Paper accepted at 14th Asian Conf. on Computer Vision (ACCV) Perth, 2018.

Continuous-time Intensity Estimation Using Event Cameras ⋆

Cedric Scheerlinck[1], Nick Barnes[1,2], and Robert Mahony[1]

[1] The Australian National University, Canberra ACT 2600, Australia
[2] Data61, CSIRO, https://www.data61.csiro.au/

Abstract. Event cameras provide asynchronous, data-driven measurements of local temporal contrast over a large dynamic range with extremely high temporal resolution. Conventional cameras capture low-frequency reference intensity information. These two sensor modalities provide *complementary* information. We propose a computationally efficient, asynchronous filter that continuously fuses image frames and events into a single high-temporal-resolution, high-dynamic-range image state. In absence of conventional image frames, the filter can be run on events only. We present experimental results on high-speed, high-dynamic-range sequences, as well as on new ground truth datasets we generate to demonstrate the proposed algorithm outperforms existing state-of-the-art methods.

Code, Datasets and Video:
https://cedric-scheerlinck.github.io/continuous-time-intensity-estimation

1 Introduction

Event cameras respond asynchronously to changes in scene-illumination at the pixel level, offering high-temporal-resolution information over a large dynamic range [8, 12, 17, 22, 27]. Conventional cameras typically acquire intensity image frames at fixed time-intervals, generating temporally-sparse, low-dynamic-range image sequences. Fusing image frames with the output of event cameras offers the opportunity to create an image state infused with high-temporal-resolution, high-dynamic-range properties of event cameras. This image state can be queried locally or globally at any user-chosen time-instance(s) for computer vision tasks such motion estimation, object recognition and tracking.

Event cameras produce events; discrete packets of information containing the timestamp, pixel-location and polarity of a brightness change [8, 22]. An event is triggered each time the change in log intensity at a pixel exceeds a preset threshold. The result is a continuous, asynchronous stream of events that encodes non-redundant information about local brightness changes. The DAVIS

⋆ This research was supported by an Australian Government Research Training Program Scholarship, and the Australian Research Council through the Australian Centre of Excellence for Robotic Vision CE140100016.

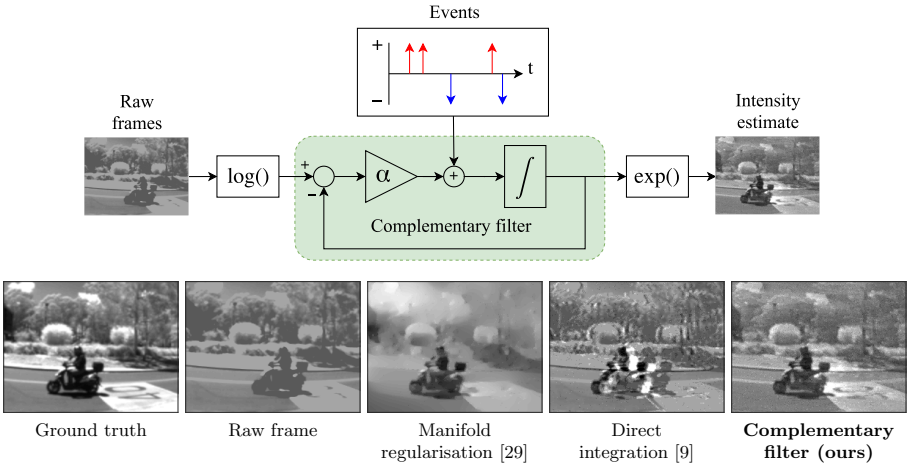

Fig. 1. The complementary filter takes image frames and events, and produces a high-dynamic-range, high-temporal-resolution, continuous-time intensity estimate.

camera [8] also provides low-frequency, full-frame intensity images in parallel. Alternative event cameras output direct brightness measurements with every event [17, 27], and may also allow user-triggered, full-frame acquisition.

Due to high availability of contrast event cameras that output polarity (and not absolute brightness) with each event, such as the DAVIS, many researchers have tackled the challenge of estimating image intensity from contrast events [3–5, 9, 29]. Image reconstruction algorithms that operate directly on the event stream typically perform spatio-temporal filtering [5,29], or take a spatio-temporal window of events and convert them into a discrete image frame [3,4]. Windowing incurs a trade-off between length of time-window and latency. SLAM-like algorithms [11, 20, 21, 28] maintain camera-pose and image gradient (or 3D) maps that can be upgraded to full intensity via Poisson integration [1, 2], however, so far these methods only work well for static scenes. Another image reconstruction algorithmic approach is to combine image frames directly with events [9]. Beginning with an image frame, events are integrated to produce inter-frame intensity estimates. The estimate is reset with every new frame to prevent growth of integration error.

In this paper, we present a continuous-time formulation of event-based intensity estimation using complementary filtering to combine image frames with events (Fig. 1). We choose an asynchronous, event-driven update scheme for the complementary filter to efficiently incorporate the latest event information, eliminating windowing latency. Our approach does not depend on a motion-model, and works well in highly dynamic, complex environments. Rather than reset the intensity estimate with arrival of a new frame, our formulation retains the high-dynamic-range information from events, maintaining an image state with greater temporal resolution and dynamic range than the image frames. Our method also

works well on a pure event stream without requiring image frames. The result is a continuous-time estimate of intensity that can be queried locally or globally at any user-chosen time.

We demonstrate our approach on datasets containing image frames and an event stream available from the DAVIS camera [8], and show that the complementary filter also works on a pure event stream without image frames. If available, synthetic frames reconstructed from events via an alternative algorithm can also be used as input to the complementary filter. Thus, our method can be used to augment any intensity reconstruction algorithm. Our approach can also be applied in a setup with any conventional camera co-located with an event camera. Additionally, we show how an adaptive gain can be used to improve robustness against under/overexposed image frames.

In summary, the key contributions of the paper are;

- a continuous-time formulation of event-based intensity estimation,
- a computationally simple, asynchronous, event-driven filter algorithm,
- a methodology for pixel-by-pixel adaptive gain tuning.

We also introduce a new ground truth dataset for reconstruction of intensities from combined image frame and event streams, and make it publicly available. Sequences of images taken on a high-speed camera form the ground truth. We retain full frames at 20Hz, and convert the inter-frame images to an event stream. We compare state-of-the-art approaches on this dataset.

The paper is organised as follows: Section 2 describes related works. Section 3 summarises the mathematical representation and notation used, and characterises the full continuous-time solution of the proposed filter. Section 4 describes asynchronous implementation of the complementary filter, and introduces adaptive gains. Section 5 shows experimental results including the new ground truth dataset, and high-speed, high-dynamic-range sequences from the DAVIS. Section 6 concludes the paper.

2 Related Works

Event cameras such as the DVS [22] and DAVIS [8] provide asynchronous, data-driven contrast events, and are widely popular due to their commercial availability. Alternative cameras such as ATIS [27] and CeleX [17, 18] are capable of providing absolute brightness with each event, but are not commercially available at the time of writing. Estimating image intensity from contrast events is important because it grants computer vision researchers a readily available high-temporal-resolution, high-dynamic-range imaging platform that can be used for tasks such as face-detection [4], SLAM [11, 20, 21], or optical flow estimation [3].

Image reconstruction from events is typically done by processing a spatio-temporal window of events [3, 4]. Barua et al. [4] learn a patch-based dictionary from simulated event data, then use the learned dictionary to reconstruct gradient images from groups of consecutive event-images. Intensity is obtained using Poisson integration [1, 2]. Optical flow [6, 7, 14, 31], together with the brightness

constancy equation can be used to estimate intensity. Bardow et al. [3] simultaneously optimise optical flow and intensity estimates within a fixed-length, sliding spatio-temporal window using the primal-dual algorithm [26]. Taking a spatio-temporal window of events imposes a latency cost at minimum equal to the length of the time window, and choosing a time-interval (or event batch size) that works robustly for all types of scenes is not trivial. Reinbacher et al. [29] integrate events over time while periodically regularising the estimate on a manifold defined by the timestamps of the latest events at each pixel; the surface of active events [6]. An alternative approach is to estimate camera pose and map in a SLAM-like framework [11,20,21,28,35]. Intensity can be recovered from the map, for example via Poisson integration of image gradients. These approaches work well for static scenes, but are not designed for dynamic scenes. Belbachir et al. [5] reconstruct intensity panoramas from a pair of 1D stereo event cameras on a single-axis rotational device by filtering events (e.g. temporal high-pass filter).

Combining different sensing modalities (e.g. a conventional camera) with event cameras [8, 10, 15, 19, 23, 30, 32] can overcome limitations of pure events, including lack of information about static or texture-less regions of the scene that do not trigger many events. Brandli et al. [9] combine image frames and event stream from the DAVIS camera to create inter-frame intensity estimates by dynamically estimating the contrast threshold (temporal contrast) of each event. Each new image frame resets the intensity estimate, preventing excessive growth of integration error, but also discarding important accumulated event information. Shedligeri et al. [30] use events to estimate ego-motion, then warp low frame-rate images to intermediate locations. Liu et al. [23] use affine motion models to reconstruct video of high-speed foreground/static background scenes.

