Event cameras

Event cameras, also called neuromorphic cameras, silicon retina or dynamic vision sensor (DVS) are image sensors that respond to changes in brightness in the scene. Unlike conventional cameras, which capture full image frames at a fixed frequency (commonly 30Hz or 60Hz), producing redundant information and require a high bandwidth for transmission, each pixel in an event-based camera operate independently and asynchronously, reacting to changes of brightness in the scene, eliminating the transmission of redundant information, allowing for much higher temporal resolution (in the order of microseconds, as opposed to the milliseconds of conventional cameras).

Event cameras are inspired by the behaviour of the cells in the retina. Though oversimplified, retinal cells respond to changes in the environment (namely brightness), generating an electrical impulse. The transient response of each retinal cell is independent. Event cameras mimic this behaviour by asynchronously and independently responding to changes in brightness in the environment, generating ON/OFF events each time a predefined threshold in brightness is exceeded.

This architecture allows for interesting properties, such as microsecond temporal resolution, high dynamic range (above 120dB), which allows for scenes with both bright and dark zones, and does not suffer from under/overexposure, nor motion blur.

Events are triggered when the brightness in a certain pixel surpasses a certain threshold. In particular, discrete brightness steps are pre-defined, and whenever brightness detected crosses the threshold, an event is generated. Positive crossings generate ON events, and negative crossings generate OFF events. In effect, each pixel is constantly working as a comparator (with corresponding electronic to support this mode of working). The figure shows the first-generation DVS sensor, with an array of 128x128 pixels (from [Lichtsteiner08]), where this idea of comparison and threshold is shown.

A close up of a device

Description automatically generated

Figure 1 Event camera (a) and corresponding mode of working (b), from [Clady15]

Events are then defined as a four-component vector:

The component refers to the spatial position of the event in the camera. The component refers to the timestamp of the event, and is of extreme importance due to the microsecond temporal resolution of the camera. Lastly, the parameter refers to the polarity of the event (ON/OFF events).

With this event structure, it is common to represent events in a three-dimensional (space-time), representation, as shown in the figure, which shows the space-time evolution of events generated from a rotating black bar on a white background.

A close up of a map

Description automatically generated

Figure 2 Space-time representation of events generated from a rotating black bar (from [Clady15])

Conventional cameras and event cameras have fundamentally different modes of operation and output. As such, a comparison of the behaviour in the same scene, and an analysis of the output, is interesting. The figure shows the response of both a conventional camera and an event camera when presented with a disk rotating at a high speed, with a black dot. The fixed capture of the conventional camera is unable to keep up with the speed of the dot, and the images suffer from motion blur and some discontinuity. The event camera, however, due to its asynchronous event generation, is able to continuously produce events relating to the movement of the dot.

A close up of text on a white background

Description automatically generated

Figure 3 Comparison between conventional and event cameras (from [Mueggler17])

Advances in camera manufacturing have allowed for cameras that have both conventional camera pixel arrays, and event camera pixel arrays.

References

[Mueggler17] : Fast Event-based Feature Detection

[Lichtsteiner08] : A 128 128 120 dB 15 us Latency Asynchronous Temporal Contrast Vision Sensor

[Clady15] : Asynchronous event-based corner detection and matching