Tutorial on Event-based Cameras:

Davide Scaramuzza

http://rpg.ifi.uzh.ch/
Event-based Vision: A Survey

Guillermo Gallego, Tobi Delbrück, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew Davison, Jörg Conradt, Kostas Daniilidis, Davide Scaramuzza

Abstract—Event cameras are bio-inspired sensors that work radically different from traditional cameras. Instead of capturing images at a fixed rate, they measure per-pixel brightness changes asynchronously. This results in a stream of events, which encode the time, location and sign of the brightness changes. Event cameras possess outstanding properties compared to traditional cameras: very high dynamic range (140 dB vs. 60 dB), high temporal resolution (in the order of μs), low power consumption, and do not suffer from motion blur. Hence, event cameras have a large potential for robotics and computer vision in challenging scenarios for traditional cameras, such as high speed and high dynamic range. However, novel methods are required to process the unconventional output of these sensors in order to unlock their potential. This paper provides a comprehensive overview of the emerging field of event-based vision, with a focus on the applications and the algorithms developed to unlock the outstanding properties of event cameras. We present event cameras from their working principle, the actual sensors that are available and the tasks that they have been used for, from low-level vision (feature detection and tracking, optic flow, etc.) to high-level vision (reconstruction, segmentation, recognition). We also discuss the techniques developed to process events, including learning-based techniques, as well as specialized processors for these novel sensors, such as spiking neural networks. Additionally, we highlight the challenges that remain to be tackled and the opportunities that lie ahead in the search for a more efficient, bio-inspired way for machines to perceive and interact with the world.


1 INTRODUCTION AND APPLICATIONS

"The brain is imagination, and that was exciting to me; I wanted to build a chip that could imagine something," that is how Misha Mahowald, a graduate student at Caltech in 1986, started to work with Prof. Carver Mead on the stereo problem from a joint biological and engineering pers-as well as new computer vision and robotic tasks. Sight is, by far, the dominant sense in humans to perceive the world, and, together with the brain, learn new things. In recent years, this technology has attracted a lot of attention from both academia and industry. This is due to the availability of prototype event cameras and the advantages that these devices offer to tackle problems that are currently unfeasible.
Open Challenges in Computer Vision

The past 60 years of research have been devoted to frame-based cameras ...but they are not good enough!

Latency & Motion blur  Dynamic Range

Event cameras do not suffer from these problems!
What is an event camera?

- Novel sensor that measures only **motion in the scene**
- **First commercialized in 2008** by T. Delbruck (UZH&ETH) under the name of Dynamic Vision Sensor (DVS)
- **Low-latency** (~ 1 μs)
- No motion blur
- **High dynamic range** (140 dB instead of 60 dB)
- **Ultra-low power** (mean: 1mW vs 1W)

Traditional vision algorithms cannot be used because:
- **Asynchronous** pixels
- **No intensity information** (only binary intensity changes)

Video from here: [https://youtu.be/LauQ6LWTkxM?t=30](https://youtu.be/LauQ6LWTkxM?t=30)

Camera vs Event Camera

- A traditional camera outputs frames at fixed time intervals:

```
frame  |  next frame
0      |  \Delta
```

- By contrast, a DVS outputs asynchronous events at microsecond resolution. An event is generated each time a single pixel detects an intensity changes value.

```
events stream
0      |  \Delta
event: \left[ t, (x, y), \text{sign} \left( \frac{dI(x, y)}{dt} \right) \right]
```

- Event polarity (or sign) (-1 or 1): increase or decrease of brightness

[Lichtsteiner, Posch, Delbruck, A 128x128 120 dB 15\mu s Latency Asynchronous Temporal Contrast Vision Sensor, 2008]
Generative Event Model

Consider the intensity at a single pixel...

\[ \pm C = \log I(x, t) - \log I(x, t - \Delta t) \]

Events are triggered asynchronously
Event cameras are inspired by the Human Eye

**Human retina:**
- 130 million **photoreceptors**
- But only 2 million **axons**!
Event Camera Output with No Motion

Without motion, only background noise is output

Standard Camera

Event Camera (ON, OFF events)

\[ \Delta T = 40 \text{ ms} \]
Event Camera Output with Relative Motion

Standard Camera

Event Camera (ON, OFF events)

$\Delta T = 10 \text{ ms}$
Event Camera Output with Relative Motion

Standard Camera

Event Camera (ON, OFF events)

$\Delta T = 40 \text{ ms}$
Examples
Low-light Sensitivity (night drive)

GoPro Hero 6

Event Camera by Prophesee
White = Positive events
Black = Negative events

Video courtesy of Prophesee: https://www.prophesee.ai
## High-speed vs Event Cameras

<table>
<thead>
<tr>
<th></th>
<th>High speed camera</th>
<th>Standard camera</th>
<th>Event Camera</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max fps or measurement rate</td>
<td>Up to 1MHz</td>
<td>100-1,000 fps</td>
<td>1MHz</td>
</tr>
<tr>
<td>Resolution at max fps</td>
<td>64x16 pixels</td>
<td>&gt;1Mpxl</td>
<td>&gt;1Mpxl</td>
</tr>
<tr>
<td>Bits per pixels (event)</td>
<td>12 bits</td>
<td>8-10 per pixel</td>
<td>~40 bits/event (t,(x,y),p)}</td>
</tr>
<tr>
<td>Weight</td>
<td>6.2 Kg</td>
<td>30 g</td>
<td>30 g</td>
</tr>
<tr>
<td>Active cooling</td>
<td>yes</td>
<td>No cooling</td>
<td>No cooling</td>
</tr>
<tr>
<td>Data rate</td>
<td>1.5 GB/s</td>
<td>32MB/s</td>
<td>~1MB/s on average (depends on dynamics)</td>
</tr>
<tr>
<td>Mean power consumption</td>
<td>150 W + external light</td>
<td>1 W</td>
<td>1 mW</td>
</tr>
<tr>
<td>Dynamic range</td>
<td>n.a.</td>
<td>60 dB</td>
<td>140 dB</td>
</tr>
</tbody>
</table>
Current commercial applications

➢ Internet of Things (IoT)
  • Low-power, always-on devices for monitoring and surveillance

➢ Automotive:
  • low-latency, high dynamic range (HDR) object detection
  • low-power training & inference
  • low-memory storage

➢ AR/VR
  • low-latency, low-power tracking

➢ Industrial automation
  • Fast pick and place
Who sells event cameras and how much are they?

