May 2020

Understanding the Performance of Neuromorphic Event-based Vision Sensors

Contents

Introduction

Neuromorphic event-based vision is a new and rapidly expanding field. The technology promises advantages in different aspects of sensor performance including the latency, temporal resolution, and dynamic range. However, it is currently difficult to compare and verify claims from different players in the field, due to a lack of standardized testing methods. In addition, there is also ongoing discussion about the relative importance of different performance aspects for different use cases. In this overview, we discuss some of the key questions on interpreting the performance specifications of neuromorphic/event-based vision sensors in a question-and-answer format.

What is temporal resolution?

Temporal resolution of a neuromorphic event-based vision sensor is defined as the discrete measurement resolution of the event detection time. The temporal resolution is characterized by the time unit of the event timestamp. A smaller unit of the event timestamp corresponds to a higher temporal resolution.

Is higher temporal resolution always better?

The usefulness of the temporal resolution is limited by the pixel front-end noises as well as the design of the timestamping scheme.

The sources of the pixel front-end noises include the photon shot noise, the circuit thermal noise, and the fixed-pattern noise due to mismatches in the temporal contrast detection threshold across pixels. All these noise sources are defined by fundamental physics and are to a large extent unavoidable. This noise introduces uncertainty and non-uniformity in the event detection time, typically in the order of hundreds of μs under normal lighting conditions, as illustrated in Figure 1.

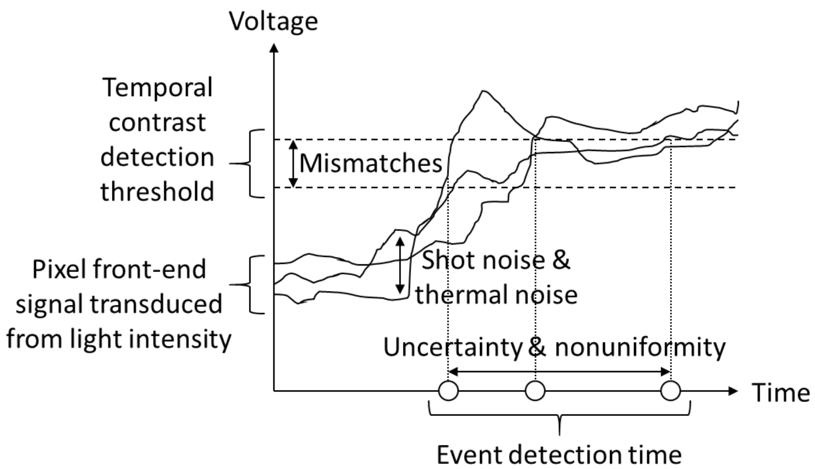

Figure 1 Uncertainty and non-uniformity in the event detection time

As an example, imagine the projection from a rigid edge in a scene moving across an array of pixels. This changes the incident light intensities in the pixels that receive the edge projection (we can call them projected pixels) as depicted in Figure 2.

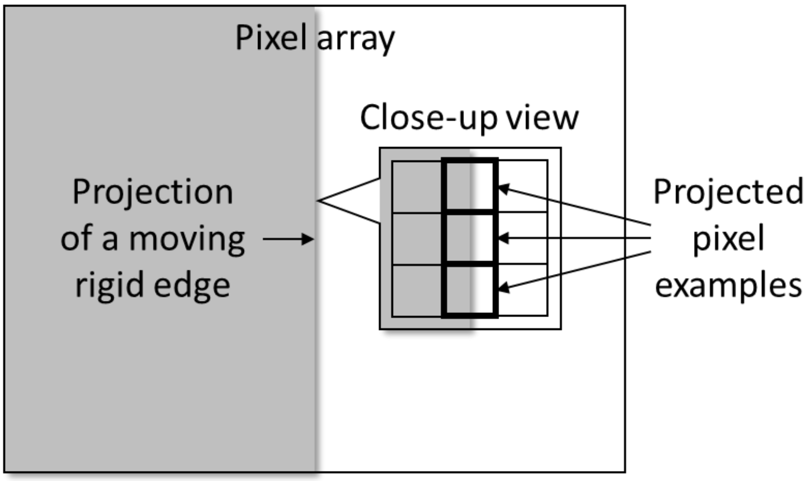

Figure 2 An example input stimulus

According to the ground truth input stimulus, at any instance in time, all of the projected pixels experience the incident light intensity changes at the same time, because these changes are caused by the movement of a single rigid edge. However, because of the pixel front-end noise, the actual event detection time differs among the projected pixels. If we are to plot the distribution of the actual event detection time in all projected pixels, it roughly resembles a curve shown in Figure 3.

