CED: Color Event Camera Dataset

Cedric Scheerlinck [†*]   Henri Rebecq [‡*]   Timo Stoffregen [§]   Nick Barnes [†]   Robert Mahony [†]

Davide Scaramuzza [‡]

Abstract

Event cameras are novel, bio-inspired visual sensors, whose pixels output asynchronous and independent timestamped spikes at local intensity changes, called 'events'. Event cameras offer advantages over conventional frame-based cameras in terms of latency, high dynamic range (HDR) and temporal resolution. Until recently, event cameras have been limited to outputting events in the intensity channel, however, recent advances have resulted in the development of color event cameras, such as the Color-DAVIS346. In this work, we present and release the first Color Event Camera Dataset (CED), containing 50 minutes of footage with both color frames and events. CED features a wide variety of indoor and outdoor scenes, which we hope will help drive forward event-based vision research. We also present an extension of the event camera simulator ESIM [1] that enables simulation of color events. Finally, we present an evaluation of three state-of-the-art image reconstruction methods that can be used to convert the Color-DAVIS346 into a continuous-time, HDR, color video camera to visualise the event stream, and for use in downstream vision applications.

Website: http://rpg.ifi.uzh.ch/CED

1. Introduction

Since their recent addition to the computer vision community [2], event cameras have challenged conventional thinking about how to solve computer vision problems. Instead of producing global-shutter images at a fixed frame-rate as in conventional cameras, event cameras have pixels that operate independently and asynchronously. When the brightness change at a given pixel exceeds a threshold, that pixel emits an event containing its (x, y) address, timestamp and polarity. Event cameras offer several advantages; they

---

[*] Equal contribution.

[†] Australian National University, Canberra, ACT, Australia.

[‡] Dept. Informatics, Univ. of Zurich and Dept. Neuroinformatics, Univ. of Zurich and ETH Zurich.

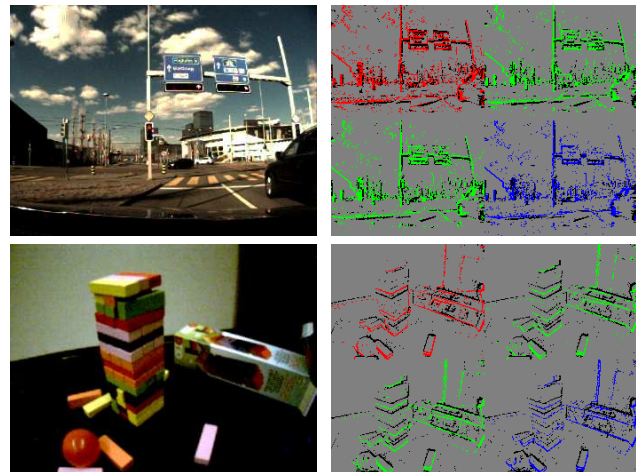[§] Monash University, Melbourne, VIC, Australia.

Figure 1: Our Color Event Camera Dataset (CED) features both outdoor (top row) and indoor (bottom row) sequences, and provides color images (left column) and color events (right row) from the Color-DAVIS346 for each sequence.

sample at the rate of scene dynamics without having to wait for an external shutter cycle, and the output is data-driven and non-redundant. This means that event cameras have extremely low latency, low power consumption and bandwidth requirements, high dynamic range and suffer essentially no motion blur. The temporal resolution of current event cameras is in the order of microseconds.

Since their introduction, event cameras have spawned a flurry of research. They have been used in feature detection and tracking [3–6], depth estimation [7–10], stereo [11–14], optical flow [15–18], image reconstruction [19–25], localization [26–29], SLAM [30–32], visual-inertial odometry [33–36], pattern recognition [37–40], and more. In response to the growing needs of the community, several important event-based vision datasets have been released, directed at popular topics such as SLAM [28], optical flow [41, 42] and recognition [37, 43]. Event camera datasets enable better benchmarking and re-producibility, and grant researchers access to high quality event data in a range of environments without necessarily having to acquire an expensive event camera.

While existing datasets are limited to monochrome events, event camera technology has since advanced to allow color events and frames [44], which opens the door to a new generation of color event processing.

Figure 2: "DAVIS346 Red Color" camera used for dataset collection.

The addition of color information to event-based vision has the potential to improve performance of many tasks, such as segmentation [45] and recognition, where it is known that color is an important source of visual information [46]. Early works have shown promising results using prototype color event cameras [47–49], or a mirrored-rig with three monochrome cameras and three color filters [45], however, to-date there are no publicly available color event datasets. Further, the wider research community has limited access to color event cameras, hindering progress into color event vision research.

We present the first *Color Event Camera Dataset* (Fig. 1) that aims to spur research into color event vision by providing the community with high quality color event data, alongside color frames from the Color-DAVIS346 [44]. The Color-DAVIS346 (Fig. 2) is the latest color event camera, built upon the popular line of DAVIS cameras that many existing datasets and research is based off. Rather than directing our focus at a specific target application, we aim to cater for general purpose vision research by including a diverse range of scenes (simple objects, indoor/outdoor scenes, people), lighting conditions (daylight, indoor light, low-light), camera motions (linear, 6-DOF motion) and dynamics. While we do not provide ground truth labels for any specific task (*e.g.* optical flow estimation, object detection, etc.), we provide color images from the sensor that are naturally synchronized and registered to events. These images may be used to generate proxy labels for any task of interest (using either conventional computer vision, or manual annotation) that can be transferred to the events.