We introduce the concept of a continuous-time image state that is asynchronously updated with every event. Our method is motion-model free and can be used with events only, or with image frames to complement events.

3 Approach

3.1 Mathematical Representation and Notation

Let $Y(\boldsymbol{p}, t)$ denote the intensity or irradiance of pixel $\boldsymbol{p}$ at time t of a camera. We will assume that the same irradiance is observed at the same pixel in both a classical and event camera, such as is the case with the DAVIS camera [8]. A classical image frame (for a global shutter camera) is an average of the received intensity over the exposure time

$$Y_j(\boldsymbol{p}) := \frac{1}{\epsilon} \int_{t_j - \epsilon}^{t_j} Y(\boldsymbol{p}, \tau) \mathrm{d}\tau, \quad j \in 1, 2, 3... \,, \tag{1}$$

where t_j is the time-stamp of the image capture and ϵ is the exposure time. In the sequel we will ignore the exposure time in the analysis and simply consider a classical image as representing image information available at time t_j. Although there will be image blur effects, especially for fast moving scenes in low light

conditions (see the experimental results in §5), a full consideration of these effects is beyond the scope of the present paper.

The approach taken in this paper is to analyse image reconstruction for event cameras in the continuous-time domain. To this end, we define a continuous-time intensity signal $Y^F(\boldsymbol{p}, t)$ as the zero-order hold (ZOH) reconstruction of the irradiance from the classical image frames:

$$Y^F(\boldsymbol{p}, t) := Y_j(\boldsymbol{p}) = Y(\boldsymbol{p}, t_j), \quad t_j \leq t < t_{j+1} \tag{2}$$

Since event cameras operate with log intensity we convert the image intensity into log-intensity:

$$L(\boldsymbol{p}, t) := \log(Y(\boldsymbol{p}, t)) \tag{3}$$

$$L_j(\boldsymbol{p}) := \log(Y_j(\boldsymbol{p})) \tag{4}$$

$$L^F(\boldsymbol{p}, t) := \log(Y^F(\boldsymbol{p}, t)). \tag{5}$$

Note that converting the zero-hold signal into the log domain is not the same as integrating the log intensity of the irradiance over the shutter time. We believe the difference will be insignificant in the scenarios considered and we do not consider this further in the present paper.

Dynamic vision sensors (DVS), or event cameras, are biologically-inspired vision sensors that respond to changes in scene illumination. Each pixel is independently wired to continuously compare the current log intensity level to the last reset-level. When the difference in log intensity exceeds a predetermined threshold (contrast threshold), an event is transmitted and the pixel resets, storing the new illumination level. Each event contains the pixel coordinates, timestamp, and polarity ($\sigma = \pm 1$ for increasing or decreasing intensity). An event can be modelled in the continuous-time[3] signal class as a Dirac-delta function $\delta(t)$. We define an event stream $e_i(\boldsymbol{p}, t)$ at pixel $\boldsymbol{p}$ by

$$e_i(\boldsymbol{p}, t) := \sigma_i^{\boldsymbol{p}} c\, \delta(t - t_i^{\boldsymbol{p}}),\ i \in 1, 2, 3... , \tag{6}$$

where $\sigma_i^{\boldsymbol{p}}$ is the polarity and $t_i^{\boldsymbol{p}}$ is the time-stamp of the i^{th} event at pixel $\boldsymbol{p}$. The magnitude c is the *contrast threshold* (brightness change encoded by one event). Define an *event field* $E(\boldsymbol{p}, t)$ by

$$E(\boldsymbol{p}, t) := \sum_{i=1}^{\infty} e_i(\boldsymbol{p}, t) = \sum_{i=1}^{\infty} \sigma_i^{\boldsymbol{p}} c\, \delta(t - t_i^{\boldsymbol{p}}). \tag{7}$$

The event field is a function of all pixels $\boldsymbol{p}$ and ranges over all time, capturing the full output of the event camera.

A quantised log intensity signal $L^E(\boldsymbol{p}, t)$ can be reconstructed by integrating the event field

$$L^E(\boldsymbol{p}, t) := \int_0^t E(\boldsymbol{p}, \tau) d\tau = \int_0^t \sum_{i=1}^{\infty} \sigma_i^{\boldsymbol{p}} c\, \delta(\tau - t_i^{\boldsymbol{p}}) d\tau. \tag{8}$$

[3] Note that events are continuous-time signals even though they are not continuous functions of time; the time variable t on which they depend varies continuously.

The result is a series of log intensity steps (corresponding to events) at each pixel. In the absence of noise, the relationship between the log-intensity $L(\boldsymbol{p}, t)$ and the quantised signal $L^E(\boldsymbol{p}, t)$ is

$$L(\boldsymbol{p}, t) = L^E(\boldsymbol{p}, t) + L(\boldsymbol{p}, 0) + \mu(\boldsymbol{p}, t; c), \tag{9}$$

where $L(\boldsymbol{p}, 0)$ is the initial condition and $\mu(\boldsymbol{p}, t; c)$ is the quantisation error. Unlike $L^F(\boldsymbol{p}, t)$, the quantisation error associated with $L^E(\boldsymbol{p}, t)$ is bounded by the contrast threshold; $|\mu(\boldsymbol{p}, t; c)| < c$.

Remark 1. Events can be interpreted as the temporal derivative of $L^E(\boldsymbol{p}, t)$

$$E(\boldsymbol{p}, t) = \frac{\partial}{\partial t} L^E(\boldsymbol{p}, t). \tag{10}$$

3.2 Complementary Filter

We will use a complementary filter structure [13, 16, 24] to fuse the event field $E(\boldsymbol{p}, t)$ with ZOH log-intensity frames $L^F(\boldsymbol{p}, t)$. Complementary filtering is ideal for fusing signals that have complementary frequency noise characteristics; for example, where one signal is dominated by high-frequency noise and the other by low-frequency disturbance. Events are a temporal derivative measurement (10) and do not contain reference intensity $L(\boldsymbol{p}, 0)$ information. Integrating events to obtain $L^E(\boldsymbol{p}, t)$ amplifies low-frequency disturbance (drift), resulting in poor low-frequency information. However, due to their high-temporal-resolution, events provide reliable *high-frequency* information. Classical image frames $L^F(\boldsymbol{p}, t)$ are derived from discrete, temporally-sparse measurements and have poor high-frequency fidelity. However, frames typically provide reliable *low-frequency* reference intensity information. The proposed complementary filter architecture combines a *high-pass* version of $L^E(\boldsymbol{p}, t)$ with a *low-pass* version of $L^F(\boldsymbol{p}, t)$ to reconstruct an (approximate) all-pass version of $L(\boldsymbol{p}, t)$.

The proposed filter is written as a continuous-time ordinary differential equation (ODE)

$$\boxed{\frac{\partial}{\partial t} \hat{L}(\boldsymbol{p}, t) = E(\boldsymbol{p}, t) - \alpha\big(\hat{L}(\boldsymbol{p}, t) - L^F(\boldsymbol{p}, t)\big),} \tag{11}$$

where $\hat{L}(\boldsymbol{p}, t)$ is the continuous-time log-intensity state estimate and α is the complementary filter gain, or crossover frequency [24] (Fig. 1).

The fact that the input signals in (11) are discontinuous poses some complexities in solving the filter equations, but does not invalidate the formulation. The filter can be understood as integration of the event field with an innovation term $-\alpha\big(\hat{L}(\boldsymbol{p}, t) - L^F(\boldsymbol{p}, t)\big)$, that acts to reduce the error between $\hat{L}(\boldsymbol{p}, t)$ and $L^F(\boldsymbol{p}, t)$.

The key property of the proposed filter (11) is that although it is posed as a continuous-time ODE, one can express the solution as a set of asynchronous-update equations. Each pixel acts independently, and in the sequel we will consider the action of the complementary filter on a single pixel $\boldsymbol{p}$. Recall the sequence $\{t_i^{\boldsymbol{p}}\}$ corresponding to the time-stamps of all events at $\boldsymbol{p}$. In addition,

there is the sequence of classical image frame time-stamps $\{t_j\}$ that apply to all pixels equally. Consider a combined sequence of monotonically increasing *unique* time-stamps $\hat{t}_k^{\boldsymbol{p}}$ corresponding to event $\{t_i^{\boldsymbol{p}}\}$ or frame $\{t_j\}$ time-stamps.

