of 2, until eventually we get a single image. This allows us to progressively combine photons along spatial-temporal motion trajectories to increase the overall signal to noise ratio (SNR) and also preserve high frequency details that were lost in the CPV.

Our method fails when motion becomes too large to properly align images, especially at low resolutions. It can also fail when not many flux changepoints are detected, this will occur mainly due to a lack of photons per pixel of movement. In the worst case, if not enough changepoints are detected, the result of the algorithm will look similar to a single long exposure image.

The method of aligning and adding photon frames is similar to contrast maximization algorithms used for event cameras [10]. However, unlike event camera data, our method relies on the CPV which contains both intensity and flux changepoints derived from single-photon timestamps.

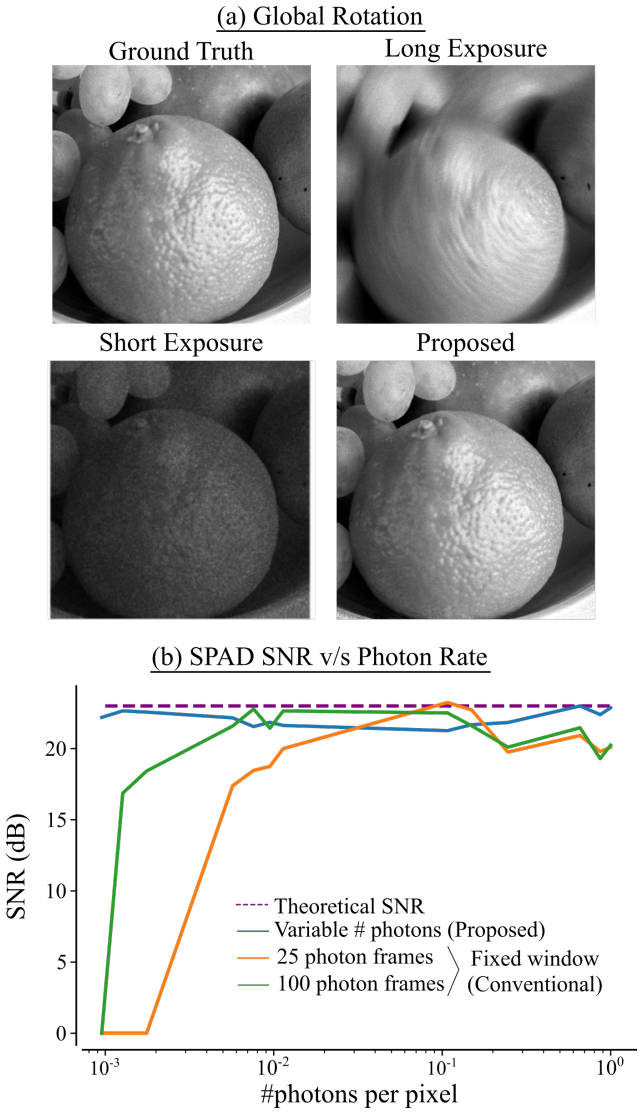


Figure 4. **Comparison of SNR for different deblurring window sizes.** (a) A ground truth video of a rotating orange used for creating simulated SPAD photon frames. A long exposure image is quite blurry while a short exposure image is very noisy. Our deblurring algorithm strikes a balance between noise and blur to get a sharp high-quality image. (b) By varying the brightness of the orange scene, we compare the simulated SNR using our method compared to a conventional method with fixed windows. Notice the proposed method stays above 20 dB at all photon rates, while the fixed photon window SNRs decrease in the low photon count regime. (Original image from FreeImages.com)

Handling Multiple Moving Objects To handle multiple moving objects on a static background, we implement a method similar to Gyögny et al. [14] and combine it with our CPV method. We cluster the changepoints for different objects using a density-based spatial clustering algorithm (DBSCAN) [6]. For each cluster, we then create a bounding box, isolating different moving objects. We then apply our motion deblurring algorithm on each object individually, before stitching together each object with the areas that are not moving in the CPV. The clustering step also denoises by rejecting flux changepoints not belonging to a cluster.

4.3. Simulations

Starting with a ground truth high resolution, high frame rate video, we scale the video frames to units of photons per second and generate photon frames using exponentially distributed arrival times. We model a photon frame readout SPAD array with 8000 bins and bin width of 256 ps.

We first simulate a rotating orange by applying successive known rigid transformations to an image and generating SPAD data by scaling the transformed images between 10^4 and 10^8 photons per second. We rotate the image by 0.1° for every 10 generated photons for a total of 1000 photons. We use the BOTTOMUP algorithm [29] with $\lambda = 5$ for the changepoint detection step. The results are shown in Fig. 4(a). Our method captures sharp details on the orange skin while maintaining high quality.