- **Inivation**:
  - DAVIS sensor: frames, events, IMU.
  - Resolution: ~QVGA (346x260 pixels)
  - **Cost: 6,000 USD**

- **Insightness**:
  - RINO sensor: frames, events, IMU.
  - Resolution: ~QVGA (320x262 pixels)
  - **Cost: 6,000 USD**

- **Prophesee**:
  - ATIS sensor: events, IMU, absolute intensity at the event pixel
  - Resolution: 1M pixels
  - **Cost: 4,000 USD**

- **CelexPixel Technology**:
  - Celex One: events, IMU, absolute intensity at the event pixel
  - Resolution: 1M pixels
  - **Cost: 1,000 USD**

- **Samsung Electronics**
  - Samsung DVS: events, IMU
  - Resolution: up to 1Mpxl
  - **Cost: not listed**
Comparison of current event cameras

<table>
<thead>
<tr>
<th>Supplier</th>
<th>Camera model</th>
<th>DVS128</th>
<th>iniVision DAVIS240</th>
<th>DAVIS346</th>
<th>Propsee ATIS</th>
<th>Gen3 CD</th>
<th>Gen3 ATIS</th>
<th>Gen 4 CD</th>
<th>Samsung DVS-Gen2</th>
<th>DVS-Gen3</th>
<th>DVS-Gen4</th>
<th>CelePix</th>
<th>CeleXi-IV</th>
<th>CeleX-V</th>
<th>Insighites Rino 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution (pixels)</td>
<td></td>
<td>128×128</td>
<td>240×180</td>
<td>346×260</td>
<td>304×240</td>
<td>640×480</td>
<td>480×360</td>
<td>1280×720</td>
<td>640×480</td>
<td>150</td>
<td>8</td>
<td>150</td>
<td>120</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Latency (μs)</td>
<td></td>
<td>12μs @ 1kHz</td>
<td>12μs @ 1kHz</td>
<td>20</td>
<td>70</td>
<td>40-200</td>
<td>40-200</td>
<td>20-150</td>
<td>65-410</td>
<td>50</td>
<td>0</td>
<td>50</td>
<td>0</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Dynamic range (dB)</td>
<td></td>
<td>120</td>
<td>120</td>
<td>120</td>
<td>143</td>
<td>&gt; 120</td>
<td>&gt; 120</td>
<td>&gt; 124</td>
<td>90</td>
<td>90</td>
<td>90</td>
<td>90</td>
<td>15</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Min. contrast sensitivity (%)</td>
<td></td>
<td>17</td>
<td>11</td>
<td>14.3-22.5</td>
<td>13</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>9</td>
<td>15</td>
<td>10</td>
<td>-</td>
<td>15</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Power consumption (mW)</td>
<td></td>
<td>23</td>
<td>11-14</td>
<td>10-120</td>
<td>50-175</td>
<td>36-95</td>
<td>25-87</td>
<td>32-84</td>
<td>27-50</td>
<td>40</td>
<td>130</td>
<td>-</td>
<td>400</td>
<td>20-70</td>
<td></td>
</tr>
<tr>
<td>Chip size (mm²)</td>
<td></td>
<td>6.3×6</td>
<td>5×5</td>
<td>5×6</td>
<td>9.9×8.2</td>
<td>9.6×7.2</td>
<td>9.6×7.2</td>
<td>6.2×3.5</td>
<td>8×5.8</td>
<td>8×5.8</td>
<td>8.4×7.6</td>
<td>4.95×4.95</td>
<td>22</td>
<td>15.5×15.8</td>
<td>14.3×11.6</td>
</tr>
<tr>
<td>Pixel size (μm²)</td>
<td></td>
<td>40×40</td>
<td>18.5×18.5</td>
<td>18.5×18.5</td>
<td>30×30</td>
<td>15×15</td>
<td>20×20</td>
<td>4.86×4.86</td>
<td>11×9</td>
<td>12×9</td>
<td>9×9</td>
<td>22</td>
<td>18</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Fill factor (%)</td>
<td></td>
<td>8.1</td>
<td>22</td>
<td>22</td>
<td>20</td>
<td>25</td>
<td>20</td>
<td>&gt;77</td>
<td>12×12</td>
<td>12×12</td>
<td>12×12</td>
<td>12×12</td>
<td>12×12</td>
<td>12×12</td>
<td></td>
</tr>
<tr>
<td>Supply voltage (V)</td>
<td></td>
<td>3.3</td>
<td>1.8 &amp; 3.3</td>
<td>1.8 &amp; 3.3</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
<td>1.1 &amp; 2.5</td>
<td>1.2 &amp; 2.8</td>
<td>1.2 &amp; 2.8</td>
<td>1.2 &amp; 2.8</td>
<td>1.2 &amp; 2.8</td>
<td>1.2 &amp; 2.8</td>
<td>1.2 &amp; 2.8</td>
<td></td>
</tr>
<tr>
<td>Stationary noise (ev/pix/s) at 25°C</td>
<td></td>
<td>0.05</td>
<td>0.1</td>
<td>0.1</td>
<td>-</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.03</td>
<td>0.03</td>
<td>0.1</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>CMOS technology (μm)</td>
<td></td>
<td>2P4M</td>
<td>1P6M MIM</td>
<td>1P6M MIM</td>
<td>1P6M MIM</td>
<td>1P6M CIS</td>
<td>1P6M CIS</td>
<td>1P6M CIS</td>
<td>1P6M BSI</td>
<td>1P6M BSI</td>
<td>1P6M CIS</td>
<td>1P6M CIS</td>
<td>1P6M CIS</td>
<td>1P6M CIS</td>
<td></td>
</tr>
<tr>
<td>Max. Bandwidth (Meps)</td>
<td></td>
<td>USB 2</td>
<td>USB 2</td>
<td>USB 3</td>
<td>-</td>
<td>USB 2</td>
<td>USB 3</td>
<td>USB 2</td>
<td>USB 3</td>
<td>USB 2</td>
<td>USB 3</td>
<td>USB 3</td>
<td>USB 2</td>
<td>USB 2</td>
<td></td>
</tr>
<tr>
<td>IMU output</td>
<td></td>
<td>no</td>
<td>1kHz</td>
<td>1kHz</td>
<td>no</td>
<td>1kHz</td>
<td>1kHz</td>
<td>no</td>
<td>1kHz</td>
<td>no</td>
<td>1kHz</td>
<td>no</td>
<td>1kHz</td>
<td>1kHz</td>
<td></td>
</tr>
</tbody>
</table>

