

An Asynchronous Kalman Filter for Hybrid Event Cameras

Supplementary Materials

Ziwei Wang Yonhon Ng Cedric Scheerlinck Robert Mahony
The Australian National University Systems Theory and Robotics Group
ziwei.wang1@anu.edu.au {firstname.lastname}@anu.edu.au
<https://github.com/ziweiWWANG/AKF>

1. Additional Results

1.1. Metric Analysis of HDR Performance:

In this section, we provide a detailed metric analysis of HDR performance over state-of-the-art event-based video reconstruction method ECNN [9], E2VID [6] and CF [7]. To demonstrate the performance of reconstructing raw scene radiance, we believe that the Mean Square Error (MSE) of raw intensity is the most relevant metric. We also evaluate the structural similarity of the reconstructed images using the Structural Similarity Index Measure (SSIM) [10] and Q-score [3] to provide a complete picture of the algorithm performance.

Existing hybrid event/frame datasets are not targeted at High Dynamic Range (HDR) scenarios and do not provide HDR reference images for evaluation. We have collected a new dataset, the HDR dataset documented in §4.1, that provides high quality event/frame sequences with HDR reference. On this dataset, we provide a comparison of our algorithm with the compared method ECNN [9], E2VID [6] and CF [7]. Fig. 1 shows a boxplot of the metrics MSE, SSIM and Q-score evaluated on the full HDR dataset (§4.1) using the reference HDR image as the ground truth. This provides a clear indication of the relative performance gains of our algorithm, both in *improved average* metric performance and *reduced variance* of the results.

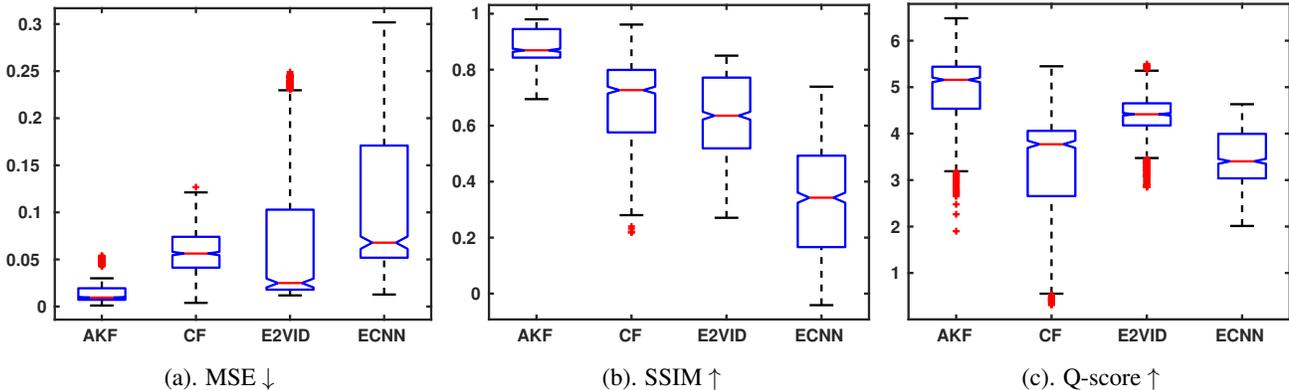


Figure 1. Boxplots of raw intensity Mean Square Error (MSE), Structural Similarity Index Measure (SSIM) [10] and the HDR Q-score measure [3], for ECNN [9], E2VID [6], CF [7] and our algorithm AKF. Results are evaluated over the full HDR datasets documented in §4.1.

The detailed results of the evaluation on the HDR dataset (§4.1) are summarised in Table 1 below. This table expands on the summary results for the HDR dataset provided in Table 1 (main paper). In the main paper, we also present quantitative results on the Artificial HDR dataset (§4.2) that we collected. The expanded results for this experiment are also shown in Table 1. The quantitative results clearly demonstrate the superior performance of our algorithm versus the compared algorithms ECNN [9], E2VID [6] and CF [7].

In addition to our HDR/AHDR dataset, we also quantitatively evaluated on the DAVIS IJRR dataset (§4.3 #1). Though it is not targeted at HDR data, it is still possible to evaluate the ability of an algorithm to reconstruct an unknown ground