Fig. 4(b) shows quantitative comparisons of SNR for different deblurring methods. The conventional approach to deblur photon data uses a fixed frame rate; we use two different window lengths for comparison. We compute the SNR of the deblurred imaging using the $\ell 2$ (root-mean-squared) distance from the ground truth to and repeat this over a range of photon flux levels. We keep the total number of photons captured approximately constant by extending the capture time for darker flux levels. Our method dynamically adapts to motion and lighting so we are able to reconstruct with high SNR even in photon starved regimes where the SNR of the fixed window methods degrades rapidly.

To simulate multi-object motion, we captured a ground truth video of two toy cars rolling down a ramp at different speeds. The video frame pixels are then scaled between 10^5 and 10^6 photons per second and a total of 690 photon frames are generated. A bright slow moving car has a contrast of 1.2 with respect to the background, and moves 48 pixels over the duration of video. The dark car has a contrast of 5.0 with the background, and moves 143 pixels. We use the PELT algorithm [19] with $\lambda = 6$ for the changepoint detection step. We use $\epsilon = 7.5$ and MinPts $= 40$ in the DBSCAN clustering algorithm. The resulting deblurred images are shown in Fig. 5. Observe that the method of [14] blurs out the low contrast white car in the back. Our method assigns dynamically changing integration windows extracted from the CPV to successfully recover both cars simultaneously with negligible motion blur. The changepoint clusters used for segmenting cars from the static background in our method are are shown in Fig. 6.

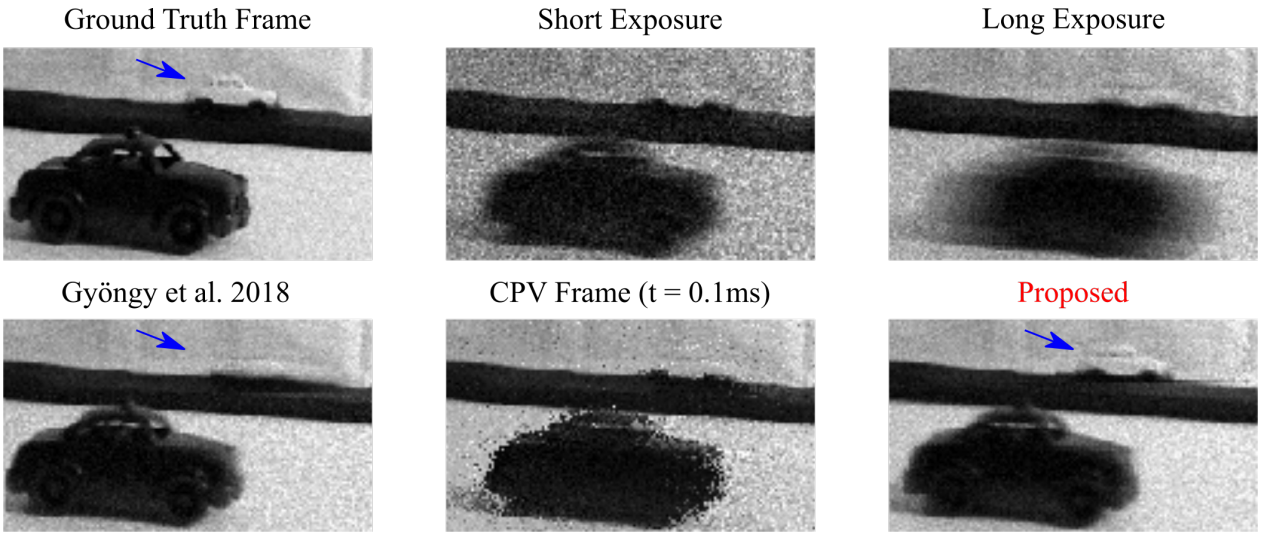


Figure 5. **Simulated Motion Deblurring for Multiple Moving Objects.** We simulate SPAD data from video of toy cars, a fast moving black car and slow moving white car. (Top row) A sample frame from the ground truth frame sequence is shown. The short and long exposure images show the results of using integrating the first 75 and 250 photon frames, respectively. Notice that the short exposure preserves the black car while the white car is quite noisy, on the other hand, the long average blurs the black car but preserves details of the white one better. (Bottom row) The method of Gyögny et al. [14], fails to reconstruct the white car (blue arrow) due to its low contrast. A sample frame from our changepoint video shows both moving cars. Finally, our deblurring algorithm is able to reconstruct both the black and white car with negligible motion blur.