Within a time-interval $t \in [\hat{t}_k^{\boldsymbol{p}}, \hat{t}_{k+1}^{\boldsymbol{p}})$ there are (by definition) no new events or frames, and the ODE (11) is a constant coefficient linear ordinary differential equation

$$\frac{\partial}{\partial t}\hat{L}(\boldsymbol{p},t) = -\alpha\big(\hat{L}(\boldsymbol{p},t) - L^F(\boldsymbol{p},t)\big), \quad t \in [\hat{t}_k^{\boldsymbol{p}}, \hat{t}_{k+1}^{\boldsymbol{p}}). \tag{12}$$

The solution to this ODE is given by

$$\hat{L}(\boldsymbol{p},t) = e^{-\alpha(t-\hat{t}_k^{\boldsymbol{p}})}\hat{L}(\boldsymbol{p},\hat{t}_k^{\boldsymbol{p}}) + (1 - e^{-\alpha(t-\hat{t}_k^{\boldsymbol{p}})})L^F(\boldsymbol{p},t), \qquad t \in [\hat{t}_k^{\boldsymbol{p}}, \hat{t}_{k+1}^{\boldsymbol{p}}). \tag{13}$$

It remains to paste together the piece-wise smooth solutions on the half-open intervals $[\hat{t}_k^{\boldsymbol{p}}, \hat{t}_{k+1}^{\boldsymbol{p}})$ by considering the boundary conditions. Let

$$(\hat{t}_{k+1}^{\boldsymbol{p}})^- := \lim_{t \to (\hat{t}_{k+1}^{\boldsymbol{p}})} t, \quad \text{for } t < \hat{t}_{k+1}^{\boldsymbol{p}} \tag{14}$$

$$(\hat{t}_{k+1}^{\boldsymbol{p}})^+ := \lim_{t \to (\hat{t}_{k+1}^{\boldsymbol{p}})} t, \quad \text{for } t > \hat{t}_{k+1}^{\boldsymbol{p}}, \tag{15}$$

denote the limits from below and above. There are two cases to consider:

New frame: When the index $\hat{t}_{k+1}^{\boldsymbol{p}}$ corresponds to a new image frame then the right hand side (RHS) of (11) has bounded variation. It follows that the solution is continuous at $\hat{t}_{k+1}^{\boldsymbol{p}}$ and

$$\hat{L}(\boldsymbol{p},\hat{t}_{k+1}^{\boldsymbol{p}}) = \hat{L}(\boldsymbol{p},(\hat{t}_{k+1}^{\boldsymbol{p}})^-). \tag{16}$$

Event: When the index $\hat{t}_{k+1}^{\boldsymbol{p}}$ corresponds to an event then the solution of (11) is *not* continuous at $\hat{t}_{k+1}^{\boldsymbol{p}}$ and the Dirac delta function of the event must be integrated. Integrating the RHS and LHS of (11) over an event

$$\int_{(\hat{t}_{k+1}^{\boldsymbol{p}})^-}^{(\hat{t}_{k+1}^{\boldsymbol{p}})^+} \frac{d}{d\tau}\hat{L}(\boldsymbol{p},\tau)d\tau = \int_{(\hat{t}_{k+1}^{\boldsymbol{p}})^-}^{(\hat{t}_{k+1}^{\boldsymbol{p}})^+} E(\boldsymbol{p},\tau) - \alpha\big(\hat{L}(\boldsymbol{p},\tau) - L^F(\boldsymbol{p},\tau)\big)d\tau \tag{17}$$

$$\hat{L}(\boldsymbol{p},(\hat{t}_{k+1}^{\boldsymbol{p}})^+) - \hat{L}(\boldsymbol{p},(\hat{t}_{k+1}^{\boldsymbol{p}})^-) = \sigma_{k+1}^{\boldsymbol{p}}c, \tag{18}$$

yields a unit step scaled by the contrast threshold and sign of the event. Note the integral of the second term $\int_{(\hat{t}_{k+1}^{\boldsymbol{p}})^-}^{(\hat{t}_{k+1}^{\boldsymbol{p}})^+} \alpha\big(\hat{L}(\boldsymbol{p},\tau) - L^F(\boldsymbol{p},\tau)\big)d\tau$ is zero since the integrand is bounded. We use the solution

$$\hat{L}(\boldsymbol{p},\hat{t}_{k+1}^{\boldsymbol{p}}) = \hat{L}(\boldsymbol{p},(\hat{t}_{k+1}^{\boldsymbol{p}})^-) + \sigma_{k+1}^{\boldsymbol{p}}c, \tag{19}$$

as initial condition for the next time-interval. Eqns. (13), (16) and (19) characterise the full solution to the filter equation (11).

Remark 2. The filter can be run on *events only* without image frames by setting $L^F(\boldsymbol{p}, t) = 0$ in (11), resulting in a high-pass filter with corner frequency α

$$\boxed{\frac{\partial}{\partial t}\hat{L}(\boldsymbol{p}, t) = E(\boldsymbol{p}, t) - \alpha\hat{L}(\boldsymbol{p}, t).} \tag{20}$$

This method can efficiently generate a good quality image state estimate from pure events. Furthermore, it is possible to use alternative pure event-based methods to reconstruct a temporally-sparse image sequence from events and fuse this with raw events using the proposed complementary filter. Thus, the proposed filter can be considered a method to augment any event-based image reconstruction method to obtain a high temporal-resolution image state.

4 Method

4.1 Adaptive Gain Tuning

The complementary filter gain α is a parameter that controls the relative information contributed by image frames or events. Reducing the magnitude of α decreases the dependence on image frame data while increasing the dependence on events ($\alpha = 0 \rightarrow \hat{L}(\boldsymbol{p}, t) = L^E(\boldsymbol{p}, t)$). A key observation is that the gain can be time-varying at pixel-level ($\alpha = \alpha(\boldsymbol{p}, t)$). One can therefore use $\alpha(\boldsymbol{p}, t)$ to dynamically adjust the relative dependence on image frames or events, which can be useful when image frames are compromised, e.g. under- or overexposed.

We propose to reduce the influence of under/overexposed image frame pixels by decreasing $\alpha(\boldsymbol{p}, t)$ at those pixel locations. We use the heuristic that pixels reporting an intensity close to the minimum $L_{\min}$ or maximum $L_{\max}$ output of the camera may be compromised, and we decrease $\alpha(\boldsymbol{p}, t)$ based on the reported log intensity. We choose two bounds L_1, L_2 close to $L_{\min}$ and $L_{\max}$, then we set $\alpha(\boldsymbol{p}, t)$ to a constant (α_1) for all pixels within the range $[L_1, L_2]$, and linearly decrease $\alpha(\boldsymbol{p}, t)$ for pixels outside of this range:

$$\alpha(\boldsymbol{p}, t) = \begin{cases} \lambda\alpha_1 + (1-\lambda)\alpha_1 \frac{(L^F(\boldsymbol{p},t) - L_{\min})}{(L_1 - L_{\min})} & L_{\min} \leq L^F(\boldsymbol{p}, t) < L_1 \\ \alpha_1 & L_1 \leq L^F(\boldsymbol{p}, t) \leq L_2 \\ \lambda\alpha_1 + (1-\lambda)\alpha_1 \frac{(L^F(\boldsymbol{p},t) - L_{\max})}{(L_2 - L_{\max})} & L_2 < L^F(\boldsymbol{p}, t) \leq L_{\max} \end{cases} \tag{21}$$

where λ is a parameter determining the strength of our adaptive scheme (we set $\lambda = 0.1$). For α_1, typical suitable values are $\alpha_1 \in [0.1, 10]$ rad/s. For our experiments we choose $\alpha_1 = 2\pi$ rad/s.

4.2 Asynchronous Update Scheme

Given temporally sparse image frames and events, and using the continuous-time solution to the complementary filter ODE (11) outlined in §3 one may compute the intensity state estimate $\hat{L}(\boldsymbol{p}, t)$ at any time. In practice it is sufficient to

Algorithm 1 Per-pixel, Asynchronous Complementary Filter

At each pixel:
Initialise $\hat{L}_\diamond$, $\hat{t}_\diamond$, $L^F_\diamond$ to zero
Initialise $\alpha_\diamond$ to α_1
for each new event **or** image frame **do**
    $\Delta t \leftarrow t - \hat{t}_\diamond$
    $\hat{L}_\diamond \leftarrow \exp(-\alpha_\diamond \cdot \Delta t) \cdot \hat{L}_\diamond + (1 - \exp(-\alpha_\diamond \cdot \Delta t)) \cdot L^F_\diamond$ based on (13)
    if event **then**
        $\hat{L}_\diamond \leftarrow \hat{L}_\diamond + \sigma c$ based on (19)
    else if image frame **then**
        Replace $L^F_\diamond$ with new frame
        Update $\alpha_\diamond$ based on (21)
    $\hat{t}_\diamond \leftarrow t$

compute the image state $\hat{L}(\boldsymbol{p}, t)$ at the asynchronous time instances $\hat{t}^{\boldsymbol{p}}_k$ (event or frame timestamps). We propose an asynchronous update scheme whereby new events cause state updates (19) *only* at the event pixel-location. New frames cause a global update (16) (note this is not a reset as in [9]) [4]. Algorithm 1 describes a per-pixel complementary filter implementation. At a given pixel $\boldsymbol{p}$, let $\hat{L}_\diamond$ denote the latest estimate of $\hat{L}(\boldsymbol{p}, t)$ stored in computer memory, and $\hat{t}_\diamond$ denote the time-stamp of the latest update at $\boldsymbol{p}$. Let $L^F_\diamond$ and $\alpha_\diamond$ denote the latest image frame and gain values at $\boldsymbol{p}$. To run the filter in events only mode (high-pass filter (20)), simply let $L^F_\diamond = 0$.