Table from [Guillermo et al., T-PAMI’20], Table 1


How do we unlock the outstanding potential of event cameras:

• Low latency
• High dynamic range
• No motion blur
Recall the Generative Event Model

An event is triggered at a single pixel if

\[ \log I(x, t) - \log I(x, t - \Delta t) = \pm C \]
1st Order Approximation

• Let us define $L(x, y, t) = \log(I(x, y, t))$

• Consider a given pixel $p(x, y)$ with gradient $\nabla L(x, y)$ undergoing the motion $u = (u, v)$ in pixels, induced by a moving 3D point $P$.

• Then, it can be shown that:

$$-\nabla L \cdot u = C$$
Proof

The proof comes from the **brightness constancy assumption**, which says that the intensity value of \( p \), before and after the motion, must remain unchanged:

\[
L(x, y, t) = L(x + u, y + v, t + \Delta t)
\]

By replacing the right-hand term by its 1\(^{\text{st}}\) order approximation at \( t + \Delta t \), we get:

\[
L(x, y, t) = L(x, y, t + \Delta t) + \frac{\partial L}{\partial x} u + \frac{\partial L}{\partial y} v
\]

\[
\Rightarrow L(x, y, t + \Delta t) - L(x, y, t) = -\frac{\partial L}{\partial x} u - \frac{\partial L}{\partial y} v
\]

\[
\Rightarrow \Delta L = C = -\nabla L \cdot \mathbf{u}
\]

This equation describes the **linearized** event generation equation for an event generated by a gradient \( \nabla L \) that moved by a motion vector \( \mathbf{u} \) (optical flow) during a time interval \( \Delta t \).
Example 1: Image Reconstruction from events

➢ Probabilistic simultaneous, gradient & rotation estimation from $C = -\nabla L \cdot u$
➢ Obtain intensity from gradients via Poisson reconstruction
➢ The reconstructed image has super-resolution and high dynamic range (HDR)
➢ In real time on a GPU

Kim et al., Simultaneous Mosaicing and Tracking with an Event Camera, BMVC’14
Example 2: 6DoF Tracking from Photometric Map

- Probabilistic, simultaneous motion & contrast estimation from $C = -\nabla L \cdot u$
- Assumes photometric map (x,y,z, grayscale Intensity) is given
- Useful for VR/AR applications (low-latency, HDR, no motion blur)
- Requires GPU to run in real time

Gallego et al., Event-based 6-DOF Camera Tracking from Photometric Depth Maps, T-PAMI, 2018. [PDF Video]
Example 2: 6DoF Tracking from Photometric Map

estimated pose  ground truth

raw events: ON / OFF

events on projected map

Bryner et al., Event-based, Direct Camera Tracking from a Photometric Depth Map, ICRA'19.  PDF  Video
Event camera

Motion estimation

Gallego et al., Event-based 6-DOF Camera Tracking from Photometric Depth Maps, T-PAMI’18. PDF Video
Example 3: Parallel Tracking & Mapping (SLAM)

- Tracking: EKF in 6 DOF pose
  - Uses random walk model & inverse depth
  - Use 1st order approximation of generative event model to update pose
- Runs in real time on a GPU

Kim et al., Real-Time 3D Reconstruction and 6-DoF Tracking with an Event Camera, ECCV’16
What if we combined the complementary advantages of event and standard cameras?
Why combining them?

<table>
<thead>
<tr>
<th></th>
<th>Event Camera</th>
<th>Standard Camera</th>
</tr>
</thead>
<tbody>
<tr>
<td>Update rate</td>
<td>High (asynchronous): 1 MHz</td>
<td>Low (synchronous)</td>
</tr>
<tr>
<td>Dynamic Range</td>
<td>High (140 dB)</td>
<td>Low (60 dB)</td>
</tr>
<tr>
<td>Motion Blur</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Static motion</td>
<td>No (event camera is a high pass filter)</td>
<td>Yes</td>
</tr>
<tr>
<td>Absolute intensity</td>
<td>No (reconstructable up to a constant)</td>
<td>Yes</td>
</tr>
<tr>
<td>Maturity</td>
<td>&lt; 10 years of research</td>
<td>&gt; 60 years of research!</td>
</tr>
</tbody>
</table>
DAVIS sensor: Events + Images + IMU

➢ Combines an **event and a standard camera** in the **same pixel array** (→ the same pixel can both trigger events and integrate light intensity).