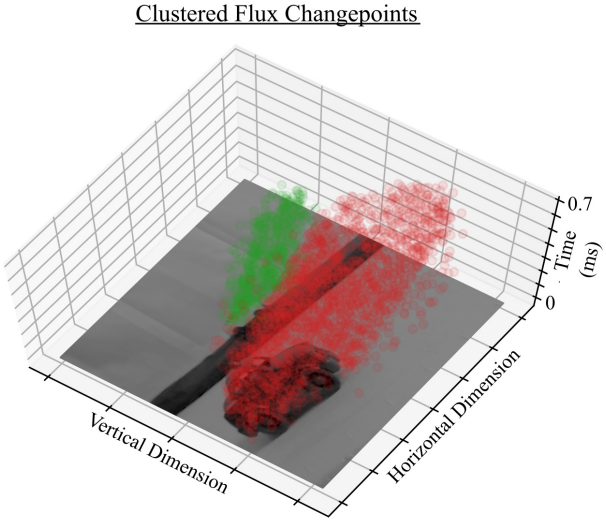


Figure 6. **Clustered Flux Changepoints** Two clusters of flux changepoints are detected using frames from the changepoint video for the toy car scene. These changepoint clusters are used for segmenting the moving cars from the static background.

5. Experiments

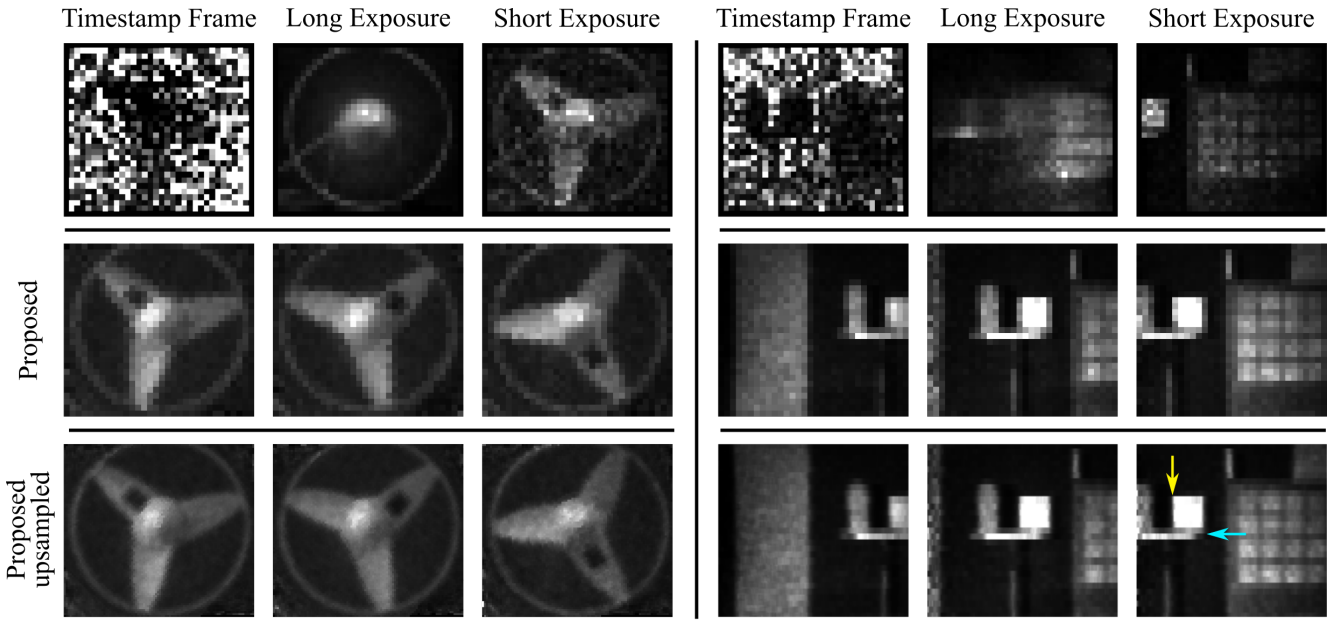
We validate our method using experimental data captured using a 32×32 InGaAs SPAD array from Princeton Lightwave Inc., USA. The pixels are sensitive to near infrared and shortwave infrared (900 nm–1.6 µm wavelengths). The SPAD array operates in a frame readout mode at 50,000 photon frames per second. Each photon frame exposure window is 2 µs and sub-divided into 8000 bins, giving a temporal resolution of 250 ps per bin [27].