Metrics	MSE ($\times 10^{-2}$) ↓				SSIM ↑				Q score ↑			
Methods	E2VID	ECNN	CF	AKF (ours)	E2VID	ECNN	CF	AKF (ours)	E2VID	ECNN	CF	AKF (ours)
Dataset	HDR sequences											
City	2.02	1.60	3.40	0.33	0.61	0.47	0.44	0.87	3.99	2.42	3.74	5.40
Trees 1	2.45	5.73	8.07	1.79	0.73	0.53	0.53	0.82	4.54	4.26	2.85	5.44
Trees 2	9.96	16.28	5.83	0.85	0.54	0.11	0.83	0.97	4.50	3.32	3.88	5.18
Trees 3	22.48	26.84	8.38	4.79	0.41	0.15	0.78	0.86	4.52	3.19	3.03	4.77
Building	1.91	6.71	5.41	0.81	0.79	0.30	0.74	0.94	4.05	3.86	1.54	3.36
Mean	7.76	11.43	6.22	1.71	0.616	0.31	0.66	0.89	4.32	3.41	3.01	4.83
Dataset	Artificial HDR sequences (AHDR)											
Mountain slow	17.19	33.03	7.43	6.49	0.53	-0.14	0.57	0.71	5.36	2.83	5.19	5.87
Mountain fast	17.75	36.69	7.87	6.47	0.56	-0.23	0.52	0.70	5.34	2.76	4.47	5.19
Lake slow	4.83	7.99	2.72	1.81	0.48	0.22	0.70	0.80	5.18	3.82	5.01	5.76
Lake fast	6.47	7.21	3.11	1.93	0.43	0.30	0.70	0.80	5.07	4.01	4.45	5.35
Mean	11.56	21.23	5.28	4.18	0.50	0.04	0.62	0.75	5.24	3.36	4.78	5.54

Table 1: Comparison of state-of-the-art event-based video reconstruction methods E2VID [6], ECNN [9] and CF [7] on our HDR (§4.1) and AHDR dataset (§4.2). Our AKF outperforms the compared methods in most scenarios.

Metrics	MSE ($\times 10^{-2}$) ↓				SSIM [10] ↑				LPIPS [12] ↓			
Methods	E2VID	ECNN	CF	AKF (ours)	E2VID	ECNN	CF	AKF (ours)	E2VID	ECNN	CF	AKF (ours)
boxes_6dof	11.87	3.99	0.33	0.26	0.51	0.62	0.78	0.80	0.34	0.24	0.26	0.26
calibration	23.46	3.05	0.11	0.09	0.43	0.65	0.92	0.92	0.31	0.18	0.09	0.07
dynamic_6dof	30.96	14.11	0.13	0.12	0.24	0.30	0.88	0.87	0.46	0.36	0.18	0.20
office_zizang	17.14	3.95	0.26	0.22	0.40	0.49	0.85	0.86	0.40	0.25	0.22	0.22
poster_6dof	21.68	6.86	0.29	0.26	0.34	0.46	0.78	0.79	0.35	0.22	0.22	0.24
shapes_6dof	19.44	8.77	0.16	0.11	0.68	0.76	0.92	0.94	0.31	0.18	0.16	0.15
slider_depth	20.13	4.16	0.21	0.18	0.44	0.61	0.83	0.86	0.40	0.23	0.22	0.21
Mean	20.67	6.41	0.21	0.18	0.43	0.56	0.85	0.86	0.37	0.24	0.19	0.19

Table 2: Comparison of state-of-the-art methods of event-based video reconstruction on IJRR [2] DAVIS datasets (§4.3 #1). Though both CF [7] and AKF perform well in the structural similarity metrics SSIM [10] and LPIPS [12], our AKF outperforms other methods with a significant margin in the absolute intensity metrics MSE.

truth image by sub-sampling the frame data. To quantitatively evaluate on DAVIS datasets, we use every second image frame as input for the algorithm and then take the intermediate image frame as ground truth for evaluation of the quality of reconstruction using the quantitative metrics. The full event stream is used in the algorithm.

We have used the same sequences and frames as used in [9] and [6] for the evaluation study. In addition to the MSE and SSIM, we also evaluated on the learned perceptual image patch similarity (LPIPS) [12]. Table 2 shows that for MSE and SSIM, the AKF is almost always the best, though CF [7] is roughly equal on LPIPS. We believe the reduced gap in performance between CF [7] and AKF is due to ‘cleaner’ frame data that is not under/overexposed as in our HDR dataset (§4.1). The large difference in MSE between AKF and E2VID [6]/ECNN [9] is not as apparent in SSIM and LPIPS, agreeing with the intuition that pure event reconstruction methods are relatively more faithful to scene structure than the absolute intensity.

1.2. Visual Analysis of HDR Performance:

It is extremely difficult to measure HDR performance from reproduced images visually, especially after tonemapping or normalisation for print reproduction. In the main paper, we have chosen a set of public DAVIS event camera datasets that highlight specific issues in the HDR reconstruction. `Boxes_6dof` and `Outdoors_running` in Dataset #1, `Night_drive` in Dataset #2 and `Shadow` in Dataset #3 of §4.3 include a range of challenging scenarios such as night/bright outdoor lighting conditions, driving, running, fast/slow motions, static/moving background, dynamic/static objects.