To visually unveil the color information contained in color events, we evaluate and compare three state-of-the-art event-based image reconstruction methods [22, 24, 50] on our Color Event Camera Dataset. Image reconstruction is an active field of event-based vision research [6, 19–24, 50] that allows visualisation of the event stream, and enables application of decades of computer vision research and expertise on event data, which in its raw form is inaccessible to powerful tools such as convolutional neural networks. Further, event reconstructed images have the potential to retain desirable qualities of event cameras, such as high dynamic range, high temporal resolution and immunity to motion blur.

Contributions:

1. We present CED: Color Event Camera Dataset containing 50 minutes of both color events and frames in a wide range of natural scenes with static and dynamic objects, and covering a variety of camera-motions from simple translations and rotations to unconstrained 6-DOF motions.

2. We release a color event camera simulator, based on ESIM [1].

3. We present color video reconstructions from a color event camera, comparing three state-of-the-art reconstruction methods. Video reconstruction provides a natural way to visualize the event stream and enable image-based processing on events.

2. Related Works

Many event-based vision datasets have been published since the introduction of the DVS [2]. Most of these datasets were recorded using a DAVIS [51] event camera or similar and have a particular use-case in mind, such as image reconstruction [24], recognition [37, 43, 52], optical flow [21, 42, 53], driving/SLAM [26, 29, 41]. The dataset perhaps most similar to ours is the *Event-Camera Dataset and Simulator* [28]. All of the above datasets are limited to monochrome temporal contrast or gray-level events. Our Color Event Camera Dataset (CED) doesn't have a particular use-case in mind and aims simply to cover a wide range of scenarios and motions that can be used in a broad swathe of research topics.

The need for publicly available datasets of arbitrary event data is partly driven by the fact that event cameras are scarce and expensive hardware acquisitions. For this reason several event camera simulators have been developed in previous years, the most sophisticated of which is the *ESIM* [1]. While ESIM provides high quality, realistic event data and ground-truth from a free moving simulated camera in an arbitrary 3D modeled environment, it does not support color events. Nor does (to our knowledge) any other contemporary, publicly available event simulator. We propose an extension of ESIM to simulate color events and make it publicly available.

Thus far there have been few works that use color events. One particular counterexample is Marcireau *et al.* [45], who perform color segmentation on color events. However, in this work the authors felt compelled to build their own color event camera using a complex array of beam splitting mirrors and filters to channel light into three separate event cameras. Further, this setup did not allow capturing color frames, which had to be instead reconstructed from the event streams of the three sensors. Our dataset hopes to save future researchers this kind of effort.
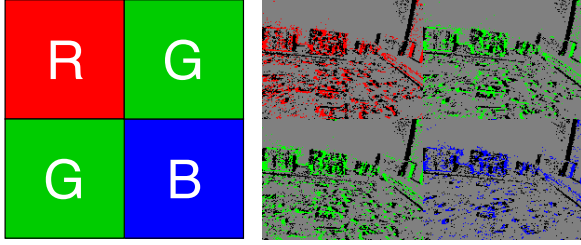
The C-DAVIS [49] was one of the first color event

Figure 3: Left: 2×2 RGBG Bayer pattern in the Color-DAVIS346. Right: Events from the Color-DAVIS346 split into each color. Positive (ON) events are colored by the corresponding filter color, negative (OFF) events are black.

cameras, based on the DAVIS [51] with VGA resolution color (RGBW) frames and QVGA monochrome events. The SDAVIS192 [48] had improved sensitivity over the DAVIS, able to output color (RGBW) events and frames at 188×192 pixel resolution. Moeys *et al*. [47] used the SDAVIS192 to demonstrate color image reconstruction from events using 1) naïve integration and 2) Poisson integration [54] of a gradient field based on the surface of active events [15]. The Color-DAVIS346 [44] is the latest color event camera at the time of writing, and outputs color (RGBG) events and frames at 346×260 resolution.

3. CED: Color Event Camera Dataset

The Color-DAVIS346 [44] consists of an 8×6mm CMOS chip patterned with RGBG filters (Fig. 3), able to output color events and standard frames at 346×260 pixel resolution. Table 1 displays the camera bias settings used (based off the defaults provided in the DAVIS ROS driver[1]). Events generated by the DAVIS are reported with microsecond timestamp precision. We provide time-stamped, raw frames from the DAVIS, as well as color frames obtained via demosaicing [55]. To minimize motion blur in the DAVIS frames, we use fixed exposure fine-tuned for each indoor sequence. We use auto-exposure for outdoor sequences since it is bright enough to drive exposure time down. No infrared filter is used unless otherwise specified. We provide binary (rosbag) files containing synchronized and time-stamped events, raw images and color images.

The Color Event Camera Dataset (Fig. 4) contains 50 minutes of footage consisting of 100k color DAVIS frames and over one billion color events. The sequences cover a wide variety of scenes that showcase some of the key properties of the technology, namely high dynamic range, high temporal resolution and immunity to motion-blur. We include five categories (Table 2): Simple, Indoors, People, Driving and Calibration. Simple contains sequences in favorable conditions, *i.e.* well-lit, moderate camera motions, where the DAVIS frame is typically sharp and well-

Table 1: Bias settings used for the Color-DAVIS346.