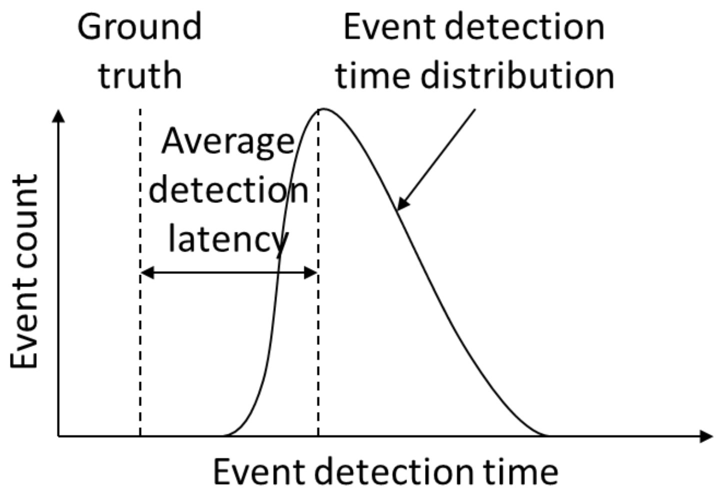

Figure 3 Distribution of the event detection time

This event detection time distribution can be quite faithfully captured if each event is timestamped as soon as it is generated, with high temporal resolution (e.g. 1 µs or even 1 ns). However, such faithful timestamping comes at a high cost – the timestamping circuit would need to be built into every pixel, which drastically increases pixel complexity and size. While the event detection timing distribution provides statistical information about the pixel front-end noises and the average detection latency, the chronological order of the event detection has no correlation with the ground truth. In practice, it is not clear that such precise timestamping is needed in any real-world use case.

Hence, most neuromorphic/event-based sensors have adopted a timestamping-during-readout scheme outside of the pixel array. Events are timestamped as they are being read from the pixel array, either one by one or in groups (rows/columns/whole array). Timestamping-during-readout means the resulting timestamp of an event contains an error caused by the readout latency. The readout latency of an event could range from <100 ns to >1 ms, depending on the maximum readout speed and the instantaneous event detection rate. As a result, the final event timestamps via a timestamping-during-readout scheme could look like Figure 4. Even if a 1 μs or even 1 ns timestamp unit is used, the timestamp distribution cannot reliably provide statistical information about the pixel front-end noise (if the instantaneous event detection rate is higher than the maximum readout speed), nor has the chronological order of the timestamps has any correlation with the ground truth.

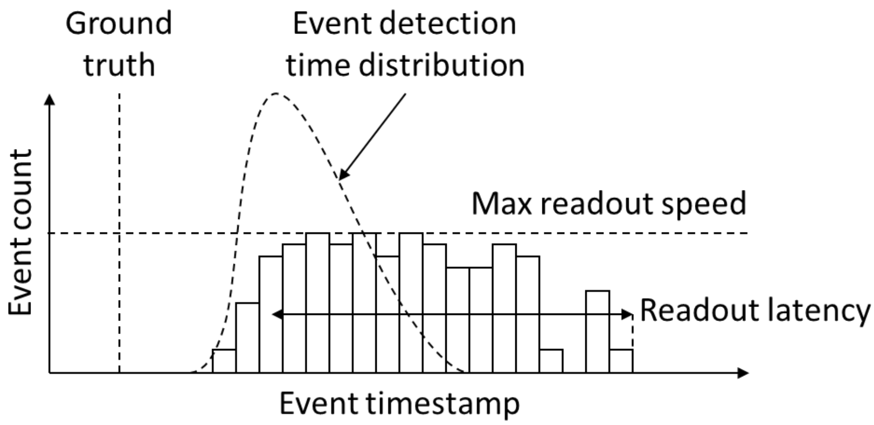

Figure 4 Distribution of the event timestamps using a timestamping-during-readout scheme

In summary, due to the pixel front-end noise, and the most commonly adopted timestamping-during-readout scheme, a small timestamp unit of 1 μs, or even 1 ns, provides little to no additional information about the ground truth than a coarser timestamp unit such as in the order of 100 μs or even 1 ms, depending on the use case. Instead, having an unnecessarily high temporal resolution requires a more complex timestamping system, as well as higher communication and processing costs to handle the extra information-less data. An unnecessarily high temporal resolution is costly on the system level, with no proven benefit. In fact, the usefulness of the event timestamps in a timestamping-during-readout scheme also largely depends on the maximum readout speed. The faster the event can be read out, the smaller the timestamp errors, as will be explained later in this document.

Taking these considerations into account, a sensor with a low quoted temporal resolution but high maximum readout speed (e.g. our DVXplorer camera with a temporal resolution of 200 μs, and a maximum event throughput of 165 MEPS) can be actually much more practical than a sensor with a much higher quoted temporal resolution but much lower maximum readout speed (e.g. our DAVIS346 camera with a temporal resolution of 1μs, and a maximum event throughput of 12 MEPS).

What are the pros and cons of an asynchronous event readout scheme and a synchronous event readout scheme?