5 Results

We compare the reconstruction performance of our complementary filter, both with frames ($\mathrm{CF_f}$) and without frames in *events only* mode ($\mathrm{CF_e}$), against three state-of-the-art methods: manifold regularization (MR) [29], direct integration (DI) [9] and simultaneous optical flow and intensity estimation (SOFIE) [3]. We introduce new datasets: two new ground truth sequences (Truck and Motorbike); and four new sequences taken with the DAVIS240C [8] camera (Night drive, Sun, Bicycle, Night run). We evaluate our method, MR and DI against our ground truth dataset using quantitative image similarity metrics. Unfortunately, as code is not available for SOFIE we are unable to evaluate its performance on our new datasets. Hence, we compare it with our method on the jumping sequence made available by the authors.

Ground truth is obtained using a high-speed, global-shutter, frame-based camera (mvBlueFOX USB 2) running at 168Hz. We acquire image sequences of dynamic scenes (Truck and Motorbike), and convert them into events following the methodology of Mueggler et al. [25]. To simulate event camera noise, a number of random noise events are generated (5% of total events), and distributed

[4] The filter can also be updated (using (13)) at any user-chosen time instance (or rate). In our experiments we update the entire image state whenever we export the image for visualisation.

randomly throughout the event stream. To simulate low-dynamic-range, low-temporal-resolution input-frames, the upper and lower 25% of the maximum intensity range is truncated, and image frames are subsampled at 20Hz. In addition, a delay of 50ms is applied to the frame time-stamps to simulate the latency associated with capturing images using a frame-based camera.

The complementary filter gain $\alpha(\boldsymbol{p}, t)$ is set according to (21) and updated with every new image frame (Algorithm 1). We set $\alpha_1 = 2\pi$ rad/s for all sequences. The bounds $[L_1,\ L_2]$ in (21) are set to $[L_{\min} + \kappa,\ L_{\max} - \kappa]$, where $\kappa = 0.05(L_{\max} - L_{\min})$. The contrast threshold (c) is not easy to determine and in practice varies across pixels, and with illumination, event-rate and other factors [9]. Here we assume two constant contrast thresholds (ON and OFF) that are calibrated for each sequence using APS frames. We note that error arising from the variability of contrast thresholds appears as noise in the final estimate, and believe that more sophisticated contrast threshold models may benefit future works. For MR [29], the number of events per output image (events/image) is a parameter that impacts the quality of the reconstructed image. For each sequence we choose events/image to give qualitatively best performance. We set events/image to 1500 unless otherwise stated. All other parameters are set to defaults provided by [29].

Night drive (Fig. 2) investigates performance in high-speed, low light conditions where the conventional camera image frame (Raw frame) is blurry and underexposed, and dark details are lost. Data is recorded through the front wind shield of a car, driving down an unlit highway at dead of night. Our method (CF) is able to recover motion-blurred objects (e.g. roadside poles), trees that are lost in Raw frame, and road lines that are lost in MR. MR relies on spatial smoothing to reduce noise, hence faint features such as distant trees (Fig. 2; zoom) may be lost. DI loses features that require more time for events to accumulate (e.g. trees on the right), because the estimate is reset upon every new image frame. The APS (active pixel sensor in DAVIS) frame-rate was set to 7Hz.

Sun investigates extreme dynamic range scenes where conventional cameras become overexposed. CF recovers features such as leaves and twigs, even when the camera is pointed directly at the sun. Raw frame is largely over-saturated, and the black dot (Fig. 2; Sun) is a camera artifact caused by extreme brightness, and marks the position of the sun. MR produces a clean looking image, though some features (small leaves/twigs) are smoothed out (Fig. 2; zoom). DI is washed out in regions where the frame is overexposed, due to the latest frame reset. Because the sun generates so many events, MR requires more events/image to recover fine features (with less events the image looks oversmoothed), so we increase events/image to 2500. The APS frame-rate was set to 26Hz.

Bicycle explores the scenario of static background, moving foreground. Raw frame is underexposed in shady areas because of large intra-scene dynamic range. When the event camera is stationary, almost no events are generated by the static background and it cannot be recovered by pure event-based reconstruction methods such as MR and $\mathrm{CF_e}$. In contrast, $\mathrm{CF_f}$ recovers both stationary and

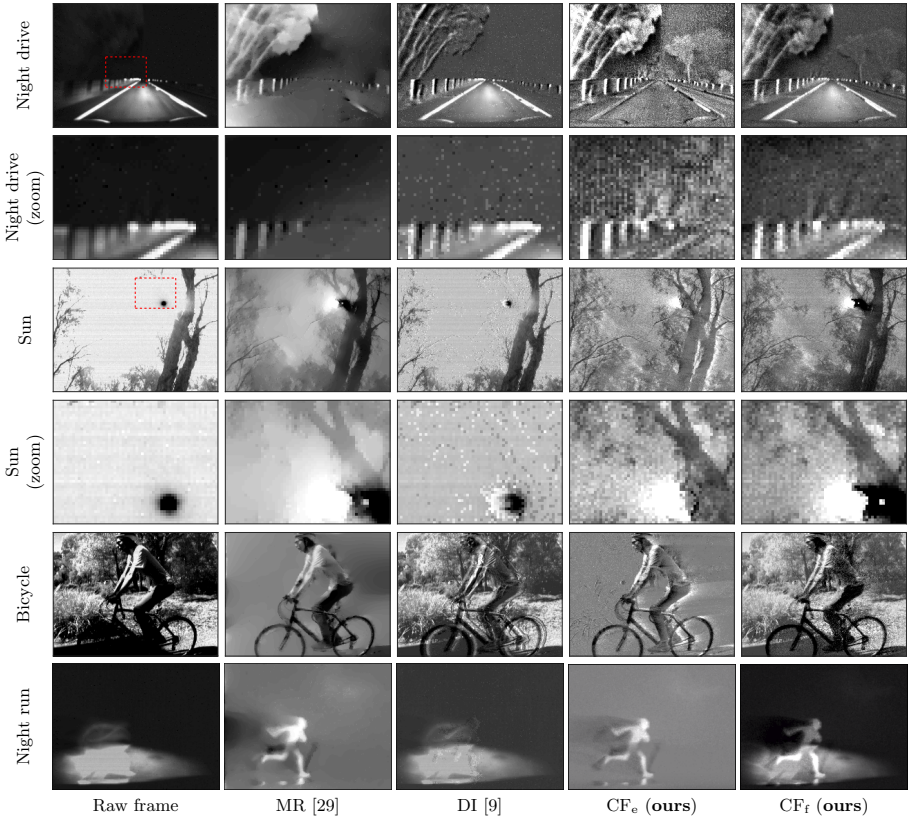

Fig. 2. Night drive: Raw frame is motion-blurred and contains little information in dark areas, but captures road markings. MR is unable to recover some road markings. CF$_f$ recovers sharp road markings, trees and roadside poles. **Sun:** Raw frame is overexposed when pointing directly at the sun (black dot is a camera artifact caused by the sun). In MR, some features are smoothed out (see zoom). DI is washed out due to the latest frame reset. CF captures detailed leaves and twigs. **Bicycle:** Static background cannot be recovered from events alone in MR and CF$_e$. CF$_f$ recovers both background and foreground. **Night run:** Pedestrian is heavily motion-blurred and delayed in Raw frame. DI may be compromised by frame-resets. MR and CF recover sharp detail despite high-speed, low-light conditions.

non-stationary features, as well as high-dynamic-range detail. The APS framerate was set to 26Hz.