Bias	Indoors		Outdoors	
	Coarse	Fine	Coarse	Fine
DiffBn	4	39	4	39
OFFBn	4	0	4	0
ONBn	6	200	6	200
PrBp	2	58	3	0
PrSFBp	1	33	1	33
RefrBp	4	25	4	25

exposed. Indoors contains challenging conditions such as low-light, fast camera motion, as well as natural indoor office scenes. People consists of pre-determined actions such as sitting, waving, dancing with both static and dynamic camera. Driving is filmed through the windshield of a car in sunny conditions and contains a range of environments including highways, tunnels, city and country. Calibration shows a ColorChecker and density step target in various lighting conditions including fluorescent, low-light, outdoors, with and without an infrared filter.

Color Event Simulator. In addition to the real event datasets, we extended the event camera simulator ESIM [1] to allow simulation of color events[2]. Our extension operates on the ground-truth color (RGB) frames generated by the rendering engine, and simulates a color filter array (specifically, an RGBG Bayer pattern, as in the DAVIS346 used for this dataset). The simulated Bayered frames are then processed by the event simulation code in ESIM, thus producing color events in the same way as the DAVIS346. ESIM can readily provide multiple ground truth modalities, such as color frames, depth maps, optical flow maps, camera poses and camera velocities. Our extension is compatible with all the rendering engines already bundled with ESIM, including a photorealistic rendering engine. Figure 5 shows an example of color event data and ground truth modalities simulated by our extension of ESIM.

4. Color Video Reconstruction

Image reconstruction from events serves two primary functions: 1) as a way to visualise events and 2) for use in downstream vision applications *e.g.* object detection.

4.1. Method

We evaluate and compare three state-of-the-art event-based image reconstruction methods on our Color Event Camera Dataset. While these methods were originally designed for monochrome events, we found that with minimal modification all three were able to produce convincing color reconstructions. While "ground-truth" color DAVIS frames were available, only color events were used as input to each method.

---

[1] https://github.com/uzh-rpg/rpg_dvs_ros

[2] https://github.com/uzh-rpg/rpg_esim

Simple

Color Keyboard

Fruits

Wires

People

Air guitar

Jumping

Selfie

Indoors

Kitchen

Office

Foosball

Corridor

Driving

City 1

City 2

Country

Tunnel

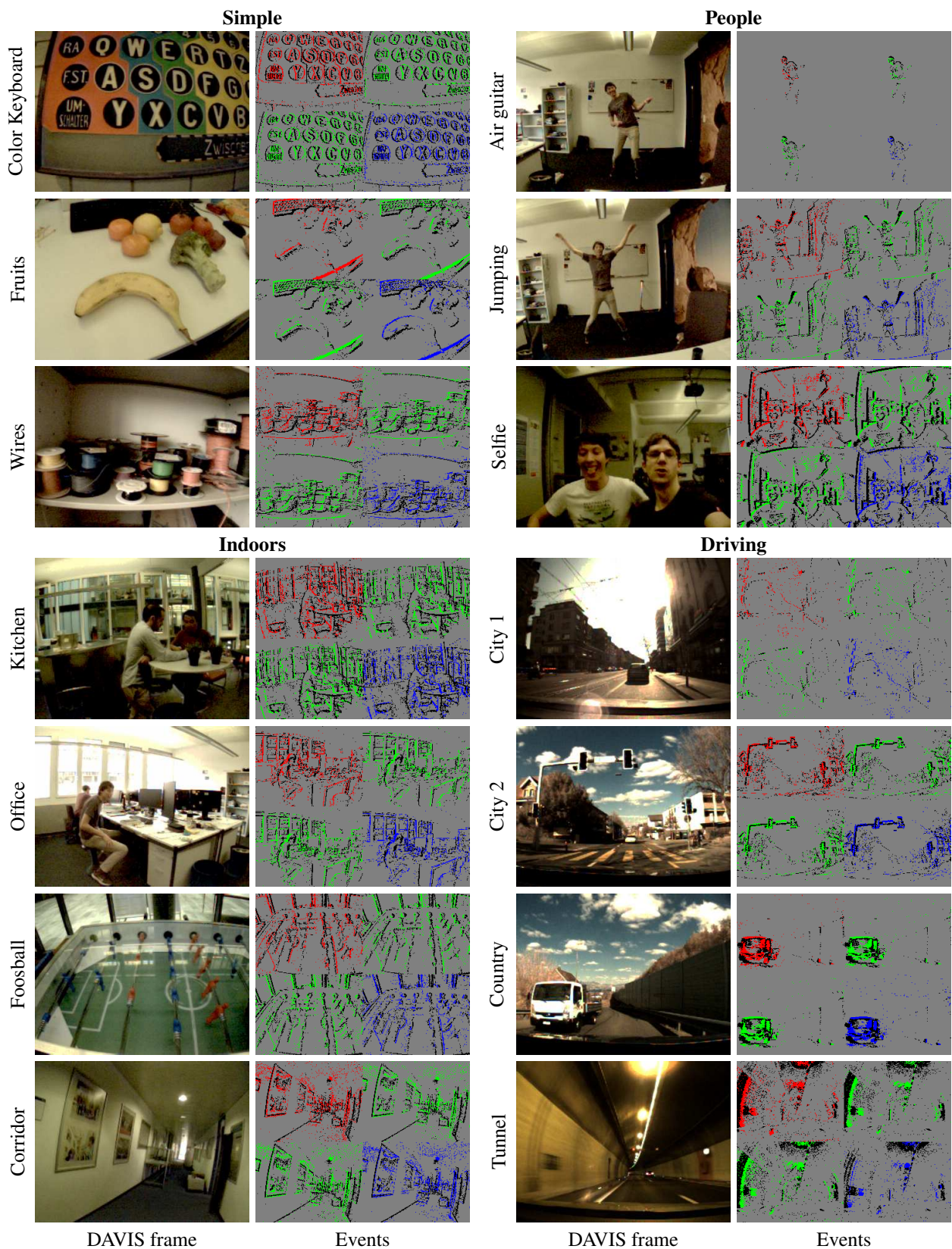DAVIS frame   Events   DAVIS frame   Events

Figure 4: Impressions of the scenes from our dataset. Left column: color DAVIS frame; right column: color events.