➢ **It also has an IMU**

Brandli et al. *A 240x180 130dB 3us latency global shutter spatiotemporal vision sensor*. IEEE JSSC, 2014
Example 1: Deblurring a blurry video

➢ A blurry image can be regarded as the integral of a sequence of latent images during the exposure time, while the events indicate the changes between the latent images.

➢ Finding: sharp image obtained by subtracting the double integral of event from input image

\[ \log \text{Input blur image} - \int \int \text{Input events} = \log \text{Output sharp image} \]
Example 1: Deblurring a blurry video

- **Blurry image**: can be regarded as the integral of a sequence of latent images during the exposure time, while the **events** indicate the changes between the latent images.

- **Finding**: Sharp image obtained by subtracting the double integral of event from input image.

![Input blur image](image1.png) ![Output sharp video](image2.png)

Pan et al., Bringing a Blurry Frame Alive at High Frame-Rate with an Event Camera, CVPR’19
Example 1: Deblurring a blurry video

- **A blurry image** can be regarded as the integral of a sequence of *latent images* during the exposure time, while the **events** indicate the changes between the latent images.

- **Finding**: sharp image obtained by subtracting the double integral of event from input image

---

Pan et al., Bringing a Blurry Frame Alive at High Frame-Rate with an Event Camera, CVPR’19
What about an asynchronous Luca-Kanade-Tomasi (KLT) Tracker for Event Cameras?

Asynchronous, Photometric Feature Tracking using Events and Frames

- **Goal:** Extract features on **frames** and track them using only **events** in the **blind time** between two **frames**
- Uses the event generation model via **joint estimation of patch warping and optic flow**

Source code: [https://github.com/uzh-rpg/rpg_eklt](https://github.com/uzh-rpg/rpg_eklt)


[PDF], [YouTube], [Evaluation Code], [Tracking Code]
High-Speed, Near-Eye Gaze Tracking

**Task:** Estimate gaze-vector of the eye at >1'000Hz in real-time on portable system

- Update Rate beyond 10 kHz
- Accuracy comparable to commercial product (EyeLink)
- Portable and low-power

Example Applications:
- AR/VR
- Driver drowsiness detection

Angelopoulos et al., “Event Based, Near-Eye Gaze Tracking Beyond 10,000Hz”, Arxiv20. [PDF]
High-Speed, Near-Eye Gaze Tracking

**Task:** Estimate gaze-vector of the eye at >1'000Hz in real-time on portable system

- Update Rate beyond 10 kHz
- Accuracy comparable to commercial product (EyeLink)
- Portable and low-power

<table>
<thead>
<tr>
<th>system</th>
<th>update rate (Hz)</th>
<th>accuracy (°)</th>
<th>portable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pupil Labs [1]</td>
<td>200</td>
<td>~ 1</td>
<td>✓</td>
</tr>
<tr>
<td>Tobii [3]</td>
<td>120</td>
<td>0.5–1.1</td>
<td>✓</td>
</tr>
<tr>
<td>EyeLink [2,14]</td>
<td>1,000</td>
<td>~ 0.5</td>
<td></td>
</tr>
<tr>
<td>Ours</td>
<td>&gt; 10,000</td>
<td>0.45–1.75</td>
<td>✓</td>
</tr>
</tbody>
</table>

Example Applications:
- AR/VR
- Driver drowsiness detection

Angelopoulos et al., “Event Based, Near-Eye Gaze Tracking Beyond 10,000Hz”, Arxiv20. [PDF]
High-Speed Human-Motion Tracking

**Task:** 3D human motion capture at 1'000 Hz

- 30x lower data-bandwidth than high-speed frame-based approach
- Works even in low-light
- Utilize frames (25 Hz) and events

---


Recap

- All the approaches seen so far enable asynchronous, low-latency (~10μs) algorithmic update on an event-by-event fashion.

- However:
  - Event-by-event update requires GPU for real-time processing.
  - Additionally, they make use of the generative event model:

\[
\pm C = \log I(x, t) - \log I(x, t - \Delta t)
\]

or its 1st order approximation

\[
\pm C = -\nabla L \cdot u
\]

which requires knowledge of the contrast sensitivity \( C \) (which is scene dependent and might differ from pixel to pixel).
Focus Maximization for:

- Motion estimation
- 3D reconstruction
- SLAM
- Optical flow estimation
- Feature tracking
- Motion segmentation
- Unsupervised learning
Idea: Warp spatio-temporal volume of events to maximize focus (e.g., sharpness) of the resulting image

Gallego et al., A Unifying Contrast Maximization Framework for Event Cameras, CVPR18, [PDF](https://example.com), [YouTube](https://example.com)
Gallego et al., Focus Is All You Need: Loss Functions for Event-based Vision, CVPR19, [PDF](https://example.com)
Focus Maximization Framework

Input Events

Image of

Focus score

\[
I(x; \theta) = \sum_{k=1}^{N_e} b_k \delta(x - x_k')
\]

Optimize point trajectories

\[
\sigma^2(I(x; \theta))
\]

Can be implemented in a sliding-window fashion to enable per low-latency, per-event update rate

Runs in real time on a CPU

Gallego et al., A Unifying Contrast Maximization Framework for Event Cameras, CVPR18, [PDF, YouTube]
Gallego et al., Focus Is All You Need: Loss Functions for Event-based Vision, CVPR19, [PDF].
Application 1: Image Stabilization

- Problem: Estimate rotational motion (3DoF) of an event camera
- Can process millions of events per second in real time on a smartphone CPU (OdroidXU4)
- Works up to over $\sim 1,000 \text{ deg/s}$