Night run illustrates the benefit in challenging low-light pedestrian scenarios. Here a pedestrian runs across the headlights of a (stationary) car at dead of night. Raw frame is not only heavily motion-blurred, but also significantly delayed, since a large exposure duration is required for image acquisition in low-light conditions. DI is unreliable as an unfortunately timed new image frame could reset the image (Fig. 2). MR and CF manage to recover the pedes-

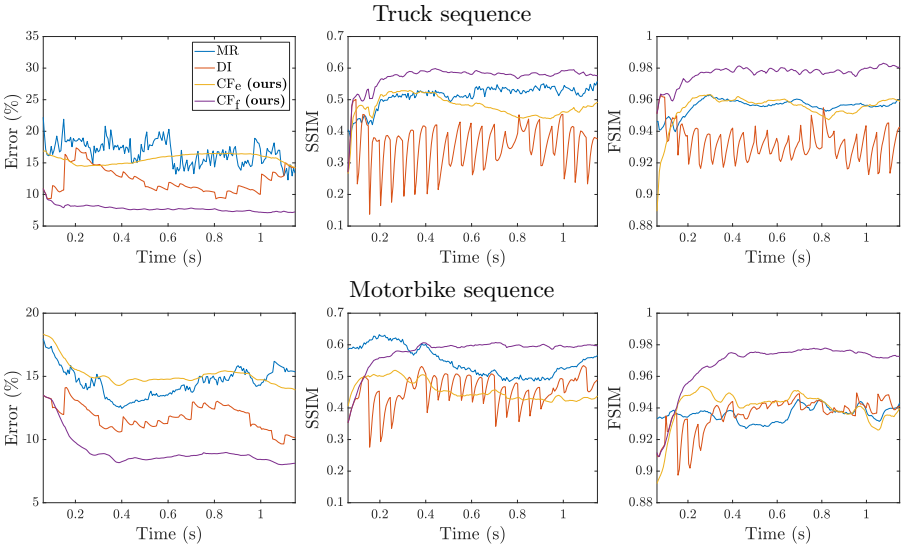

Fig. 3. Full-reference quantitative evaluation of each reconstruction method on our ground truth datasets using photometric error (%), SSIM [33] and FSIM [34].

Table 1. Overall performance of each reconstruction method on our ground truth dataset (Truck and Motorbike). Values are reported as mean ± standard deviation. Our method (CF_f) outperforms state-of-the-art on all metrics.

	Truck sequence			Motorbike sequence		
Method	Photometric Error (%)	SSIM	FSIM	Photometric Error (%)	SSIM	FSIM
Direct integration [9]	12.25 ± 1.94	0.36 ± 0.07	0.93 ± 0.01	11.78 ± 0.99	0.45 ± 0.05	0.94 ± 0.01
Manifold regularisation [29]	16.81 ± 1.58	0.51 ± 0.03	0.96 ± 0.00	14.53 ± 1.13	0.55 ± 0.05	0.94 ± 0.00
CF_e (ours)	15.67 ± 0.73	0.48 ± 0.03	0.95 ± 0.01	15.14 ± 0.88	0.45 ± 0.03	0.94 ± 0.01
CF_f (**ours**)	$\mathbf{7.76 \pm 0.49}$	$\mathbf{0.57 \pm 0.03}$	$\mathbf{0.98 \pm 0.01}$	$\mathbf{9.05 \pm 1.19}$	$\mathbf{0.58 \pm 0.04}$	$\mathbf{0.97 \pm 0.02}$

trian, and CF_f also recovers the background without compromising clarity of the pedestrian. The APS frame-rate was set to 4.5Hz.

Ground truth evaluation. We evaluate our method with (CF_f) and without (CF_e) 20Hz input-frames, and compare against DI [9] and MR [29]. To assess similarity between ground truth and reconstructed images, each ground truth frame is matched with the corresponding reconstructed image with the closest time-stamp. Average absolute photometric error (%), structural similarity (SSIM) [33], and feature similarity (FSIM) [34] are used to evaluate performance (Fig. 3 and Table 1). We initialise DI and CF_f using the first input-frame, and MR and CF_e to zero.

Fig. 3 plots the performance of each reconstruction method over time. Our method shows an initial improvement as useful information starts to accumulate, then maintains good performance over time as new events and frames are incorporated into the estimate. The oscillations apparent in DI arise from image frame

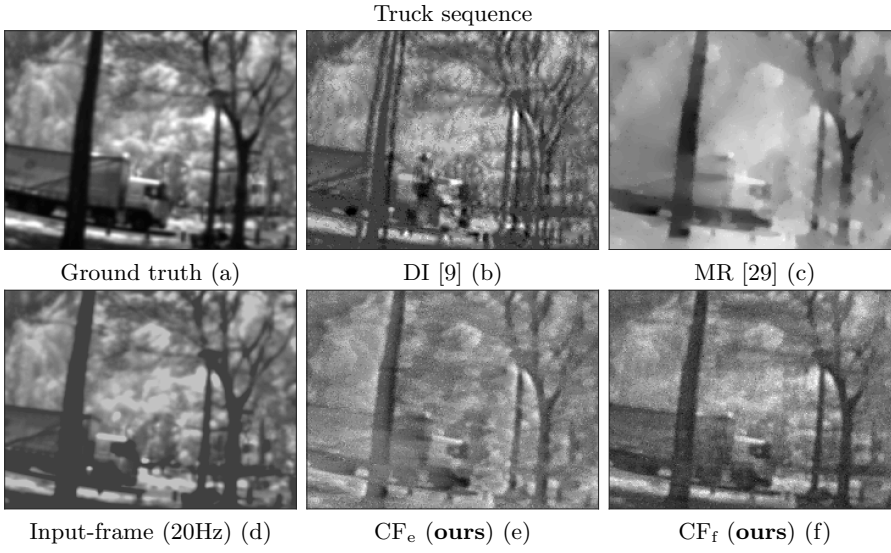

Fig. 4. Reconstructed image for each method (DI, MR, CF) on ground truth dataset (Truck) with raw input-frame (d) and ground truth (a) for comparison. DI (b) displays edge artifacts where new events are added directly to the latest input-frame. MR (c) produces smoothed images compared to CF (e), (f).

resets. Table 1 summarises average performance for each sequence. Our method CF_f achieves the lowest photometric error, and highest SSIM and FSIM scores for all sequences. Fig. 4 shows the reconstructed image halfway between two input-frames of Truck ground truth sequence. Pure event-based methods (MR and CF_e) do not recover absolute intensity in some regions (truck body) due to sparsity of events. DI displays artifacts around edges, where many events are generated, because events are directly added to the latest input-frame. In CF_f, event and frame information is continuously combined, reducing edge artifacts (Fig. 4) and producing a more consistent estimate over time (Fig. 3).

SOFIE. The code for SOFIE [3] was not available at the time of writing, however the authors kindly share their pre-recorded dataset (using DVS128 [22]) and results. We use their dataset to compare our method to SOFIE (Fig. 5), and since no camera frames are available, we first demonstrate our method by setting input-frames to zero (CF_e), then show that reconstructed image frames output from an alternative reconstruction algorithm such as SOFIE can be used as input-frames to the complementary filter (CF_f (events + SOFIE)) to generate intensity estimates.

6 Conclusion

We have presented a continuous-time formulation for intensity estimation using an event-driven complementary filter. We compare complementary filtering

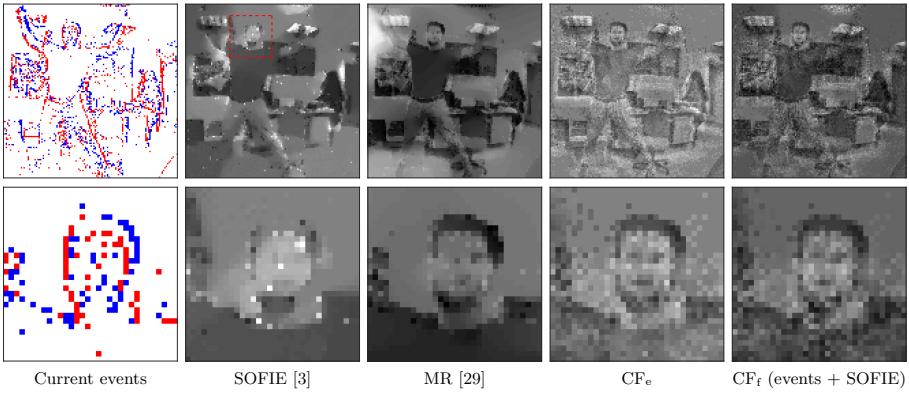

Fig. 5. SOFIE [3] and MR [29] reconstruct images from a pure event stream. CF can reconstruct images from a pure event stream by either setting input-frames to zero CF$_e$, or by taking reconstructed images from other methods (e.g. SOFIE) as input-frames CF$_f$ (events + SOFIE).

with existing reconstruction methods on sequences recorded on a DAVIS camera, and show that the complementary filter outperforms current state-of-the-art on a newly presented ground truth dataset. Finally, we show that the complementary filter can estimate intensity based on a pure event stream, by either setting the input-frame signal to zero, or by fusing events with the output of a computationally intensive reconstruction method.