For this low resolution experimental data, we zero-order hold upsample both the timestamp frames and the CPV before step (b) in Fig. 3 in the spatial dimensions. We then use upsampled photon frames for step (c) resulting in sharper images. Upsampling before motion integration allows photons that are captured in the same pixel to land in a larger space during the motion integration step.

Fig. 7 shows results from two different scenes. The "fan" scene shows the performance of our algorithm with fast rotation[†]. The optimal exposure time in this case depends on the rotation speed. Our method preserves the details of the fan blades including the small black square patch on one of the fan blades. The "checkerboard" scene shows deblurring result with purely horizontal global motion. Note that our method is able to resolve details such as the outlines of the squares on the checkerboard.

The last row in Fig. 7 shows the upsampled results. Benefits of upsampling are restricted to the direction of motion. The "fan" dataset is upsampled 9× compared to the original resolution. The "checkerboard" dataset is upsampled 4×; this is because the motion is limited to the horizontal dimension. Note that some of the details in the vertical edges are sharp, but horizontal edges remain blurry.

[†]The background behind the fan is covered with absorbing black felt material. This allows us to treat the data as having global rotation, because there is hardly any light captured from the background.



(a) "Fan" Scene (b) "Checkerboard" Scene

Figure 7. **Experimental Results.** The first row shows a single raw data frame from the photon timestamp tensor; each photon has an associated timestamp with a 250 ps bin resolution. Integration over a long exposure (2 ms for the fan and 0.2 ms for checkerboard scene) this gives a low noise but blurry result. Using a short exposure time (40 µs for both scenes) produces very noisy results. The second row shows a sequence of three frames from the final deblurred video. The third row shows the result with upsampling. Note that in the checkerboard scene due to purely horizontal motion, some vertical edges (yellow arrow) are sharper but not the horizontal edges (cyan arrow). See supplementary video results.

6. Discussion and Future Work

Euclidean Motion Assumption In the case of camera shake, most of the motion will come from small rotations in the camera that result in 2D translations and rotations in the image frames. The short distance translations of shaking camera would cause translations in the frames that are similar in nature, but smaller in magnitude.

Translation of the camera over larger distances would result in parallax while motion within the scene can result in more complex changes. In these cases our model captures scene changes only approximately. It applies to frame to frame motion over short time-scales and limited to regions in the scene. Ideas for extending our model to deal with larger motion will be the subject of future work.

Dealing with Local Motion The techniques presented in this paper assume multiple moving objects exhibiting euclidean motion, with no occlusions. We can extend our approach to more complex motions. We can use a patch-wise alignment and merging methods to deal with more complex local motion and occlusions [16, 22].

Deblurring algorithms developed for event camera data can be adapted to SPAD data, because the flux changepoints represent changes in brightness similar to the output of event cameras. Current event camera algorithms are able to recover complex motion in scenes [10, 25], and they could be improved with a fusion based approach where image intensity information is also available [11].

Data Compression With increasing number of pixels, processing photon frames with high spatio-temporal resolution will be quite resource intensive. Our online changepoint method takes some initial steps towards a potential real-time implementation. The CPV can be used for video compression with variable frame rate: by tuning the regularization parameter of the changepoint detection algorithm a tradeoff between image fidelity and data rate can be achieved.

Acknowledgments This material is based upon work supported by the Department of Energy / National Nuclear Security Administration under Award Number DE-NA0003921, the National Science Foundation (GRFP DGE-1747503, CAREER 1846884), DARPA (REVEAL HR0011-16-C-0025) and the Wisconsin Alumni Research Foundation.

References

[1] S. Afshar, T. J. Hamilton, L. Davis, A. Van Schaik, and D. Delic. Event-based processing of single photon avalanche diode sensors. *IEEE Sensors Journal*, pages 1–1, 2020.

[2] Alan Agresti and Brent A. Coull. Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician*, 52(2):119–126, 1998.

[3] Patrick Bardow, Andrew J Davison, and Stefan Leutenegger. Simultaneous optical flow and intensity estimation from an event camera. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 884–892, 2016.

[4] Michèle Basseville, Igor V Nikiforov, et al. *Detection of abrupt changes: theory and application*, volume 104. Prentice Hall Englewood Cliffs, 1993.

[5] Vinit Dhulla, Sapna S Mukherjee, Adam O Lee, Nanditha Dissanayake, Booshik Ryu, and Charles Myers. 256 x 256 dual-mode cmos spad image sensor. In *Advanced Photon Counting Techniques XIII*, volume 10978, page 109780Q. International Society for Optics and Photonics, 2019.