In Fig. 2, we demonstrate more examples on sequence `Night_drive` in ACD dataset (§4.3 Dataset #2) and `Boxes_6dof`

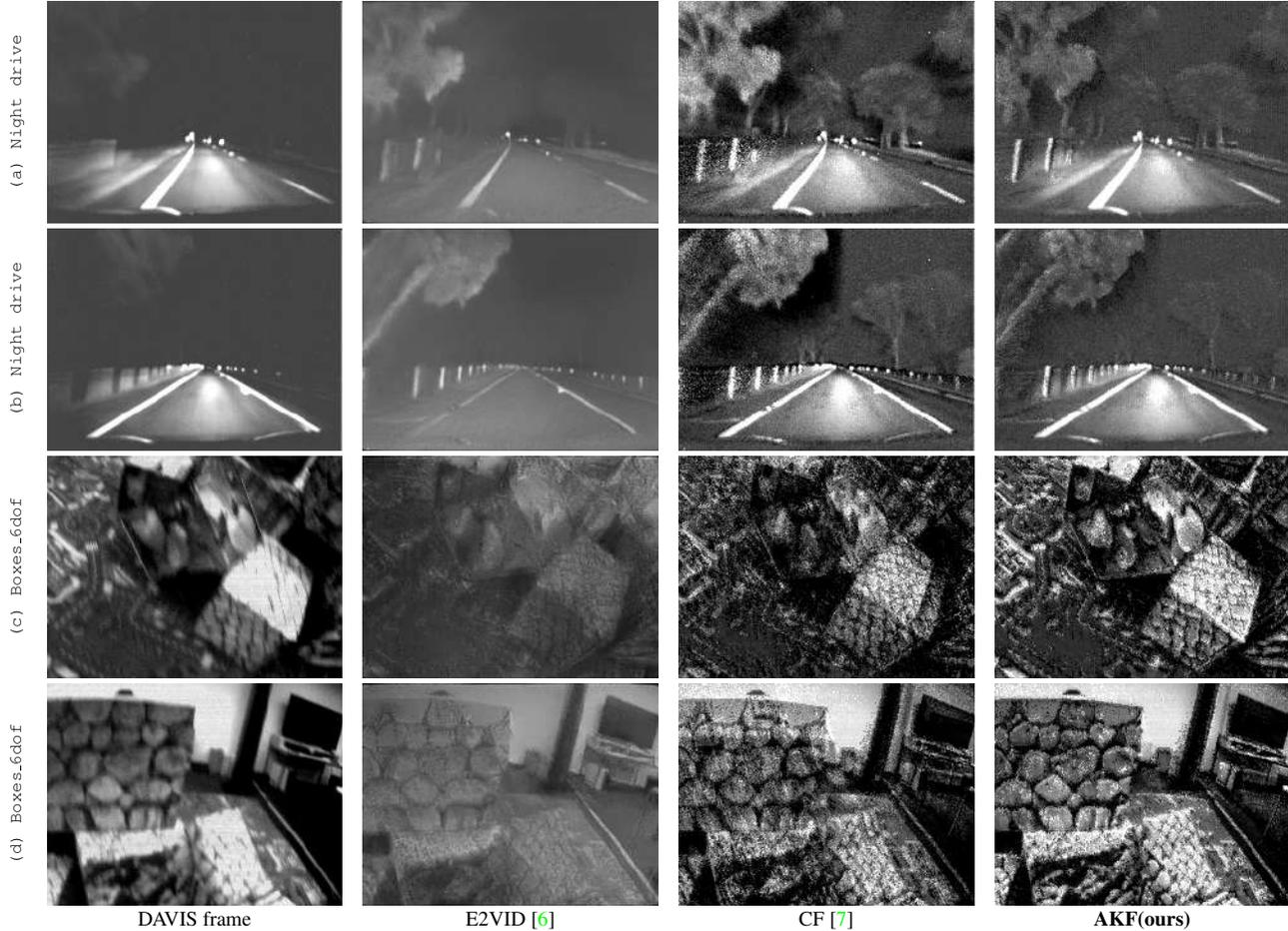


Figure 2. Comparison of state-of-the-art event-based video reconstruction methods on sequences with challenging lighting conditions and fast motions, drawn from the open-source datasets ACD [7] and IJRR [2]. We demonstrate the different time instances of the same dataset in the main paper. E2VID [6] fails to recover trees on the right-hand side in *Night_drive* and details on the carpet in *Boxes_6dof*. The method provides washed-out reconstructions that fail to capture the true image intensity. CF [7] leads to dark shadows trailing behind the moving objects (*e.g.* trees and roadside poles in *Night_drive*). AKF provides sharp, clean and high dynamic reconstructions under the challenging scenarios. Video comparisons are provided in the supplementary video materials.

in IJRR dataset (§4.3 Dataset #1) at different time instances of the same dataset in the main paper. In Fig. 2(a) and (b), the DAVIS frames only capture blurry roadside poles and a small part of the background trees. E2VID [6] produces washed-out reconstructions of the near trees and the roadside poles. The background trees on the right-hand side can not be captured properly. By fusing DAVIS frames, CF [7] captures more detailed background trees and provides more accurate intensities. However, the reconstructions of CF [7] include ‘hot pixels’ (constantly fired pixels) and noisy shadows trailing behind the fast-moving objects. Our AKF overcomes the limitations of CF [7] by dynamically adjusting the Kalman gain based on the event and frame uncertainty. For CF [7], the trade-off between fusing image frames and event stream data is fixed by a chosen constant gain in the filter. In comparison, our AKF encodes the uncertainty in underexposed and overexposed image pixels and allows the reconstruction to exploit the full information in the event stream where the image frame information is negligible, while still exploiting the image frame information where it is useful, *e.g.*, the visible trees, the roadside poles, and the road markings. For CF [7], the ‘hot pixels’ affect the reconstruction performance, but for AKF, the ‘hot pixels’ have high isolated pixel uncertainty because they are spatially and temporally isolated from other events. Therefore, AKF relies more on frame data/state on these pixels so that is not affected by the ‘hot pixels’ in *Night_drive* sequence.