Table 2: Types of scenes in our Color Event Camera Dataset.

Type	# Seq	Length (mins)	Lux	Description	Possible Applications
Simple	16	5	80 - 1e3	Simple camera motions looking at simple objects and scenes with vibrant colors such as fruit, blocks and posters.	Image reconstruction
Indoors	15	5	0.8 - 1e3	Natural indoor scenes including office, kitchen, rooms and corridors	Object detection
People	27	10	400	Common actions and gestures such as sitting, waving, jumping, air guitar.	Action recognition
Driving	12	28	200 - 1e5	Footage from front windshield of car driving around country, suburban and city landscapes. Features tunnels, traffic lights, vehicles and pedestrians during the day in sunny conditions.	Segmentation, Optical flow
Calibration	14	2	80 - 1e5	ColorChecker and density step target: indoors, outdoors, with and without infrared filter.	Color calibration
Simulated	-	-	-	Color ESIM (adapted from [1]). Simulator can be used to generate unlimited sequences with ground truth depth, ego-motion, optical flow and more.	Optical flow, SLAM, Image reconstruction

Synthetic Room

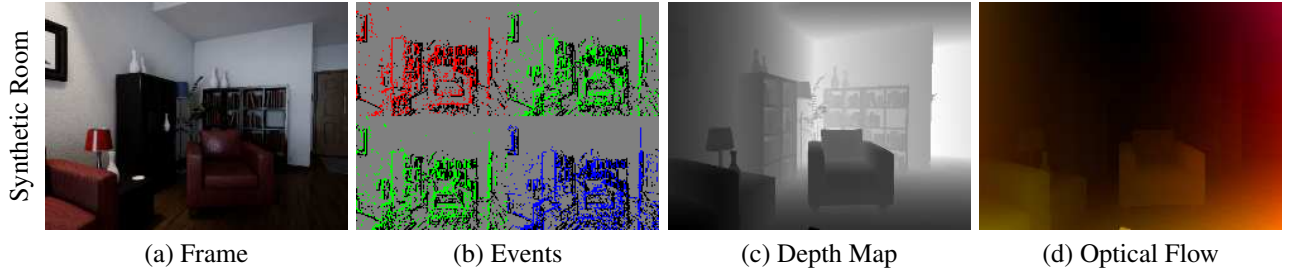(a) Frame   (b) Events   (c) Depth Map   (d) Optical Flow

Figure 5: Example color events and ground truth modalities simulated with our color extension for ESIM. This scene was generated using the photorealistic rendering engine based on Unreal Engine.

1. Manifold Regularisation (MR).[3] Reinbacher *et al.* [22] use integration with spatio-temporal smoothing to recover image frames from events. They use the surface of active events [15] to define a manifold that guides regularisation. We use default parameters provided by the authors; the integration window length is set to to $1,000$ events.

2. High-pass Filter (HF).[4] Scheerlinck *et al.* [24] show that a lightweight, asynchronous complementary filter can be used to obtain a continuous-time video from events and frames. If desired, the frame input to the filter can be set to zero, resulting in a simple high-pass filter that produces reasonable results from only events. Since each pixel is treated independently without spatial smoothing, the Bayer pattern is preserved, and demosaicing [56] can be used to recover an RGB image at any point in time.

We use a gain of 0.06 for both `cutoff_frequency` and `cutoff_frequency_per_event_component`. As a final post-processing step, we apply a 5×5 bilateral filter with `spatial_filter_sigma` set to 1.0 for each output reconstruction.

3. E2VID Neural Network (E2VID). Rebecq *et al.* [50] show that a recurrent neural network trained on a large amount of event data simulated with ESIM [1] can generate high quality video reconstructions from event data only. E2VID converts the stream of events into a sequence of "event tensors", each consisting of a fixed batch of events represented as a 3D spatio-temporal voxel grid. The sequence of event tensors is passed to a recurrent UNet that outputs a sequence of reconstructed image frames.

Manifold regularization (MR) and E2VID utilize spatial smoothing, which destroys the Bayer pattern if applied directly to events. For both of these methods, we found that color images can still be obtained by reconstructing red,