Gallego et al., Accurate Angular Velocity Estimation with an Event Camera, IEEE RAL’16. [PDF](#). [Video](#).
Application 3: 3D Reconstruction from a Train at 200km/h

In collaboration with SIEMENS and the Swiss Railway company, SBB

Rebecq et al., EVO: A Geometric Approach to Event-based 6-DOF Parallel Tracking and Mapping, RAL’17. PDF Video
Application 4: Motion Segmentation

Sequence: Fan and Coin

One motion model is used per cluster; one for the fan, modelling rotation, one for the coin, modelling optic flow

Motion-Compensated Segmented Events

Stoffregen et al., Motion Segmentation by Motion Compensation, ICCV’19. PDF. Video.
Application 4: Motion Segmentation

Stoffregen et al., Motion Segmentation by Motion Compensation, ICCV’19. PDF. Video.
Application 5:
Drone Dodging Dynamic Obstacles
Event-based Dynamic Obstacle Detection & Avoidance

➢ Works with relative speeds of up to 10 m/s
➢ Perception latency: 3.5 ms

Falanga et al., *Dynamic Obstacle Avoidance for Quadrotors with Event Cameras*, *Science Robotics*, 2020. PDF. Video
Falanga et al. *How Fast is too fast? The role of perception latency in high speed sense and avoid*, RAL’19. PDF. Video
How can we separate events triggered by ego-motion from events triggered by the moving object?

Falanga et al., *Dynamic Obstacle Avoidance for Quadrotors with Event Cameras*, Science Robotics, 2020. [PDF](#). [Video](#)

Falanga et al. *How Fast is too fast? The role of perception latency in high speed sense and avoid*, RAL’19. [PDF](#). [Video](#)
**Idea:** Warp spatio-temporal volume of events to maximize contrast of the resulting image: Static objects will appear sharp, while moving ones will appeared blurred.

Falanga et al., *Dynamic Obstacle Avoidance for Quadrotors with Event Cameras*, Science Robotics, 2020. [PDF](#). [Video](#)

Falanga et al. *How Fast is too fast? The role of perception latency in high speed sense and avoid*, RAL’19. [PDF](#). [Video](#)
Uses Rino event camera from Insightness

Falanga et al., *Dynamic Obstacle Avoidance for Quadrotors with Event Cameras*, *Science Robotics*, 2020. [PDF. Video]
Falanga et al. *How Fast is too fast? The role of perception latency in high speed sense and avoid*, RAL’19. [PDF. Video]
UltimateSLAM: combining events, images, and IMU for robust visual SLAM in HDR and High Speed Scenarios
UltimateSLAM: combining Events + Frames + IMU

Front End:
Feature tracking from Events and Frames

Back-End
State-of-the-art Sliding-Window Visual-inertial Fusion

Rosinol et al., Ultimate SLAM? RAL’18 – Best RAL’18 Paper Award Honorable Mention PDF. Video. IEEE Spectrum.
Mueggler et al., Continuous-Time Visual-Inertial Odometry for Event Cameras, TRO’18. PDF
Application: Autonomous Drone Navigation in Low Light

UltimateSLAM running on board (CPU: Odroid XU4)

Rosinol et al., Ultimate SLAM? RAL’18 – Best RAL’18 Paper Award Honorable Mention PDF, Video, IEEE Spectrum.
Mueggler et al., Continuous-Time Visual-Inertial Odometry for Event Cameras, TRO’18. PDF
UltimateSLAM: Frames + Events + IMU

85% accuracy gain over standard Visual-Inertial SLAM in HDR and high speed scenes

Rosinol et al., Ultimate SLAM? RAL’18 – Best RAL’18 Paper Award Honorable Mention PDF. Video. IEEE Spectrum.
Mueggler et al., Continuous-Time Visual-Inertial Odometry for Event Cameras, TRO’18. PDF
Learning with Event Cameras

• **Synchronous, Dense**, Artificial Neural Networks (ANNs) designed for standard images

• **Asynchronous, Sparse** ANNs

• **Asynchronous, Spiking** Neural Networks (SNNs)
Input representation

How do we pass sparse events into a convolutional neural network designed for standard images?

Video from Zhu et al. (link)
Input representation

[Maqueda CVPR’18], [Zhu’RSS’18]
• Aggregate positive and negative events into separate channels
• Discards temporal information

[Zhu ECCVW’18], [Rebecq, CVPR’19], [Zhu, CVPR’19]
• Represent events in space-time into a 3D voxel grid \((x,y,t)\)
• Each voxel contains sum of ON and OFF events falling within the voxel
• Preserves temporal information but discards polarity information

[Gehrig, ICCV’19]
• Represent events in space-time as a 4D Event Spike Tensor \((x,y,t,p)\)
• Polarity information is preserved
Input representation

Discretized 3D volume (x,y,t): events are inserted into the volume with trilinear interpolation, resulting in minimal loss in resolution

[Video from Zhu et al., CVPR’19]

[Zhu, ECCVW’18], [Zhu, CVPR’19], [Gehrig, ICCV’19], [Rebecq, CVPR’19]
Focus used as loss: maximize sharpness of the aggregated event image.

Zhu, Unsupervised Event-based Learning of Optical Flow, Depth and Egomotion, CVPR 19. [PDF]
Gallego et al., Focus Is All You Need: Loss Functions for Event-based Vision, CVPR19, [PDF].

Video from here
Focus as Loss Function for Unsupervised Learning

• We proposed and benchmarked **22 focus loss functions**

• Focus is the “data fidelity” term

Gallego et al., Focus Is All You Need: Loss Functions for Event-based Vision, CVPR19, [PDF].
Application: Unsupervised Learning of Optical Flow, Depth and Ego Motion

Focus used as loss: maximize sharpness of the aggregated event image.