In Fig. 2(c) and (d), the fast camera motion causes blurry DAVIS frames and the lack of textures on boxes and the carpet. The batch method E2VID [6] accumulates events within a certain time period before generating a frame reconstruction from the event batch. The reconstruction quality is highly dependent on the quantity and quality of the corresponding event data.

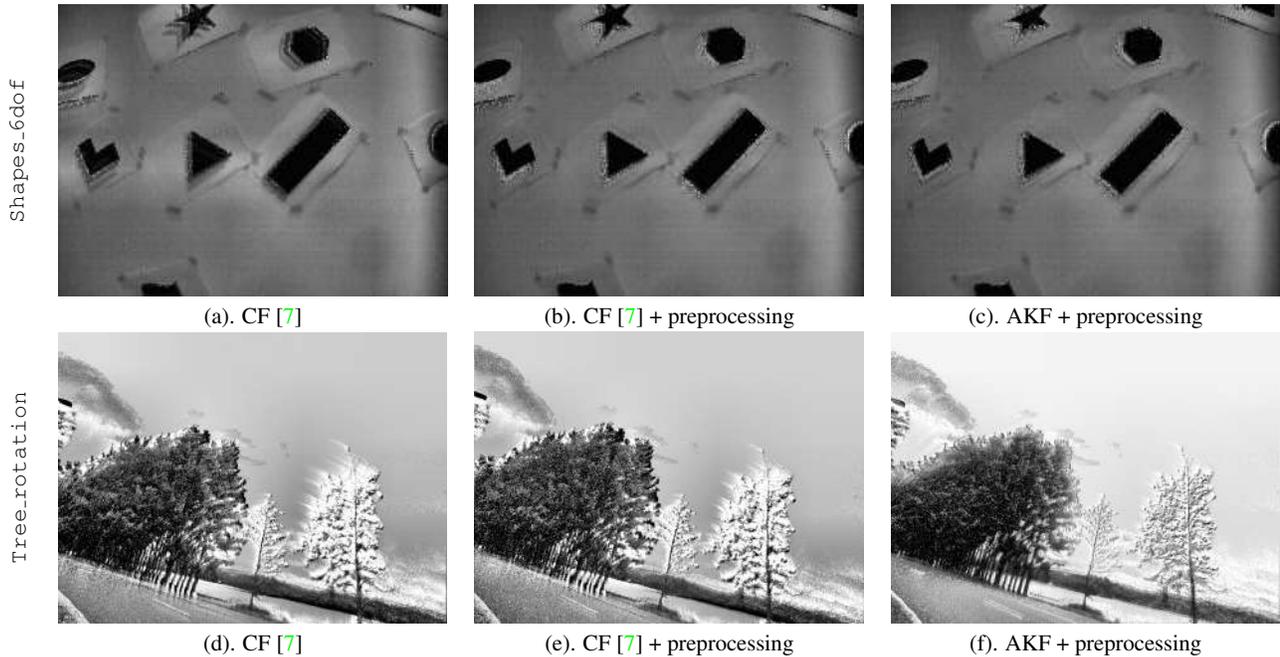


(a). LDR frame

(b). Frame Augmentation

(c). Full AKF Pipeline

Figure 3. An example of image reconstruction using our frame augmentation based on EDI [5] and our full AKF pipeline. The de-blurring method EDI [5] computes the grid in the overexposed region using events that happens in the frame exposure time but fails to recover the intensity from white (overexposed) to grey. Our AKF recovers the intensity difference on the grid successfully using the state ‘memory’ that associates with the previous frame/event data.



(a). CF [7]

(b). CF [7] + preprocessing

(c). AKF + preprocessing

(d). CF [7]

(e). CF [7] + preprocessing

(f). AKF + preprocessing

Figure 4. An example of applying our frame augmentation preprocessing algorithm on CF [7]. Adding the preprocessing step helps to reduce the ‘double edges’ in the *Shapes_6dof* with fast camera motion. However, it fails to recover the HDR region (*e.g.* the white trialling shadows in *Tree.rotation*).