References

1. Agrawal, A., Chellappa, R., Raskar, R.: An algebraic approach to surface reconstruction from gradient fields. In: Int. Conf. Comput. Vis. (ICCV). pp. 174–181 (2005). https://doi.org/10.1109/iccv.2005.31
2. Agrawal, A., Raskar, R., Chellappa, R.: What is the range of surface reconstructions from a gradient field? In: Eur. Conf. Comput. Vis. (ECCV). pp. 578–591 (2006)
3. Bardow, P., Davison, A.J., Leutenegger, S.: Simultaneous optical flow and intensity estimation from an event camera. In: IEEE Conf. Comput. Vis. Pattern Recog. (CVPR). pp. 884–892 (2016). https://doi.org/10.1109/CVPR.2016.102
4. Barua, S., Miyatani, Y., Veeraraghavan, A.: Direct face detection and video reconstruction from event cameras. In: IEEE Winter Conf. Appl. Comput. Vis. (WACV). pp. 1–9 (2016). https://doi.org/10.1109/WACV.2016.7477561
5. Belbachir, A.N., Schraml, S., Mayerhofer, M., Hofstaetter, M.: A novel HDR depth camera for real-time 3D 360-degree panoramic vision. In: IEEE Conf. Comput. Vis. Pattern Recog. Workshops (CVPRW) (Jun 2014)
6. Benosman, R., Clercq, C., Lagorce, X., Ieng, S.H., Bartolozzi, C.: Event-based visual flow. IEEE Trans. Neural Netw. Learn. Syst. **25**(2), 407–417 (2014). https://doi.org/10.1109/TNNLS.2013.2273537

7. Benosman, R., Ieng, S.H., Clercq, C., Bartolozzi, C., Srinivasan, M.: Asynchronous frameless event-based optical flow. Neural Netw. **27**, 32–37 (2012). https://doi.org/10.1016/j.neunet.2011.11.001
8. Brandli, C., Berner, R., Yang, M., Liu, S.C., Delbruck, T.: A 240x180 130dB 3us latency global shutter spatiotemporal vision sensor. IEEE J. Solid-State Circuits **49**(10), 2333–2341 (2014). https://doi.org/10.1109/JSSC.2014.2342715
9. Brandli, C., Muller, L., Delbruck, T.: Real-time, high-speed video decompression using a frame- and event-based DAVIS sensor. In: IEEE Int. Symp. Circuits Syst. (ISCAS). pp. 686–689 (Jun 2014). https://doi.org/10.1109/ISCAS.2014.6865228
10. Censi, A., Scaramuzza, D.: Low-latency event-based visual odometry. In: IEEE Int. Conf. Robot. Autom. (ICRA). pp. 703–710 (2014). https://doi.org/10.1109/ICRA.2014.6906931
11. Cook, M., Gugelmann, L., Jug, F., Krautz, C., Steger, A.: Interacting maps for fast visual interpretation. In: Int. Joint Conf. Neural Netw. (IJCNN). pp. 770–776 (2011). https://doi.org/10.1109/IJCNN.2011.6033299
12. Delbruck, T., Linares-Barranco, B., Culurciello, E., Posch, C.: Activity-driven, event-based vision sensors. In: IEEE Int. Symp. Circuits Syst. (ISCAS). pp. 2426–2429 (May 2010). https://doi.org/10.1109/ISCAS.2010.5537149
13. Franklin, G.F., Powell, J.D., Workman, M.L.: Digital control of dynamic systems, vol. 3. Addison-wesley Menlo Park, CA (1998)
14. Gallego, G., Rebecq, H., Scaramuzza, D.: A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation. In: IEEE Conf. Comput. Vis. Pattern Recog. (CVPR). pp. 3867–3876 (2018). https://doi.org/10.1109/CVPR.2018.00407
15. Gehrig, D., Rebecq, H., Gallego, G., Scaramuzza, D.: Asynchronous, photometric feature tracking using events and frames. In: Eur. Conf. Comput. Vis. (ECCV) (2018). https://doi.org/10.1007/978-3-030-01258-8_46
16. Higgins, W.T.: A comparison of complementary and kalman filtering. IEEE Transactions on Aerospace and Electronic Systems (3), 321–325 (1975)
17. Huang, J., Guo, M., Chen, S.: A dynamic vision sensor with direct logarithmic output and full-frame picture-on-demand. In: IEEE Int. Symp. Circuits Syst. (ISCAS). pp. 1–4 (May 2017). https://doi.org/10.1109/iscas.2017.8050546
18. Huang, J., Guo, M., Wang, S., Chen, S.: A motion sensor with on-chip pixel rendering module for optical flow gradient extraction. In: IEEE Int. Symp. Circuits Syst. (ISCAS). pp. 1–5 (May 2018). https://doi.org/10.1109/iscas.2018.8351312
19. Huang, J., Wang, S., Guo, M., Chen, S.: Event-guided structured output tracking of fast-moving objects using a CeleX sensor. IEEE Trans. Circuits Syst. Video Technol. **28**(9), 2413–2417 (Sep 2018). https://doi.org/10.1109/tcsvt.2018.2841516
20. Kim, H., Handa, A., Benosman, R., Ieng, S.H., Davison, A.J.: Simultaneous mosaicing and tracking with an event camera. In: British Machine Vis. Conf. (BMVC) (2014). https://doi.org/10.5244/C.28.26
21. Kim, H., Leutenegger, S., Davison, A.J.: Real-time 3D reconstruction and 6-DoF tracking with an event camera. In: Eur. Conf. Comput. Vis. (ECCV). pp. 349–364 (2016). https://doi.org/10.1007/978-3-319-46466-4_21
22. Lichtsteiner, P., Posch, C., Delbruck, T.: A 128×128 120 dB 15 μs latency asynchronous temporal contrast vision sensor. IEEE J. Solid-State Circuits **43**(2), 566–576 (2008). https://doi.org/10.1109/JSSC.2007.914337
23. Liu, H.C., Zhang, F.L., Marshall, D., Shi, L., Hu, S.M.: High-speed video generation with an event camera. The Visual Computer **33**(6-8), 749–759 (May 2017). https://doi.org/10.1007/s00371-017-1372-y

24. Mahony, R., Hamel, T., Pflimlin, J.M.: Nonlinear complementary filters on the special orthogonal group. IEEE Transactions on automatic control **53**(5), 1203–1218 (2008)
25. Mueggler, E., Rebecq, H., Gallego, G., Delbruck, T., Scaramuzza, D.: The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and SLAM. Int. J. Robot. Research **36**, 142–149 (2017). https://doi.org/10.1177/0278364917691115
26. Pock, T., Chambolle, A.: Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In: Int. Conf. Comput. Vis. (ICCV). pp. 1762–1769 (Nov 2011). https://doi.org/10.1109/iccv.2011.6126441
27. Posch, C., Matolin, D., Wohlgenannt, R.: A QVGA 143 dB dynamic range frame-free PWM image sensor with lossless pixel-level video compression and time-domain CDS. IEEE J. Solid-State Circuits **46**(1), 259–275 (Jan 2011). https://doi.org/10.1109/JSSC.2010.2085952
28. Rebecq, H., Horstschäfer, T., Gallego, G., Scaramuzza, D.: EVO: A geometric approach to event-based 6-DOF parallel tracking and mapping in real-time. IEEE Robot. Autom. Lett. **2**, 593–600 (2017). https://doi.org/10.1109/LRA.2016.2645143
29. Reinbacher, C., Graber, G., Pock, T.: Real-time intensity-image reconstruction for event cameras using manifold regularisation. In: British Machine Vis. Conf. (BMVC) (2016). https://doi.org/10.5244/C.30.9
30. Shedligeri, P.A., Shah, K., Kumar, D., Mitra, K.: Photorealistic image reconstruction from hybrid intensity and event based sensor. arXiv preprint arXiv:1805.06140 (2018)
31. Stoffregen, T., Kleeman, L.: Simultaneous optical flow and segmentation (SOFAS) using dynamic vision sensor. In: Australasian Conf. Robot. Autom. (ACRA) (2017)
32. Vidal, A.R., Rebecq, H., Horstschaefer, T., Scaramuzza, D.: Ultimate SLAM? combining events, images, and IMU for robust visual SLAM in HDR and high-speed scenarios. IEEE Robot. Autom. Lett. **3**(2), 994–1001 (Apr 2018). https://doi.org/10.1109/lra.2018.2793357
33. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (Apr 2004). https://doi.org/10.1109/tip.2003.819861
34. Zhang, L., Zhang, L., Mou, X., Zhang, D.: FSIM: A feature similarity index for image quality assessment. IEEE Trans. Image Process. **20**(8), 2378–2386 (Aug 2011). https://doi.org/10.1109/tip.2011.2109730
35. Zhou, Y., Gallego, G., Rebecq, H., Kneip, L., Li, H., Scaramuzza, D.: Semi-dense 3D reconstruction with a stereo event camera. In: Eur. Conf. Comput. Vis. (ECCV) (2018). https://doi.org/10.1007/978-3-030-01246-5_15

Continuous-time Intensity Estimation Using Event Cameras **Supplementary Material**

Section 1 provides a more in depth analysis of the results, and compares the time-evolution of our proposed Complementary filter (CF) to state-of-the-art. Section 2 (Appendix 1) provides a short review of complementary filtering and the underlying principles of frequency based data fusion.