---

[3]https://github.com/VLOGroup/dvs-reconstruction
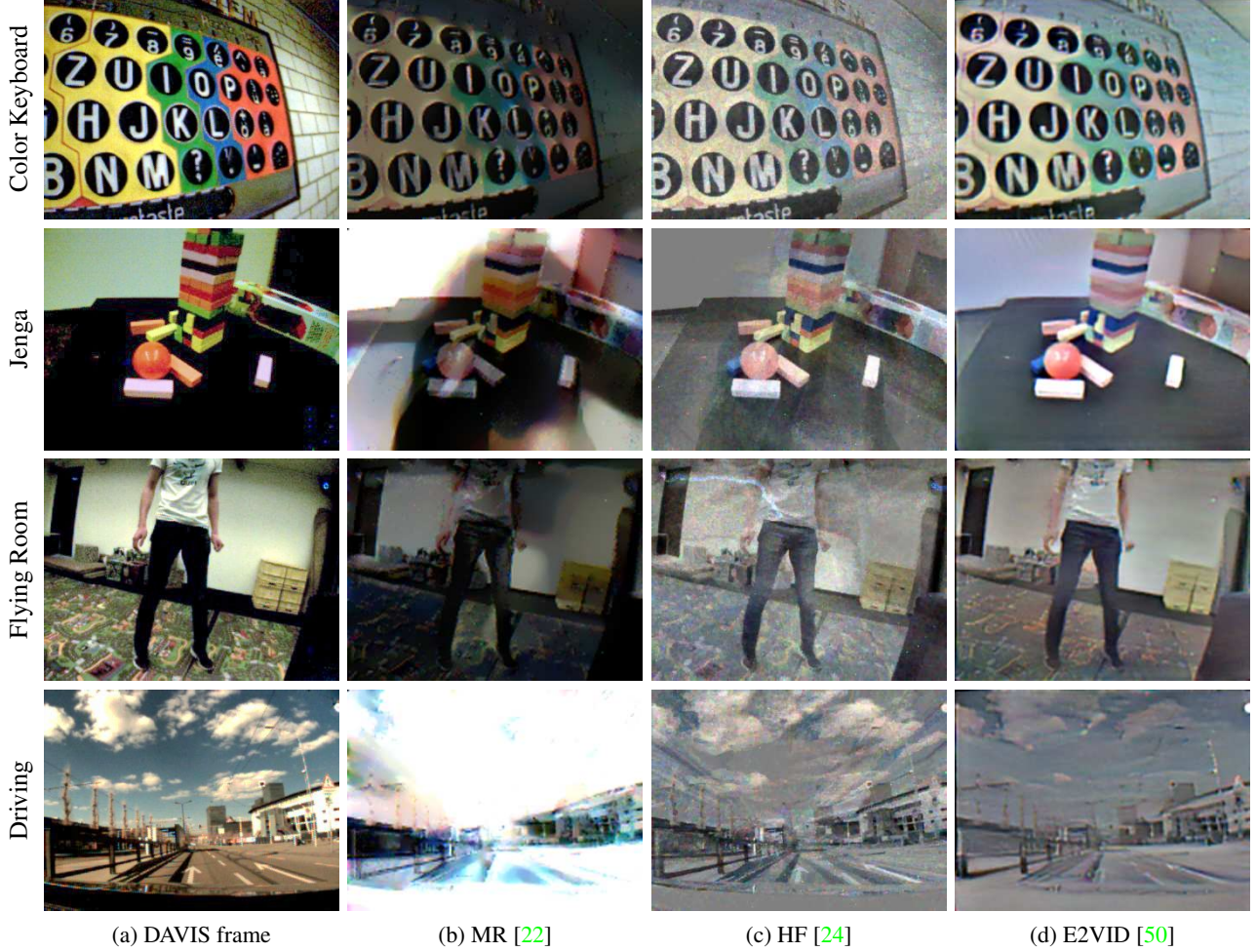[4]https://github.com/cedric-scheerlinck/dvs_image_reconstruction

Figure 6: Qualitative comparison of different color video reconstruction methods on our dataset (images randomly selected). Only events were used for each reconstruction method. Results (c), (d) qualitatively match the DAVIS frame (a).

(a) DAVIS frame   (b) MR [22]   (c) HF [24]   (d) E2VID [50]

green and blue channels independently (at quarter resolution), then upsampling to the original resolution using bicubic interpolation. Because of the Bayer pattern, the four different (upsampled) color channels will not be exactly aligned. Therefore, we shift each color channel by one pixel horizontally and/or vertically so that all four color channels are geometrically aligned. We fuse both green channels (after alignment) by simply taking the mean. In contrast, the High-pass filter (HF) treats each pixel independently and does not perform spatial smoothing. Thus, it can be applied directly to events, then converted to color using demosaicing [56].

4.2. Results

Figure 6 shows reconstruction results of all three methods; Manifold regularisation (MR), High-pass filter (HF) and events-to-video neural network (E2VID), alongside DAVIS frames from the Color-DAVIS346. HF and E2VID preserve color well and qualitatively match the DAVIS frame. We encourage the reader to watch the accompanying video, which convey our results better than still-images.

Figure 7 displays edge cases such as high-speed, HDR *etc.* that highlight strengths and weaknesses of each reconstruction method and the DAVIS frames:

Initialisation (first row). Both MR and HF are initialised at zero and rely on integration of events to build a consistent image over time. Thus, they are prone to producing edge-like images, particularly within the first few milliseconds after initialisation, until enough events 'fill in' the missing information. In contrast, E2VID is good at filling in gaps and can hallucinate color accurately in places with no events.

Fast Motion (second row). HF is a temporal high-pass filter, and is sensitive to temporal components in the input signal, such as frequency and speed. Thus, the quality of the reconstruction can be adversely affected by extremely fast (or slow) motions.

Figure 7: Edge cases for different reconstruction methods. First row: initialization, all method but E2VID fail. Second row: fast motion, HF accumulates more noise. Third row: zoom on carpet, HF preserves fine details better. Fourth row: Low apparent motion *e.g.* in the sky, HF preserves slow moving objects better. Fifth row: HDR scene, DAVIS cannot capture entire intensity range, reconstructions can. Sixth row: dark room (2 lux), DAVIS suffers motion blur, not the reconstructions.

In addition, fast motions tend to generate noise in the event stream that is accumulated without discrimination by the integrator in HF. MR and E2VID are good at rejecting noise from fast motion and showcase the attractive properties of event cameras for challenging scenarios.