When the camera motion is high under a highly textured scene, pixels will not be triggered fast enough due to the refractory period [11], so the event quality is degraded. This leads to the poor reconstruction for E2VID [6], while our AKF models this type of noise (main paper §3.1.1) and managed to produce better image reconstruction. The E2VID [6] algorithm also fails to reconstruct the correct intensity information from DAVIS frames as expected when the scene brightness to be reconstructed varies significantly from the training data. CF [7] fuses the image frame data and consequently remains faithful to the true intensity values in the image. However, the simple zero-order hold assumption (fusing event data with the previous frame data) cannot account for fast camera motion. This leads to obvious ‘double edges’ where image frame data from previous frames is not fully compensated in the reconstruction. In comparison, our AKF provides sharp and clean reconstruction in the challenging scenarios with high dynamic range and fast motions. Please refer to our supplementary video for more comparisons.

2. Frame Augmentation versus AKF

The proposed algorithm benefits from two key contributions outlined in the main paper, the frame augmentation process and the asynchronous Kalman filter. The frame augmentation is based directly on the established performance of the EDI scheme [5] and also provides a more sophisticated temporal interpolation.

To analyse the effect of AKF above and beyond the frame augmentation step, we run an experiment with frame augmentation and the full AKF pipeline in Fig. 3. In our frame augmentation step, EDI [5] generates sharp images by associating frames to the event data during the exposure time. Since EDI [5] is not designed for HDR task [4] and the exposure time of each frame is small, only pixels with events that occurred during the short time can be recovered properly (around the grid edges) in Fig. 3. As a de-blurring method, it recovers some of the HDR regions surprisingly well although the overall reconstruction is clearly still overexposed. The AKF recovers the correct intensities for the overexposed region by exploiting the inherent ‘memory’ in the state of the Kalman filter.

We also compare the CF [7] with and without the frame augmentation in Fig. 4(a)-(b) and (d)-(e). It demonstrates that our AKF outperforms the simple modelled filter algorithm CF [7] even with frame augmentation. The fast camera motion of the dataset *Shapes_6dof* leads to the large difference between two consecutive input frames. Without temporal frame interpolation between two frames, CF [7] uses the outdated frame data (in a zero-order hold assumption) that leads to the ‘double edges’ in Fig. 4(a). The addition of preprocessing provides CF [7] with more accurate reference images which get rid of most of the ‘double edges’ in Fig. 4(b). The performance of CF [7] with the preprocessing step is similar to our full AKF pipeline when the DAVIS frames are reliable (with no over or underexposed region). However, in HDR scenario, adding the preprocessing step still cannot provide an accurate HDR reference image to CF [7]. The ‘shadowing effect’ (e.g. white trialling shadows behind trees) is still obvious in Fig. 4(d)-(e). Our AKF produces a significantly better result by dynamically adjusting gain corresponding to our noise model in Fig. 4(f). This is particularly clear in the tree in the bottom right of the image.

3. Algorithm

Algorithm 1 Event-based Continuous-time Intensity Estimation Using Asynchronous Kalman Filter

- 1: Initialise variables
 - 2: **for** New i^{th} event at pixel p , $e(t_p^i)$ **do**
 - 3: **if** new image frame arrives **then**
 - 4: Deblur new image based on [5]
 - 5: Compute $c(\tau^k)$ based on Eq (10)
 - 6: **end if**
 - 7: Update augmented frame $\hat{L}_p^A(t_p^i)$ based on Eq (8)- (11)
 - 8: Update image covariance $R_p(t_p^i)$ based on Eq (4)- (7)
 - 9: Update state $\hat{L}_p(t_p^i)$ based on Eq (14)-(16)
 - 10: Update covariance $P_p(t_p^i)$ based on Eq (18)-(19), where Q is designed based on discussions in §3.1
 - 11: **if** publishing new image **then**
 - 12: **for** all pixels q **do**
 - 13: Update state $\hat{L}_q(t_q^i)$ based on Eq (16)
 - 14: Update covariance $P_q(t_q^i)$ based on Eq (18)-(19), where Q corresponds to zero event covariance
 - 15: Write image
 - 16: **end for**
 - 17: **end if**
 - 18: **end for**
-

The pseudocode of our proposed algorithm is shown above (Algorithm 1). The time complexity of asynchronous filtering methods, AKF and CF [7], are both event-wise operations, where they perform a linear combination of several $\mathcal{O}(1)$ operations, resulting in an overall $\mathcal{O}(1)$ complexity per event. Besides, for each image, the heaviest time complexity of frame augmentation is contrast threshold calibration, which has the complexity of $\mathcal{O}(N + M)$, where N is the number of events between two images, and each image contains M pixels.

4. Dataset List

In this section, we list all datasets we used in the paper and supplementary material. The datasets developed in this paper are the HDR (§4.1) and AHDR (§4.2) dataset with HDR references.

4.1. Our HDR Dataset

Though DAVIS event cameras allow easy access to dual event/frame data, they are subject to significant shutter noise due to electrical coupling in the pixel circuits [1]. The hybrid event/frame configuration of DAVIS also limits the resolution of both sensors. The current state-of-the-art hybrid event camera is DAVIS346 with a limited 0.1 Megapixels resolution of both event and frames. However, both pure event camera and frame-based camera are developing much faster than the hybrid event camera. For example, the pure event camera Samsung Gen4 has achieved a resolution of 1280×960 (around 1 Megapixel) and the conventional RGB camera can easily acquire higher video resolution. The sensor noise and low image resolution impact the utility of DAVIS camera data for HDR reconstruction and further motivates our approach and the new dataset we provide in the paper. Our stereo configuration also models the scenario where a pure event sensor camera is added to a suite of separate vision sensors such as the most likely use case for modern mobile phones and robotic systems.