[6] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD'96, page 226–231. AAAI Press, 1996.

[7] Georgios D Evangelidis and Emmanouil Z Psarakis. Parametric image alignment using enhanced correlation coefficient maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(10):1858–1865, 2008.

[8] Eric R Fossum, Jiaju Ma, Saleh Masoodian, Leo Anzagira, and Rachel Zizza. The quanta image sensor: Every photon counts. *Sensors*, 16(8):1260, 2016.

[9] G. Gallego, T. Delbruck, G. M. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. Davison, J. Conradt, K. Daniilidis, and D. Scaramuzza. Event-based vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 2020.

[10] Guillermo Gallego, Henri Rebecq, and Davide Scaramuzza. A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation. *Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 06 2018.

[11] Daniel Gehrig, Henri Rebecq, Guillermo Gallego, and Davide Scaramuzza. EKLT: Asynchronous, photometric feature tracking using events and frames. *Int. J. Comput. Vis.*, 2019.

[12] Abhiram Gnanasambandam, Omar Elgendy, Jiaju Ma, and Stanley H Chan. Megapixel photon-counting color imaging using quanta image sensor. *Optics express*, 27(12):17298–17310, 2019.

[13] I. Gyöngy, T. A. Abbas, N. Dutton, and R. Henderson. Object tracking and reconstruction with a quanta image sensor. In *Proceedings of the International Image Sensor Workshop*, 2017.

[14] Istvan Gyongy, Neale AW Dutton, and Robert K Henderson. Single-photon tracking for high-speed vision. *Sensors*, 18(2):323, 2018.

[15] Samuel W. Hasinoff. *Photon, Poisson Noise*, pages 608–610. Springer US, Boston, MA, 2014.

[16] 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 (TOG)*, 35(6):1–12, 2016.

[17] Atul Ingle, Andreas Velten, and Mohit Gupta. High flux passive imaging with single-photon sensors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6760–6769, 2019.

[18] Gyongy Istvan, Dutton Neale, Luca Parmesan, Davies Amy, Saleeb Rebecca, Duncan Rory, Rickman Colin, Dalgarno Paul, and Robert K Henderson. Bit-plane processing techniques for low-light, high speed imaging with a spad-based qis. In *International Image Sensor Workshop*, pages 1–4, 2015.

[19] R. Killick, P. Fearnhead, and I. A. Eckley. Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107(500):1590–1598, 2012.

[20] Martin Laurenzis. Single photon range, intensity and photon flux imaging with kilohertz frame rate and high dynamic range. *Optics Express*, 27(26):38391–38403, 2019.

[21] Anat Levin, Peter Sand, Taeg Sang Cho, Frédo Durand, and William T Freeman. Motion-invariant photography. *ACM Transactions on Graphics (TOG)*, 27(3):1–9, 2008.

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

[23] Kazuhiro Morimoto, Andrei Ardelean, Ming-Lo Wu, Arin Can Ulku, Ivan Michel Antolovic, Claudio Bruschini, and Edoardo Charbon. Megapixel time-gated spad image sensor for 2d and 3d imaging applications. *Optica*, 7(4):346–354, 2020.

[24] Ramesh Raskar, Amit Agrawal, and Jack Tumblin. Coded exposure photography: motion deblurring using fluttered shutter. In *ACM SIGGRAPH 2006 Papers*, pages 795–804. ACM, 2006.

[25] Timo Stoffregen, Guillermo Gallego, Tom Drummond, Lindsay Kleeman, and Davide Scaramuzza. Event-based motion segmentation by motion compensation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 7244–7253, 2019.

[26] Alexander Tartakovsky, Igor Nikiforov, and Michele Basseville. *Sequential analysis: Hypothesis testing and changepoint detection*. CRC Press, 2014.

[27] Princeton Lightwave/AMS Technologies. *32 x 32 Geiger-mode Avalanche Photodiode (GmAPD) Camera*, 2012 (accessed June 20, 2020). `http://www.amstechnologies.com/fileadmin/amsmedia/downloads/4796_gmapdcameradatasheet.pdf`.

[28] C. Truong. *ruptures Python Package*, 2017 (accessed June 20, 2020). `https://ctruong.perso.math.cnrs.fr/ruptures-docs/build/html/index.html`.

[29] Charles Truong, Laurent Oudre, and Nicolas Vayatis. Selective review of offline change point detection methods. *Signal Processing*, 167:107299, 2020.

[30] Lin Zhu, Siwei Dong, Jianing Li, Tiejun Huang, and Yonghong Tian. Retina-like visual image reconstruction via spiking neural model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1438–1446, 2020.