Fidget Spinner w/ Challenging Lighting

Grayscale Image w/ Sparse Flow Quiver
Dense Flow Output

Zhu et al., Unsupervised Learning of Optical Flow, Depth and Ego Motion, CVPR’19
Application:
Learning High-speed and HDR Video Rendering from an Event Camera

Code & datasets: https://github.com/uzh-rpg/rpg_e2vid

Rebecq et al., “Events-to-Video: Bringing Modern Computer Vision to Event Cameras”, CVPR19. PDF Video.
Image Reconstruction from Events

Rebecq et al., “Events-to-Video: Bringing Modern Computer Vision to Event Cameras”, CVPR19. [PDF Video].

Code & datasets: [https://github.com/uzh-rpg/rpg_e2vid](https://github.com/uzh-rpg/rpg_e2vid)
Overview

- **Recurrent neural network** (main module: Unet)

- Input: last reconstructed frame + **sequences of event tensors** (spatio-temporal 3D voxels grid: each voxel contains sum of ON and OFF events falling within the voxel)

- Network processes **last N events** (10,000)

- **Trained in simulation only** (without seeing a single real image) (we used our event camera simulator: [http://rpg.ifi.uzh.ch/esim.html](http://rpg.ifi.uzh.ch/esim.html))

Event Camera Simulator


Rebecq, ESIM: an Open Event Camera Simulator, CORL'18. [PDF](#), [YouTube](#), [Project page](#)
High Speed Video Reconstruction Results

Code & datasets: https://github.com/uzh-rpg/rpg_e2vid

Rebecq et al., “Events-to-Video: Bringing Modern Computer Vision to Event Cameras”, CVPR19. PDF Video.
Bullet shot by a gun (376m/s (=1,354km/h))

Recall: trained in simulation only!

Huawei P20 Pro (240 FPS)

Our reconstruction (5400 FPS)

We used Samsung DVS

Code & datasets: https://github.com/uzh-rpg/rpg_e2vid

Rebecq et al., “Events-to-Video: Bringing Modern Computer Vision to Event Cameras”, CVPR19. PDF Video.
Bullet shot by a gun (1,300 km/h)

Recall: trained in simulation only!

Huawei P20 Pro (240 FPS)

Our reconstruction (5400 FPS)
We used Samsung DVS

100 x slow motion

Code & datasets: https://github.com/uzh-rpg/rpg_e2vid

Rebecq et al., “Events-to-Video: Bringing Modern Computer Vision to Event Cameras”, CVPR19. PDF Video.
Bullet shot by a gun (1,300 km/h)

Recall: trained in simulation only!

Huawei P20 Pro (240 FPS)

Our reconstruction (4800 FPS)

We used Samsung DVS

100 x slow motion

Code & datasets: https://github.com/uzh-rpg/rpg_e2vid

Rebecq et al., “Events-to-Video: Bringing Modern Computer Vision to Event Cameras”, CVPR19. PDF Video.
Bullet shot by a gun (1,300 km/h)

Recall: trained in simulation only!

Huawei P20 Pro (240 FPS)

Our reconstruction (1,300 FPS)

We used Samsung DVS

100 x slow motion

Code & datasets: https://github.com/uzh-rpg/rpg_e2vid

Rebecq et al., “Events-to-Video: Bringing Modern Computer Vision to Event Cameras”, CVPR19. PDF Video.
Popping a water balloon

Recall: trained in simulation only!

Apple iPad (120 FPS)

* different sequences, recorded in identical conditions

Our reconstruction (4800 FPS)

We used Samsung DVS

Real time

Code & datasets: https://github.com/uzh-rpg/rpg_e2vid

Rebecq et al., “Events-to-Video: Bringing Modern Computer Vision to Event Cameras”, CVPR19. PDF Video.

Popping a water balloon

Recall: trained in simulation only!

Apple iPad (120 FPS)

* different sequences, recorded in identical conditions

Our reconstruction (4800 FPS) 100 x slow motion

Code & datasets: https://github.com/uzh-rpg/rpg_e2vid

Rebecq et al., “Events-to-Video: Bringing Modern Computer Vision to Event Cameras”, CVPR19. PDF Video.

HDR Video Reconstruction Results

Code & datasets: https://github.com/uzh-rpg/rpg_e2vid
Rebecq et al., “Events-to-Video: Bringing Modern Computer Vision to Event Cameras”, CVPR19. PDF Video.
HDR Video: Driving out of a tunnel

Recall: trained in simulation only!

Events   Our reconstruction   Phone camera

Code & datasets: [https://github.com/uzh-rpg/rpg_e2vid](https://github.com/uzh-rpg/rpg_e2vid)
Rebecq et al., “Events-to-Video: Bringing Modern Computer Vision to Event Cameras”, CVPR19. [PDF Video]
HDR Video: Night Drive

Recall: trained in simulation only!

Our reconstruction from events

GoPro Hero 6

Code & datasets: [https://github.com/uzh-rpg/rpg_e2vid](https://github.com/uzh-rpg/rpg_e2vid)

Rebecq et al., “Events-to-Video: Bringing Modern Computer Vision to Event Cameras”, CVPR19. [PDF Video](https://example.com).

Color video reconstruction

Color events

- Each pixel is sensitive to **red, green or blue** light.
- Transmits **brightness changes** in each color channel

Taverni et al., Front and back illuminated Dynamic and Active Pixel Vision Sensors comparison, TCS’18
Color Event Camera Reconstruction (HDR)

Color events

Our reconstruction

Color frame

Color Event Camera Datasets: [http://rpg.ifi.uzh.ch/CED.html](http://rpg.ifi.uzh.ch/CED.html)

Scheerlinck, Rebecq, Stoffregen, Barnes, Mahony, Scaramuzza

CED: Color Event Camera Dataset. CVPRW, 2019. [PDF][YouTube][Dataset]
Downstream Applications:
What if we input the reconstructed frames to state of the art ML algorithms?