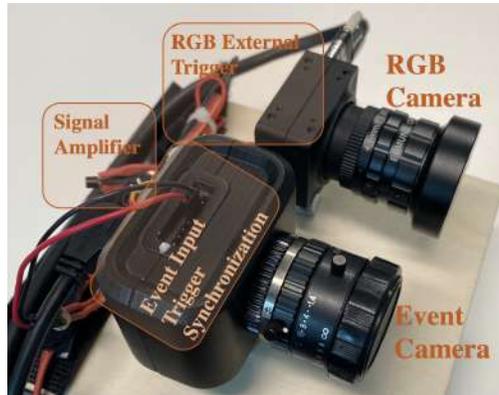


Figure 5. Our Hybrid event/frame System

The stereo hybrid event/frame camera prototype we built is shown in Fig. 5. We use the camera system to collect events, frames and HDR reference images. The dataset sequences focus on different HDR scenes with different camera motion speeds, which is challenging for all event-based HDR image reconstruction methods. The reference HDR image is generated from several low dynamic range raw images taken by the RGB camera at different exposures. The images are fused using a traditional multi-exposure image fusion method followed by an image warp to register the reference image with each frame. The details of our proposed HDR event/frame dataset are summarised below.

HDR Dataset	# of images	Speed	Description	HDR Scene	Reference Image
Dataset	HDR sequences				
city	150	medium	rooftop overlooking city buildings	dark buildings	✓
trees 1	208	slow and fast	a car under tree shadow	tree shadow	✓
trees 2	150	medium	parking lot, buildings and clouds	bright far-away buildings	✓
trees 3	150	medium	trees partially covered by shadow and far away buildings	buildings and shadow	✓
building	150	medium	mountain, tower, building and plants	dark plants	✓



(a). city

(b). tree 1

(c). tree 2

(d). tree 3

(e). building

4.2. Our AHDR Dataset

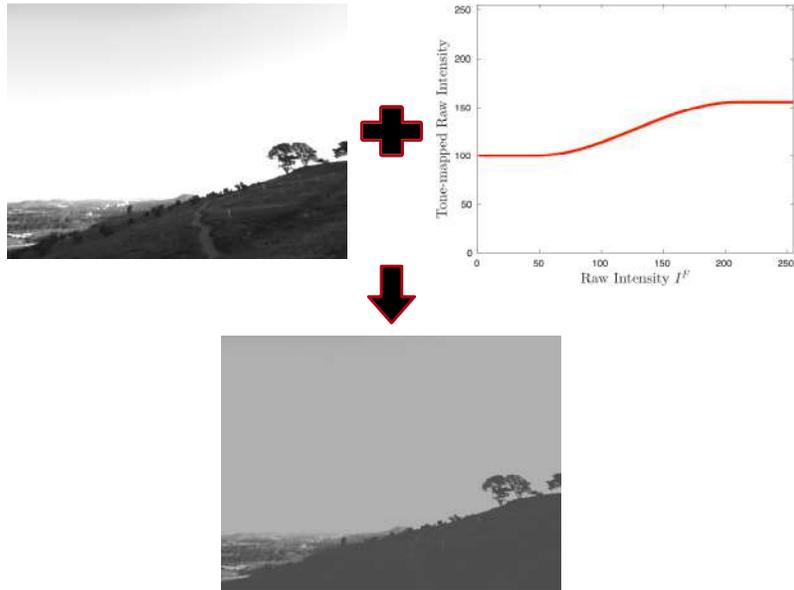
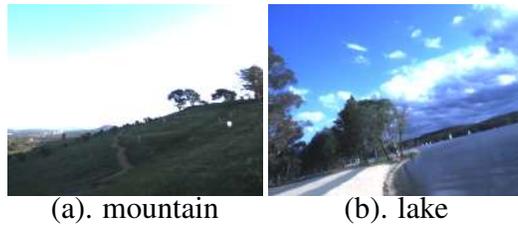


Figure 6. Process of generating our AHDR dataset. We simulate a low dynamic range camera by applying an artificial camera response function to real images.

For our AHDR dataset, we apply an artificial camera response function to RGB camera output frames to simulate a low dynamic range camera (see Fig. 6). Following the process introduced in the main paper §3.2.1, we experimentally determine the corresponding camera uncertainty function of the low dynamic range camera, where the resulting image noise covariance is high for the ‘cropped’ intensity values. Details about the AHDR dataset are as follows.