Sharpness (third row). MR and E2VID rely on spatial smoothing to filter out noise from the event stream, which can degrade sharpness of fine details. For color reconstruction, the spatial smoothing property of these two methods destroys the Bayer pattern, requiring each color to be reconstructed independently (at quarter resolution), then upscaled back to the original resolution, further losing fine details. In contrast, HF requires no spatial smoothing, so a raw intensity reconstruction at full resolution is possible, since the Bayer pattern is preserved. A demosaicing algorithm [56] can be used to convert the raw output to color without loss of resolution, resulting in a sharper reconstruction.

Memory (fourth row). The "memory" (*i.e.* the time span over which information in the event data can be propagated) is variable between all three methods. For HF, the size of the temporal receptive field (memory) is explicitly encoded through the cutoff frequency parameter. Hence, the duration across which information can be propagated can be set to an arbitrarily high amount of time, at the expense of integrating more noise, and creating "bleeding" patterns following moving objects. By contrast, MR and E2VID have an implicit memory, whose size can vary with the number of events used in each integration window (MR), or event tensor (E2VID). However, we observe that the memory of MR and HF is notably smaller than HF, which is particularly visible in the driving sequence (fourth row of Fig. 7), where HF is able to reconstruct slow moving objects, *e.g.* the clouds or the distant buildings, in contrast to MR and E2VID.

HDR (fifth row). Since the APS is limited to a uniform exposure duration for all pixels, the DAVIS frame has low dynamic range compared to events. Thus, dark regions are often underexposed while bright regions (window) are well exposed, and vice versa. Reconstructions from MR, HF and E2VID all showcase the high dynamic range property of events, *i.e.* both dark and bright regions are clear.

Low light (sixth row). Low lighting is a challenge for conventional cameras because the exposure duration must be increased to avoid underexposure, leading to motion blur. While the DAVIS frame is motion blurred, MR, HF and E2VID demonstrate immunity to motion blur, even in challenging low lighting conditions.

4.3. Application of Reconstructions

While many computer vision algorithms work on grayscale images, it is well established that incorporating color information can significantly boost performance for the task at hand [57]. This is because color images contain

Figure 8: Object detection (YOLO [58]) on reconstructed images using E2VID. Color (right) tends to improve detection performance.

more information about the scene than grayscale images, which can only encode structural information. This is particularly true in recognition tasks, where color can be an important visual cue. Figure 8 shows one example where color improves object detection performance. We apply YOLO [58] to E2VID images reconstructed from both grayscale and color events and observe that color offers qualitative improvement. While image reconstructions can be used directly for the task at hand, they may also be used to generate proxy labels (*e.g.* segmentation, optical flow, recognition) that can be transferred to events.

5. Conclusion

We present the first *Color Event Camera Dataset*, containing both frames and events across a diverse range of scenes, motions and lighting conditions. We release an open source color event camera simulator based on ESIM [1]. We show how three state-of-the-art event-based image reconstruction methods can be adapted for color video reconstruction, and compare strengths/weaknesses of each method. We hope that our Color Event Camera Dataset and simulator will inspire future work with color events, which we believe is the next step for event-based vision.

Acknowledgements

References

[1] H. Rebecq, D. Gehrig, and D. Scaramuzza, "ESIM: an open event camera simulator," in *Conf. on Robotics Learning (CoRL)*, 2018. 1, 2, 3, 5, 8

[2] P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128×128 120 dB 15 μs latency asynchronous temporal contrast vision sensor," *IEEE J. Solid-State Circuits*, vol. 43, no. 2, pp. 566–576, 2008. 1, 2

[3] D. Gehrig, H. Rebecq, G. Gallego, and D. Scaramuzza, "Asynchronous, photometric feature tracking using events and frames," in *Eur. Conf. Comput. Vis. (ECCV)*, 2018. 1

[4] F. Barranco, C. Fermuller, and E. Ros, "Real-time clustering and multi-target tracking using event-based sensors," in *IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*, 2018. 1

[5] I. Alzugaray and M. Chli, "Asynchronous corner detection and tracking for event cameras in real time," *IEEE Robot. Autom. Lett.*, vol. 3, pp. 3177–3184, Oct. 2018. 1

[6] C. Scheerlinck, N. Barnes, and R. Mahony, "Asynchronous spatial image convolutions for event cameras," *IEEE Robot. Autom. Lett.*, vol. 4, pp. 816–822, Apr. 2019. 2

[7] H. Rebecq, G. Gallego, and D. Scaramuzza, "EMVS: Event-based multi-view stereo," in *British Mach. Vis. Conf. (BMVC)*, 2016. 1

[8] H. Rebecq, G. Gallego, E. Mueggler, and D. Scaramuzza, "EMVS: Event-based multi-view stereo—3D reconstruction with an event camera in real-time," *Int. J. Comput. Vis.*, vol. 126, pp. 1394–1414, Dec. 2018. 1

[9] H. Kim, S. Leutenegger, and A. J. Davison, "Real-time 3D reconstruction and 6-DoF tracking with an event camera," in *Eur. Conf. Comput. Vis. (ECCV)*, pp. 349–364, 2016. 1

[10] G. Gallego, H. Rebecq, and D. Scaramuzza, "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.