Code & datasets: https://github.com/uzh-rpg/rpg_e2vid
Rebecq et al., “Events-to-Video: Bringing Modern Computer Vision to Event Cameras”, CVPR19. PDF Video.
Monocular Depth Estimation

Events  Our reconstruction  Monocular depth estimation (Megadepth) applied on the reconstructed frames

Code & datasets: https://github.com/uzh-rpg/rpg_e2vid
Rebecq et al., “Events-to-Video: Bringing Modern Computer Vision to Event Cameras”, CVPR19. PDF Video.
Object detection

Events

Our reconstruction + object detections (YOLOv3)

Code & datasets: [https://github.com/uzh-rpg/rpg_e2vid](https://github.com/uzh-rpg/rpg_e2vid)
Rebecq et al., “Events-to-Video: Bringing Modern Computer Vision to Event Cameras”, CVPR19. [PDF Video].
Does it mean that in order to use event cameras we must first reconstruct an image?

**NO!**

These results were only to show that it should be possible to design more efficient algorithms that process events end-to-end without passing through image reconstruction!

However, to design end-to-end approaches for event cameras, we need more data! But event cameras are new, so there is a shortage of large scale datasets compared to standard cameras!

Is it possible to recycle existing large-scale video datasets recorded with standard cameras for event cameras?
Idea: convert Standard videos to events!

Code: https://github.com/uzh-rpg/rpg_vid2e

How to we convert a standard video to events?

- Typical video has a low temporal resolution and needs to be upsampled first.
- We use off-the-shelf upsampling techniques (Super SloMo [Jiang, CVPR’18]).
- Event generation using our event camera simulator (http://rpg.ifi.uzh.ch/esim.html).
- Noise-free simulation. We randomize the contrast sensitivity.

Code: https://github.com/uzh-rpg/rpg_vid2e

Experiments on Semantic Segmentation

Code: https://github.com/uzh-rpg/rpg_vid2e

Generalization to challenging HDR scenario

Code: https://github.com/uzh-rpg/rpg_vid2e

Binas et al., “DDD17: End-To-End DAVIS Driving Dataset”, ICMLW’17.
Computational Photography
Revolution
Event-based Super-Resolution

- Given low-resolution events as input, reconstruct a high-resolution image
- For standard images, the spatial resolution is fixed and cannot change
- For event data, high spatial resolution may be hidden in the temporal resolution of the data. Networks can exploit this!

Mostafavi I. et al., “Learning to Super Resolve Intensity Images from Events”, CVPR20
Learning with Event Cameras

- **Synchronous, Dense**, Artificial Neural Networks (ANNs) designed for standard images
- **Asynchronous, Sparse** ANNs
- **Asynchronous, Spiking** Neural Networks (SNNs)
Adapting Neural Networks To Event-based Data

Event-based Asynchronous Sparse Convolutional Networks

- Convolutional neural networks process events as dense images.
- However, event data is inherently sparse and asynchronous, meaning that wasteful computation is being performed.
- We can save computation by adopting **sparse convolutions** [Graham CVPR’18] which only compute the convolution at active pixels.

Messikommer et al., “Event-based Asynchronous Sparse Convolutional Networks”, ECCV’20. [PDF Video]
Event-based Asynchronous Sparse Convolutional Networks

➢ For each new event we do not have to update the full network layers. Just the pixels which are in the receptive field of the pixel which triggered the event.

➢ For regular convolutions this receptive field grows quadratically with the depth of the network. However, for sparse convolutions it grows much more slowly.

➢ This growth rate of the receptive field is related to the fractal dimension, which is an intrinsic property of event data.

Learning with Event Cameras

- **Synchronous, Dense** Artificial Neural Networks (ANNs) designed for standard images
- **Asynchronous, Sparse** ANNs
- **Asynchronous, Spiking** Neural Networks (SNNs)
Spiking Neural Networks (SNN)

- Common processing units based on Von Neumann architectures (CPU and GPU) are inefficient & very power consuming for event-by-event processing [1]
- There exists very efficient, specialized hardware for event-by-event inference: IBM TrueNorth [1], Intel Loihi [2], DynapSE & Speck (AiCTX) [3]
- Promising for Robotics, IoT, VR/AR/MR
  - Low power
  - Low latency
  - Leverage event-based sensing
- Promise ultra-low carbon footprint!

The Cost of Current Computer Technologies is Not Sustainable

➢ In 2017, > 10 zettabytes of data were produced.

➢ IT infrastructures and consumer electronics absorbed > 10% of the global electricity supply.

➢ By 2025, over 50 billion of Internet-of-Things (IoT) devices will be interconnected.

➢ Over 180 zettabytes of data will be generated annually, potentially leading to a consumption of one-fifth of global electricity (source, Nature, Feb., 2018)

“Software companies make headlines but research on computer could bring bigger rewards. “
Training a single AI model can emit as much carbon as five cars in their lifetimes

Deep learning has a terrible carbon footprint.

by Karen Hao

Jun 6, 2019

The artificial-intelligence industry is often compared to the oil industry: once mined and refined, data, like oil, can be a highly lucrative commodity. Now it seems the metaphor may extend even further. Like its fossil-fuel counterpart, the process of deep learning has an outsize environmental impact.
Radical paradigm shift in computer hardware technologies

- Our brain is slow, noisy ("speed" is not a requirement)
- Massively parallel distributed computation, local connectivity (minimize wiring)
- Real-time interaction with the environment
- Complex spatio-temporal pattern recognition
Model of a Spiking Neuron
Origins of Spiking Neural Networks