HDR Dataset	# of images	Speed	Description	HDR Scene	Reference Image
Dataset	Artificial HDR sequences (AHDR)				
mountain × 2	150	slow and fast	mountain with road overlooking city	dark grass and road	✓
lake × 2	200	slow and fast	side of lake with trees and road during cloudy day	trees and clouds	✓



4.3. Existing DAVIS Dataset

For completeness, we also selected a number of challenging sequences (with HDR scene, fast motion or dynamic objects) from the existing open-source datasets [2, 7, 8] for evaluation, with details as follows.

HDR Dataset	# of images	Speed	Description	HDR Scene	Reference Image
Dataset #1		IJRR dataset [2]			
boxes_6dof	1298	increasing speed	highly textured environment	boxes and carpet	✗
outdoors_running	1573	running speed	sunny urban environment	buildings	✗
calibration	1422	slow	checkerboard (6x7, 70mm)	-	✗
dynamic_6dof	1267	increasing speed	office with moving person	-	✗
office_zizang	248	slow	office environment	-	✗
poster_6dof	1358	increasing speed	wallposter	-	✗
shapes_6dof	1356	increasing speed	simple shapes on a wall	-	✗
slider_depth	87	constant speed	objects at different depths	-	✗
Dataset #2		ACD dataset [7]			
night_drive	1058	high speed	low-light driving	roadside signs, poles and trees	✗
Dataset #3		CED dataset [8]			
shadow	576	slow	static background, moving board	board	✗

5. Derivation of $P_{\mathbf{p}}(t)$

Within a time-interval $t \in [t_{\mathbf{p}}^i, t_{\mathbf{p}}^{i+1})$, the ODE of $P_{\mathbf{p}}(t)$ is

$$\frac{1}{P_{\mathbf{p}}^2(t)} \frac{dP_{\mathbf{p}}(t)}{dt} = -R_{\mathbf{p}}^{-1}(t).$$

Moving dt to the right hand side yields

$$\frac{1}{P_{\mathbf{p}}^2(t)} dP_{\mathbf{p}}(t) = -R_{\mathbf{p}}^{-1}(t) dt.$$

Integrate from event time $t_{\mathbf{p}}^i$ to time t

$$\begin{aligned} \int_{t_{\mathbf{p}}^i}^t \frac{1}{P_{\mathbf{p}}^2(t)} dP_{\mathbf{p}}(t) &= \int_{t_{\mathbf{p}}^i}^t -R_{\mathbf{p}}^{-1}(t) dt, \\ -P_{\mathbf{p}}^{-1}(t) &= -R_{\mathbf{p}}^{-1}(t) \cdot (t - t_{\mathbf{p}}^i) + C_1, \\ P_{\mathbf{p}}(t) &= \frac{1}{R_{\mathbf{p}}^{-1}(t) \cdot (t - t_{\mathbf{p}}^i) - C_1}, \end{aligned}$$

where C_1 is a constant number.

Let $t = t_{\mathbf{p}}^i$ and we have

$$\begin{aligned} P(t_{\mathbf{p}}^i) &= \frac{1}{-C_1}, \\ C_1 &= -P_{\mathbf{p}}^{-1}(t_{\mathbf{p}}^i). \end{aligned}$$

The solution of the ODE of $P_{\mathbf{p}}(t)$ is

$$P_{\mathbf{p}}(t) = \frac{1}{P_{\mathbf{p}}^{-1}(t_{\mathbf{p}}^i) + R_{\mathbf{p}}^{-1}(t) \cdot (t - t_{\mathbf{p}}^i)}.$$

6. Derivation of $\hat{L}_{\mathbf{p}}(t)$

Within a time-interval $t \in [t_{\mathbf{p}}^i, t_{\mathbf{p}}^{i+1})$, the ODE of $\hat{L}_{\mathbf{p}}(t)$ is

$$\begin{aligned}\dot{\hat{L}}_{\mathbf{p}}(t) &= -\frac{R_{\mathbf{p}}^{-1}(t) \cdot [\hat{L}_{\mathbf{p}}(t) - L_{\mathbf{p}}^A(t)]}{P_{\mathbf{p}}^{-1}(t_{\mathbf{p}}^i) + R_{\mathbf{p}}^{-1}(t) \cdot (t - t_{\mathbf{p}}^i)}, \\ \frac{d[\hat{L}_{\mathbf{p}}(t) - L_{\mathbf{p}}^A(t)]}{\hat{L}_{\mathbf{p}}(t) - L_{\mathbf{p}}^A(t)} &= -\frac{R_{\mathbf{p}}^{-1}(t)}{P_{\mathbf{p}}^{-1}(t_{\mathbf{p}}^i) + R_{\mathbf{p}}^{-1}(t) \cdot (t - t_{\mathbf{p}}^i)} dt.\end{aligned}$$