[11] D. Zou, F. Shi, W. Liu, J. Li, Q. Wang, P.-K. J. Park, C.-W. Shi, Y. J. Roh, and H. E. Ryu, "Robust dense depth map estimation from sparse DVS stereos," in *British Mach. Vis. Conf. (BMVC)*, 2017. 1

[12] A. Andreopoulos, H. J. Kashyap, T. K. Nayak, A. Amir, and M. D. Flickner, "A low power, high throughput, fully event-based stereo system," in *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, pp. 7532–7542, 2018. 1

[13] A. Z. Zhu, Y. Chen, and K. Daniilidis, "Realtime time synchronized event-based stereo," in *Eur. Conf. Comput. Vis. (ECCV)*, 2018. 1

[14] Y. Zhou, G. Gallego, H. Rebecq, L. Kneip, H. Li, and D. Scaramuzza, "Semi-dense 3D reconstruction with a stereo event camera," in *Eur. Conf. Comput. Vis. (ECCV)*, 2018.

[15] R. Benosman, C. Clercq, X. Lagorce, S.-H. Ieng, and C. Bartolozzi, "Event-based visual flow," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 2, pp. 407–417, 2014. 1, 3, 5

[16] E. Mueggler, C. Forster, N. Baumli, G. Gallego, and D. Scaramuzza, "Lifetime estimation of events from dynamic vision sensors," in *IEEE Int. Conf. Robot. Autom. (ICRA)*, pp. 4874–4881, 2015. 1

[17] T. Stoffregen and L. Kleeman, "Simultaneous optical flow and segmentation (SOFAS) using Dynamic Vision Sensor," in *Australasian Conf. Robot. Autom. (ACRA)*, 2017. 1

[18] M. Liu and T. Delbruck, "Adaptive time-slice block-matching optical flow algorithm for dynamic vision sensors," in *British Mach. Vis. Conf. (BMVC)*, 2018.

[19] H. Kim, A. Handa, R. Benosman, S.-H. Ieng, and A. J. Davison, "Simultaneous mosaicing and tracking with an event camera," in *British Mach. Vis. Conf. (BMVC)*, 2014. 1, 2

[20] S. Barua, Y. Miyatani, and A. Veeraraghavan, "Direct face detection and video reconstruction from event cameras," in *IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, pp. 1–9, 2016. 1, 2

[21] P. Bardow, A. J. Davison, and S. Leutenegger, "Simultaneous optical flow and intensity estimation from an event camera," in *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, pp. 884–892, 2016. 1, 2

[22] C. Reinbacher, G. Graber, and T. Pock, "Real-time intensity-image reconstruction for event cameras using manifold regularisation," in *British Mach. Vis. Conf. (BMVC)*, 2016. 1, 2, 5, 6, 7

[23] G. Munda, C. Reinbacher, and T. Pock, "Real-time intensity-image reconstruction for event cameras using manifold regularisation," *Int. J. Comput. Vis.*, vol. 126, pp. 1381–1393, July 2018. 1, 2

[24] C. Scheerlinck, N. Barnes, and R. Mahony, "Continuous-time intensity estimation using event cameras," in *Asian Conf. Comput. Vis. (ACCV)*, 2018. 1, 2, 5, 6, 7

[25] L. Pan, C. Scheerlinck, X. Yu, R. Hartley, M. Liu, and Y. Dai, "Bringing a blurry frame alive at high frame-rate with an event camera," in *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2019.

[26] G. Gallego, J. E. A. Lund, E. Mueggler, H. Rebecq, T. Delbruck, and D. Scaramuzza, "Event-based, 6-DOF camera tracking from photometric depth maps," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, pp. 2402–2412, Oct. 2018. 1

[27] C. Reinbacher, G. Munda, and T. Pock, "Real-time panoramic tracking for event cameras," in *IEEE Int. Conf. Comput. Photography (ICCP)*, pp. 1–9, 2017. 1

[28] E. Mueggler, H. Rebecq, G. Gallego, T. Delbruck, and D. Scaramuzza, "The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and SLAM," *Int. J. Robot. Research*, vol. 36, no. 2, pp. 142–149, 2017. 1, 2

[29] S. Bryner, G. Gallego, H. Rebecq, and D. Scaramuzza, "Event-based, direct camera tracking from a photometric 3D map using nonlinear optimization," in *IEEE Int. Conf. Robot. Autom. (ICRA)*, 2019. 2

[30] B. Kueng, E. Mueggler, G. Gallego, and D. Scaramuzza, "Low-latency visual odometry using event-based feature tracks," in *IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*, pp. 16–23, 2016. 1

[31] H. Rebecq, T. Horstschäfer, G. Gallego, and D. Scaramuzza, "EVO: A geometric approach to event-based 6-DOF parallel tracking and mapping in real-time," *IEEE Robot. Autom. Lett.*, vol. 2, no. 2, pp. 593–600, 2017. 1

[32] G. Gallego and D. Scaramuzza, "Accurate angular velocity estimation with an event camera," *IEEE Robot. Autom. Lett.*, vol. 2, no. 2, pp. 632–639, 2017.

[33] E. Mueggler, G. Gallego, H. Rebecq, and D. Scaramuzza, "Continuous-time visual-inertial odometry for event cameras," *IEEE Trans. Robot.*, 2018. 1

[34] A. Z. Zhu, N. Atanasov, and K. Daniilidis, "Event-based visual inertial odometry," in *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, pp. 5816–5824, 2017. 1

[35] H. Rebecq, T. Horstschaefer, and D. Scaramuzza, "Real-time visual-inertial odometry for event cameras using keyframe-based nonlinear optimization," in *British Mach. Vis. Conf. (BMVC)*, 2017. 1

[36] A. Rosinol Vidal, H. Rebecq, T. Horstschaefer, and D. Scaramuzza, "Ultimate SLAM? combining events, images, and IMU for robust visual SLAM in HDR and high speed scenarios," *IEEE Robot. Autom. Lett.*, vol. 3, pp. 994–1001, Apr. 2018.