- **First model** (integrate-and-fire) of a spiking neuron in 1907 by Louis Lapicque [1]
- **First computational model for neural networks in 1943** [2]: Neural network research split into biological processes in the brain and the application for artificial intelligence
- First scientific model of biological spike propagation by Hodgkin and Huxley in 1951 [3] (Nobel Prize in Physiology)
- A range of more general spiking neuron models are available nowadays [4]

[1]: Lapicque L. Recherches quantitatives sur l'excitation electrique des nerfs traitee comme une polarisation. *Journal de Physiologie et de Pathologie Generale*. 1907
[2]: McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*. 1943
SNNs: Current Applications and Demos

- IBM TrueNorth
  - Targeting ultra low-power gesture control, video-surveillance and IoT with SNN on digital processor
  - At CVPR 2017 gesture recognition demo (10 gestures)
  - 96.5 % recognition accuracy
  - 200 mW power consumption (event-camera + processing)

- Intel Loihi:
  - Targeting ultra low-power surveillance and IoT with SNN on analog processor

- aiCTX (Zurich-based startup):
  - Targeting ultra low-power surveillance and IoT with SNN on analog processor
  - At CES’19 and CVPR’19, they demonstrated face recognition from event data on an SNN processor; total power consumption: 1mW

[1]: https://aictx.ai/
Spiking Neural Network (SNN) Regression

**Task:** Estimate angular velocity of an event camera with an SNN

- SNNs are *asynchronous networks* that consume events
- Hence,
  - no additional latency
  - no preprocessing necessary
- We show that SNNs are competitive to ANNs on this task

Gehrig et al., "Event-Based Angular Velocity Regression with Spiking Networks", ICRA20 [PDF](#)
Conclusions, Takeaways, Resources
Recap

➢ **Event cameras** have many **advantages**:  
  • high dynamic range (HDR)  
  • high speed  
  • low latency  
  • low power

➢ **Current commercial applications**  
  • IoT  
    • monitoring and surveillance  
  • Automotive:  
    • low-latency detection, object classification, low-power and low-memory storage  
  • AR/VR  
    • low-latency, inter-frame pose estimation, low-power  
  • Industrial automation  
    • Fast pick and place
Research Challenges with Event Cameras

➢ Quantify the **trade-offs**:
  • **Latency vs. power consumption** and **accuracy**
  • **Sensitivity vs. bandwidth** and **processing** capacity

➢ **Active parameter adaptation**

➢ **Hardware:**
  ▪ pairing event cameras with dedicated hardware (SNN hardware, e.g., Intel Loihi, aiCTX Speck)
  ▪ How do we make sparse convolution in space and time efficient?

➢ **Learning** with event cameras:
  • How do we **exploit knowledge from image-based learning to event cameras**?
  • **Asynchronous** inference
  • **Where do we find learning data**? Event data is much more rare than frames. Potential solutions: unsupervised Learning, learning in simulation, transfer learning from frames to events
Event-based Vision: A Survey

Guillermo Gallego, Tobi Delbrück, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew Davison, Jörg Conradt, Kostas Daniilidis, Davide Scaramuzza

Abstract—Event cameras are bio-inspired sensors that work radically different from traditional cameras. Instead of capturing images at a fixed rate, they measure per-pixel brightness changes asynchronously. This results in a stream of events, which encode the time, location and sign of the brightness changes. Event cameras possess outstanding properties compared to traditional cameras: very high dynamic range (140 dB vs. 60 dB), high temporal resolution (in the order of 1ms), low power consumption, and do not suffer from motion blur. Hence, event cameras have a large potential for robotics and computer vision in challenging scenarios for traditional cameras, such as high speed and high dynamic range. However, novel methods are required to process the unconventional output of these sensors in order to unlock their potential. This paper provides a comprehensive overview of the emerging field of event-based vision, with a focus on the applications and the algorithms developed to unlock the outstanding properties of event cameras. We present event cameras from their working principle, the actual sensors that are available and the tasks that they have been used for, from low-level vision (feature detection and tracking, optic flow, etc.) to high-level vision (reconstruction, segmentation, recognition). We also discuss the techniques developed to process events, including learning-based techniques, as well as specialized processors for these novel sensors, such as spiking neural networks. Additionally, we highlight the challenges that remain to be tackled and the opportunities that lie ahead in the search for a more efficient, bio-inspired way for machines to perceive and interact with the world.


1 INTRODUCTION AND APPLICATIONS

“T
e brain is imagination, and that was exciting to me; I wanted to build a chip that could imagine something!” that is how Misha Mahowald, a graduate student at Caltech in 1986, started to work with Prof. Carver Mead on the stereo problem from a joint biological and engineering per-
CVPR19 Workshop on Event-based Vision

➢ Full-day workshop with talks by 23 researchers on event-based cameras, including Samsung, Intel, and event-camera companies

➢ Slides and video recordings:

http://rpg.ifi.uzh.ch/CVPR19_event_vision_workshop.html

Second International Workshop on Event-based Vision and Smart Cameras
June 17, Long Beach

Held in conjunction with the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach.
List of Event-based Vision resources

Code, datasets, papers, videos, companies on event cameras

https://github.com/uzh-rpg/event-based_vision_resources
UZH-FPV Drone Racing Dataset & Competition

- Recorded with a drone flown by a professional pilot up to over 20m/s
- Contains over 30 sequences with images, events, IMU, and ground truth from a robotic total station: [http://rpg.ifi.uzh.ch/uzh-fpv.html](http://rpg.ifi.uzh.ch/uzh-fpv.html)

Thanks!

Code, datasets, simulators, papers, and videos:
http://rpg.ifi.uzh.ch/research_dvs.html

Research updates:
@davsca1
@davidescaramuzza