Integrate from the event time $t_{\mathbf{p}}^i$ to time t

$$\begin{aligned}\ln(\hat{L}_{\mathbf{p}}(t) - L_{\mathbf{p}}^A(t)) &= \int_{t_{\mathbf{p}}^i}^t -\frac{R_{\mathbf{p}}^{-1}(\gamma)}{P_{\mathbf{p}}^{-1}(t_{\mathbf{p}}^i) + R_{\mathbf{p}}^{-1}(\gamma) \cdot (\gamma - t_{\mathbf{p}}^i)} d\gamma \\ &= -\ln(P_{\mathbf{p}}^{-1}(t_{\mathbf{p}}^i) + R_{\mathbf{p}}^{-1}(t) \cdot (t - t_{\mathbf{p}}^i)) + C_2,\end{aligned}$$

where C_2 is a constant number. Take exponential of both sides

$$\hat{L}_{\mathbf{p}}(t) - L_{\mathbf{p}}^A(t) = \frac{1}{P_{\mathbf{p}}^{-1}(t_{\mathbf{p}}^i) + R_{\mathbf{p}}^{-1}(t) \cdot (t - t_{\mathbf{p}}^i)} \cdot e^{C_2},$$

and we define $C_3 = e^{C_2}$.

Let $t = t_{\mathbf{p}}^i$ and we have

$$\begin{aligned}\hat{L}_{\mathbf{p}}(t_{\mathbf{p}}^i) - L_{\mathbf{p}}^A(t_{\mathbf{p}}^i) &= \frac{1}{P_{\mathbf{p}}^{-1}(t_{\mathbf{p}}^i)} \cdot C_3, \\ C_3 &= [\hat{L}_{\mathbf{p}}(t_{\mathbf{p}}^i) - L_{\mathbf{p}}^A(t_{\mathbf{p}}^i)] \cdot P_{\mathbf{p}}^{-1}(t_{\mathbf{p}}^i).\end{aligned}$$

The solution of the ODE of $\hat{L}_{\mathbf{p}}(t)$ is

$$\hat{L}_{\mathbf{p}}(t) = \frac{[\hat{L}_{\mathbf{p}}(t_{\mathbf{p}}^i) - L_{\mathbf{p}}^A(t_{\mathbf{p}}^i)] \cdot P_{\mathbf{p}}^{-1}(t_{\mathbf{p}}^i)}{P_{\mathbf{p}}^{-1}(t_{\mathbf{p}}^i) + R_{\mathbf{p}}^{-1}(t) \cdot (t - t_{\mathbf{p}}^i)} + L_{\mathbf{p}}^A(t).$$

References

- [1] Christian Brandli, Raphael Berner, Minhao Yang, Shih-Chii Liu, and Tobi Delbruck. A 240×180 130 db $3 \mu\text{s}$ latency global shutter spatiotemporal vision sensor. *IEEE Journal of Solid-State Circuits*, 49(10):2333–2341, 2014. [6](#)
- [2] Elias Mueggler, Henri Rebecq, Guillermo Gallego, Tobi Delbruck, and Davide Scaramuzza. The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and SLAM. *Int. J. Robot. Research*, 36(2):142–149, 2017. [2](#), [3](#), [7](#), [8](#)
- [3] Manish Narwaria, Rafal Mantiuk, Mattheiu P Da Silva, and Patrick Le Callet. Hdr-vdp-2.2: a calibrated method for objective quality prediction of high-dynamic range and standard images. *Journal of Electronic Imaging*, 24(1):010501, 2015. [1](#)
- [4] Liyuan Pan, Richard I. Hartley, Cedric Scheerlinck, Miaomiao Liu, Xin Yu, and Yuchao Dai. High frame rate video reconstruction based on an event camera. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2020. [5](#)
- [5] Liyuan Pan, Cedric Scheerlinck, Xin Yu, Richard Hartley, Miaomiao Liu, and Yuchao Dai. Bringing a blurry frame alive at high frame-rate with an event camera. In *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2019. [4](#), [5](#)
- [6] Henri Rebecq, René Ranftl, Vladlen Koltun, and Davide Scaramuzza. High speed and high dynamic range video with an event camera. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2020. [1](#), [2](#), [3](#), [4](#)
- [7] Cedric Scheerlinck, Nick Barnes, and Robert Mahony. Continuous-time intensity estimation using event cameras. In *Asian Conf. Comput. Vis. (ACCV)*, 2018. [1](#), [2](#), [3](#), [4](#), [5](#), [7](#), [8](#)
- [8] Cedric Scheerlinck, Henri Rebecq, Timo Stoffregen, Nick Barnes, Robert Mahony, and Davide Scaramuzza. CED: Color event camera dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. [7](#), [8](#)

- [9] Timo Stoffregen, Cedric Scheerlinck, Davide Scaramuzza, Tom Drummond, Nick Barnes, Lindsay Kleeman, and Robert Mahony. Reducing the Sim-to-Real gap for event cameras. In *Eur. Conf. Comput. Vis. (ECCV)*, 2020. 1, 2
- [10] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4):600–612, Apr. 2004. 1, 2
- [11] Minhao Yang, Shih-Chii Liu, and Tobi Delbruck. A dynamic vision sensor with 1% temporal contrast sensitivity and in-pixel asynchronous delta modulator for event encoding. *IEEE Journal of Solid-State Circuits*, 50(9):2149–2160, 2015. 4
- [12] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2018. 2