[37] A. Sironi, M. Brambilla, N. Bourdis, X. Lagorce, and R. Benosman, "HATS: Histograms of averaged time surfaces for robust event-based object classification," in *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, pp. 1731–1740, 2018. 1, 2

[38] A. I. Maqueda, A. Loquercio, G. Gallego, N. García, and D. Scaramuzza, "Event-based vision meets deep learning on steering prediction for self-driving cars," in *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, pp. 5419–5427, 2018. 1

[39] A. Z. Zhu, L. Yuan, K. Chaney, and K. Daniilidis, "EV-FlowNet: Self-supervised optical flow estimation for event-based cameras," in *Robotics: Science and Systems (RSS)*, 2018. 1

[40] S. B. Shrestha and G. Orchard, "SLAYER: Spike layer error reassignment in time," in *Conf. Neural Inf. Process. Syst. (NIPS)*, Dec. 2018.

[41] A. Z. Zhu, D. Thakur, T. Ozaslan, B. Pfrommer, V. Kumar, and K. Daniilidis, "The multivehicle stereo event camera dataset: An event camera dataset for 3D perception," *IEEE Robot. Autom. Lett.*, vol. 3, pp. 2032–2039, July 2018. 1, 2

[42] B. Rueckauer and T. Delbruck, "Evaluation of event-based algorithms for optical flow with ground-truth from inertial measurement sensor," *Front. Neurosci.*, vol. 10, no. 176, 2016. 1, 2

[43] G. Orchard, A. Jayawant, G. K. Cohen, and N. Thakor, "Converting static image datasets to spiking neuromorphic datasets using saccades," *Front. Neurosci.*, vol. 9, p. 437, 2015. 1, 2

[44] G. Taverni, D. P. Moeys, C. Li, C. Cavaco, V. Motsnyi, D. S. S. Bello, and T. Delbruck, "Front and back illuminated Dynamic and Active Pixel Vision Sensors comparison," *IEEE Trans. Circuits Syst. II*, vol. 65, no. 5, pp. 677–681, 2018. 2, 3

[45] A. Marcireau, S.-H. Ieng, C. Simon-Chane, and R. B. Benosman, "Event-based color segmentation with a high dynamic range sensor," *Front. Neurosci.*, vol. 12, 2018. 2

[46] A. Trémeau, S. Tominaga, and K. N. Plataniotis, "Color in image and video processing: Most recent trends and future research directions," *EURASIP Journal on Image and Video Processing*, vol. 2008, pp. 1–26, 2008. 2

[47] D. P. Moeys, C. Li, J. N. P. Martel, S. Bamford, L. Longinotti, V. Motsnyi, D. S. S. Bello, and T. Delbruck, "Color temporal contrast sensitivity in dynamic vision sensors," in *IEEE Int. Symp. Circuits Syst. (ISCAS)*, pp. 1–4, 2017. 2, 3

[48] D. P. Moeys, F. Corradi, C. Li, S. A. Bamford, L. Longinotti, F. F. Voigt, S. Berry, G. Taverni, F. Helmchen, and T. Delbruck, "A sensitive dynamic and active pixel vision sensor for color or neural imaging applications," *IEEE Trans. Biomed. Circuits Syst.*, vol. 12, pp. 123–136, Feb. 2018. 2, 3

[49] C. Li, C. Brandli, R. Berner, H. Liu, M. Yang, S.-C. Liu, and T. Delbruck, "Design of an RGBW color VGA rolling and global shutter dynamic and active-pixel vision sensor," in *IEEE Int. Symp. Circuits Syst. (ISCAS)*, 2015. 2

[50] H. Rebecq, R. Ranftl, V. Koltun, and D. Scaramuzza, "Events-to-video: Bringing modern computer vision to event cameras," in *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2019. 2, 5, 6, 7

[51] C. Brandli, R. Berner, M. Yang, S.-C. Liu, and T. Delbruck, "A 240x180 130dB 3us latency global shutter spatiotemporal vision sensor," *IEEE J. Solid-State Circuits*, vol. 49, no. 10, pp. 2333–2341, 2014. 2

[52] T. Serrano-Gotarredona and B. Linares-Barranco, "Poker-DVS and MNIST-DVS. their history, how they were made, and other details," *Front. Neurosci.*, vol. 9, Dec. 2015.

[53] A. Mitrokhin, C. Fermuller, C. Parameshwara, and Y. Aloimonos, "Event-based moving object detection and tracking," in *IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*, 2018. 2

[54] A. Agrawal, R. Chellappa, and R. Raskar, "An algebraic approach to surface reconstruction from gradient fields," in *Int. Conf. Comput. Vis. (ICCV)*, pp. 174–181, 2005. 3

[55] "Opencv color conversions." https://docs.opencv.org/3.1.0/de/d25/imgproc_color_conversions.html. Accessed: 2019-03-27. 3

[56] R. Kimmel, "Demosaicing: Image reconstruction from color CCD samples," *IEEE Trans. Image Process.*, vol. 8, pp. 1221–1228, Sept. 1999. 5, 6, 8

[57] T. Gevers, A. Gijsenij, J. van de Weijer, and J.-M. Geusebroek, *Color in Computer Vision: Fundamentals and Applications*. Wiley Publishing, 1st ed., 2012. 8

[58] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2016. 8