1 Experimental Results

Key:

- Raw frame	Direct output of frame-based, global-shutter camera (DAVIS240C [1]).
- Current events	All events within a 5ms time window (red: positive (ON), blue: negative (OFF)).
- MR	Manifold regularisation [6] (uses events only).
- DI	Direct integration [2] (uses events and Raw frames).
- CF (events only)	Our Complementary filter with input-frames set to zero (uses events only).
- CF (events + frames)	Our Complementary filter using events and Raw frames.
- Ground truth	Ground truth image captured with the mvBlueFox global-shutter camera running at 168Hz.
- Input-frame (20Hz)	Ground truth sequence subsampled at 20Hz, with 50ms delay applied, and upper and lower 25% of intensity range truncated.

1.1 Night drive

Figure 1 shows that CF detects distant trees earlier than MR and DI, and the same trees can be seen more clearly one second later, as they draw nearer. The frame reset in DI periodically throws away accumulated information from events, making it more difficult to detect distant trees. Pure event-based reconstruction

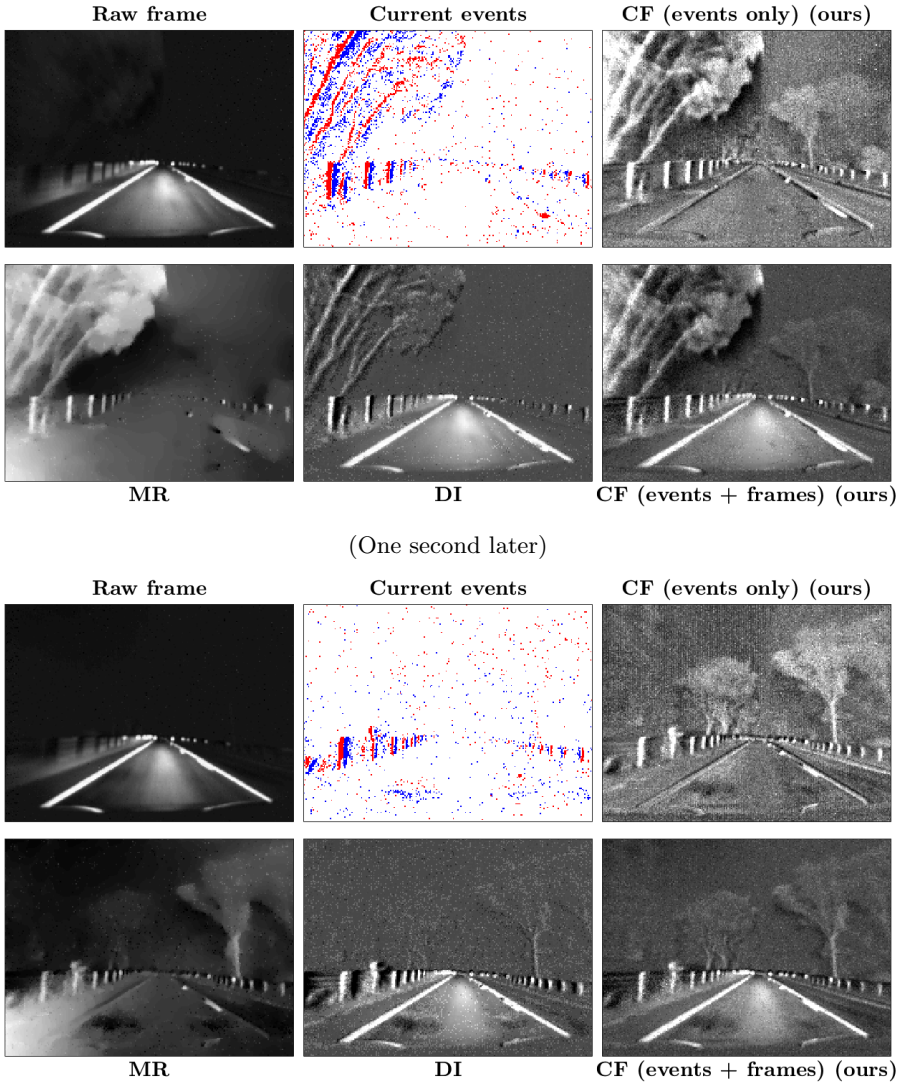

Fig. 1. Top: Night drive (at same time instance as in main paper). **Bottom:** Same sequence, one second later. Distant trees become clearer as they draw nearer.

techniques (MR, CF (events only)) struggle to recover certain aspects of the scene including road-lines and absolute intensity information (the road should be brighter than the sky).

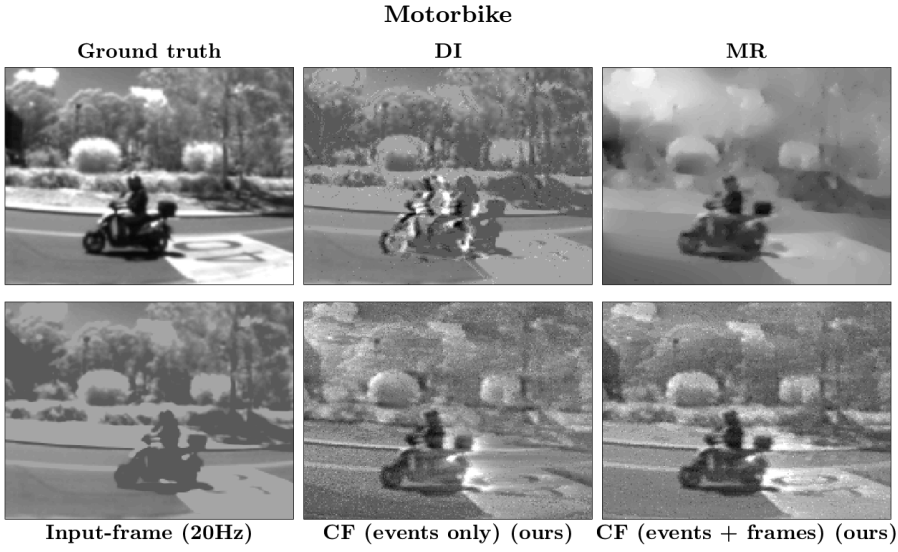

Fig. 2. CF shows good resemblance to ground truth despite added noise in the event-stream, and temporally sparse, lagged, low-dynamic-range input-frames. DI displays separation of moving objects between input-frames and event information. MR tends to over-smooth the background, which generates fewer events than the foreground.

1.2 Ground truth

Figure 2 displays a snapshot of Motorbike, halfway between two Input-frames. Due to the 50ms temporal lag added to Input-frame, the motorbike is significantly misaligned between Input-frame and events (Fig. 2; DI). Despite this, CF manages to smoothly incorporate both Input-frame and event information into a single estimate without introducing too many artifacts. The result is better (low-temporal-frequency) absolute intensity information than CF (events only). The "40" sign painted on the road is not apparent in Input-frame, and smoothed out in MR, however, it is recovered in CF.

1.3 Night run

Night run (Fig. 3) shows a person running across a parked car's headlights in pitch black conditions. By utilising the high temporal resolution and high dynamic range properties of the event camera, our method is able to clearly reconstruct sharp images of the person with much higher fidelity than the raw camera frame. The frame-based camera has low temporal resolution compared to the event camera, resulting in delayed perception of the running person. The time-lag between the first glimpse of the person by the event camera and the first raw frame that captures the person is 100ms (Fig. 3).

Figure 4 further demonstrates the lag between raw camera frame and reconstructed images. Note that combining event-stream with lagged raw frames

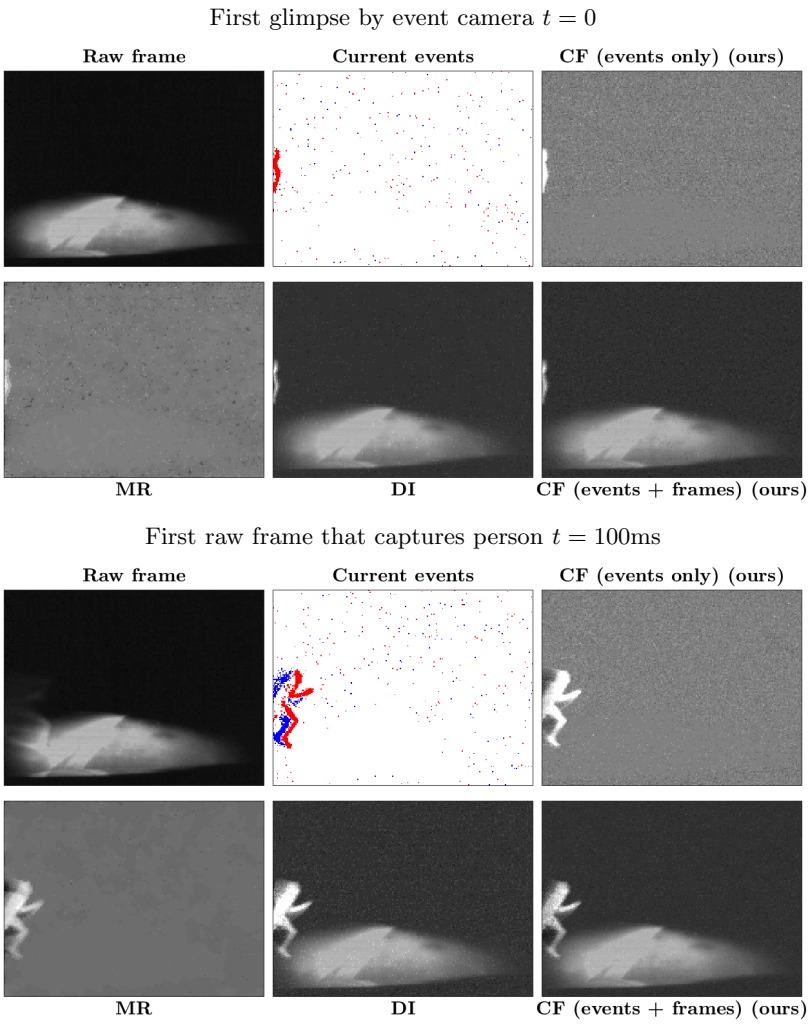

Fig. 3. Night run sequence. First glimpse of pedestrian from Raw frame is 100ms behind events. Due to high exposure-duration and thus low frame-rate of conventional camera, overall it lags events by >200ms

does not degrade temporal resolution of the CF estimate, i.e. the person is reconstructed at the temporal resolution of the event-stream. Additionally, our method (CF (events + frames)) is able to capture the area of the road lit by the headlights, as well as sharp images of the running person. In contrast, pure event-based methods capture only the running person. The frame-based camera generates heavily motion blurred images of the person even once they are in full view (Fig. 4). DI captures a sharp image most of the time, however, at each new image frame reset, the estimate is degraded to the quality of the raw frame.

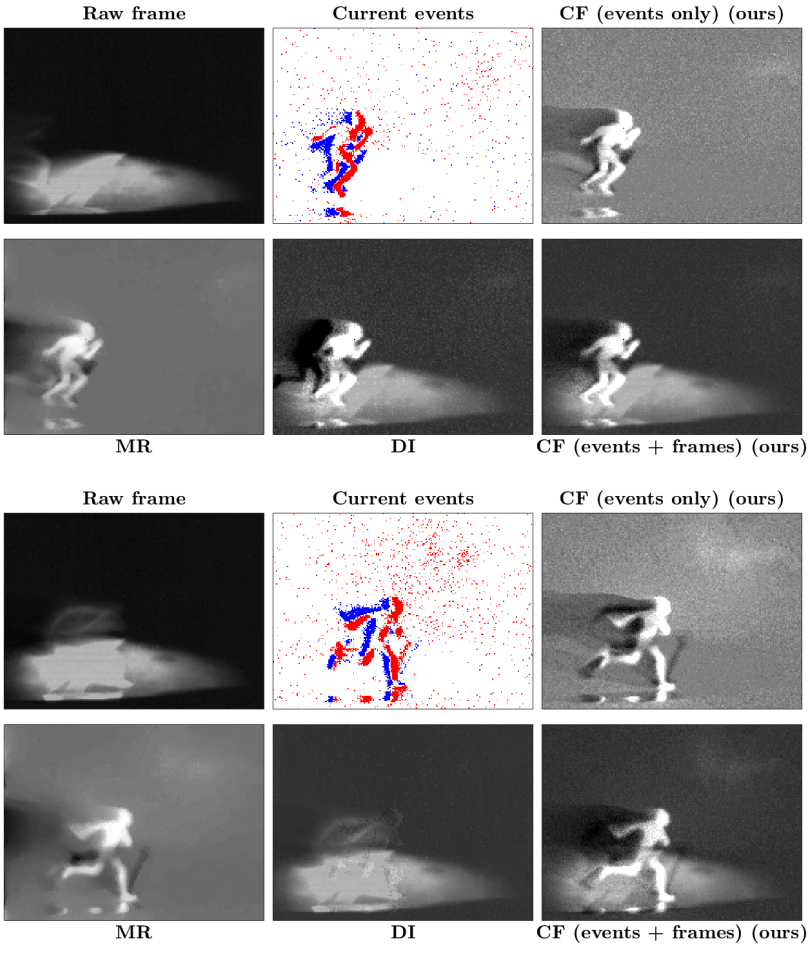

Fig. 4. Night run sequence. The frame-based camera produces a lagged, heavily motion blurred image. DI is adversely impacted by every reset (bottom). The high temporal resolution of the event camera allows sharp reconstruction of person running quickly (MR and CF). CF (events + frames) is additionally able to capture the area of the road lit by the headlights.

2 Appendix 1

A complementary filter is a continuous-time dynamical system that takes as input two or more *complementary* measurements of a state and fuses these measurements into a single state estimate [4,5,3]. In this case we treat the absolute image (log) intensity as the state that we are trying to estimate. The term complementary refers to the noise characteristics of the signals expressed in the frequency domain. The most common situation is when one of the signals is corrupted by low frequency noise such as unknown bias and slow drift while the other signal is subject to high frequency noise such as sample noise. The filter is particularly effective when the low frequency noise signal is obtained as the integral of a measured signal, such as is the case for $L^E(\boldsymbol{p}, t)$ (main paper eq. (8)). This signal is an estimate for the true log intensity $L(\boldsymbol{p}, t)$ corrupted by an unknown bias (initial condition) $L(\boldsymbol{p}, 0)$, quantisation, and other noise. The unknown bias $L(\boldsymbol{p}, 0)$ is the dominating noise effect and is low-frequency. In addition, integration of events (to obtain $L^E(\boldsymbol{p}, t)$) also amplifies low-frequency noise components. In contrast, the classical image frame information $L^F(\boldsymbol{p}, t)$ contains sample noise at camera frame rate that presents as high-frequency noise in the filter analysis.

Let $L_A(s)$ and $L_B(s)$ denote noisy measurements of $L(s)$

$$L_A(s) := L(s) + \eta(s), \quad L_B(s) := L(s) + \nu(s), \tag{1}$$

where $\eta(s)$ and $\nu(s)$ represent high- and low-frequency noise respectively in the Laplace domain. Assume that $L_B(s) = \frac{U(s)}{s}$ is the time-integral of a signal U. Let $F_L(s)$ and $F_H(s)$ denote complementary low- and high-pass filters, that is their sum $F_L(s) + F_H(s) = 1$ is an all-pass filter. The complementary state estimate of $L(s)$ is given by

$$\hat{L}(s) = F_L(s)L_A(s) + F_H(s)L_B(s) = F_L(s)L_A(s) + \frac{F_H(s)}{s}U(s). \tag{2}$$

Setting $F_L(s) := \frac{\alpha}{s+\alpha}$ to a be a causal low-pass filter and setting $F_H(s) := 1 - F_L(s) = \frac{s}{s+\alpha}$ and removing the pole-zero cancellation at the origin of $F_H(s)/s$, then the associated ordinary differential equation for the evolution of the filter state $\hat{L}$ (2) in the time-domain is given by

$$\frac{d\hat{L}(t)}{dt} + \alpha\hat{L}(t) = \alpha L_A(t) + U(t). \tag{3}$$

This is a causal differential equation that fuses the two complementary signals according to the filter characteristics shown in (2). In the proposed implementation the low-frequency signal (corresponding to $L_A(t)$) is $L^F(\boldsymbol{p}, t)$ while the high-frequency signal ($L_B(t)$) is $L^E(\boldsymbol{p}, t) = \int E(\boldsymbol{p}, \tau)\mathrm{d}\tau$. The crossover frequency α is measured in rad/s (since the filter is implemented in continuous-time) and controls how the state estimate depends on each input signal. Increasing α increases the influence of $L^F(\boldsymbol{p}, t)$. Reducing α increases dependence on $L^E(\boldsymbol{p}, t)$.

References

1. Brandli, C., Berner, R., Yang, M., Liu, S.C., Delbruck, T.: A 240x180 130dB 3us latency global shutter spatiotemporal vision sensor. IEEE J. Solid-State Circuits **49**(10), 2333–2341 (2014). https://doi.org/10.1109/JSSC.2014.2342715
2. Brandli, C., Muller, L., Delbruck, T.: Real-time, high-speed video decompression using a frame-and event-based davis sensor. In: Circuits and Systems (ISCAS), 2014 IEEE International Symposium on. pp. 686–689. IEEE (2014)
3. Franklin, G.F., Powell, J.D., Workman, M.L.: Digital control of dynamic systems, vol. 3. Addison-wesley Menlo Park, CA (1998)
4. Higgins, W.T.: A comparison of complementary and kalman filtering. IEEE Transactions on Aerospace and Electronic Systems (3), 321–325 (1975)
5. Mahony, R., Hamel, T., Pflimlin, J.M.: Nonlinear complementary filters on the special orthogonal group. IEEE Transactions on automatic control **53**(5), 1203–1218 (2008)
6. Reinbacher, C., Graber, G., Pock, T.: Real-time intensity-image reconstruction for event cameras using manifold regularisation. In: British Machine Vis. Conf. (BMVC) (2016). https://doi.org/10.5244/C.30.9