An asynchronous event readout scheme means that the events are communicated from the pixel array to a host receiver via an un-clocked handshake communication protocol. When multiple new events are detected while a previous event is being communicated, they are considered to be simultaneous. These simultaneous events are communicated in an order determined by an arbiter system. These simultaneous events are usually timestamped during readout, so they may not share the same timestamp. The key advantages of an asynchronous event readout scheme are its low power consumption and low readout latency, both of which are only true if the event detection rate is lower than the maximum readout speed. The main disadvantage of an asynchronous event readout scheme is its limited maximum readout speed.

In an asynchronous event readout scheme, the communication of an event is initiated by the detection of an event. If the event detection rate is low, the readout system is less active, or even idle at times, therefore, consuming little power. Furthermore, if the event detection rate is lower than the maximum readout speed, the communication of a newly detected event is initiated immediately by its detection with minimum delay. As an example, our DAVIS346 camera which implements an asynchronous event readout scheme needs about 80 ns to communicate one event. Therefore, if the event detection rate is lower than one event in every 80 ns (or about 12 MEPS), the DAVIS346 camera's best-case readout latency is about 80 ns.

However, as the event detection rate increases, an asynchronous event readout scheme becomes less power efficient. This is because the handshake communication protocol requires multiple exchanges of information between the pixel array and the host receiver to communicate one single event. Furthermore, if the instantaneous event detection rate is higher than the maximum readout speed, the number of simultaneous events queueing to be communicated starts to accumulate, resulting in an increased readout latency. This increase in readout latency is non-deterministic for each individual event due to the use of arbitration in determining the queueing sequence. Taking our DAVIS346 camera (which has an maximum event throughput of 12 MEPS) as an example, if an input stimulus produces 10k simultaneous events (in about 10% of the total 0.1M pixels), the worst-case readout latency of these events is about 1 ms, introducing a large timestamp error.

As shown in Figure 5, if new events are detected within this 1 ms, they are inserted into the communication queue by arbitration, potentially further increasing the readout latency and timestamp error of the previously accumulated simultaneous events.

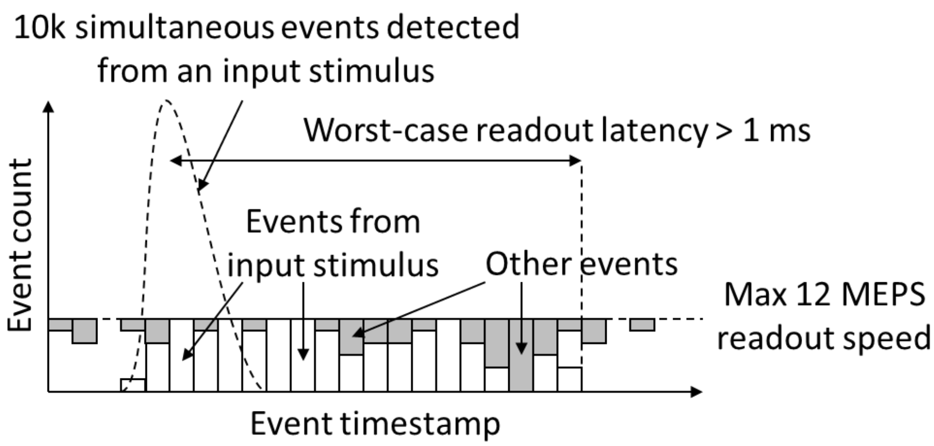

Figure 5 Worst-case readout latency in an asynchronous event readout scheme (DAVIS346)

A synchronous event readout scheme means the events are communicated from the pixel array to a host receiver via a clocked systematic scanning scheme. A synchronous event readout scheme is typically used in combination with a global-shutter-like global event sampling mechanism, which means the whole pixel array are sampled simultaneously at known (and adjustable) time intervals. When multiple events are detected and sampled in one such time interval, they are considered to be simultaneous and are communicated in a systematic order according to their positions in the pixel array, such as from top to bottom and left to right. The simultaneous events share the same timestamp and can be referred to as one event frame. The key advantages of a synchronous event readout scheme include its high maximum readout speed, as well as its low and constant readout latency at a high event detection rate. The disadvantages of a synchronous event readout scheme are its constant static power consumption and constant readout latency at a low event detection rate. The coarser timestamping as a result of the global event sampling operation is not a disadvantage, as explained previously in the answer to the temporal resolution question.

In a synchronous event readout scheme, the communication of events is dictated by a readout controller outside of the pixel array, typically following every global event sampling operation at a certain sampling rate. Therefore, the number of detected events during each global sampling interval is not known by the readout controller, which must initiate the communication of possible detected events following each global sampling operation. These periodic sampling and communication activities cause a constant non-zero static power draw even at zero event detection rate. Fortunately, through design optimization, the magnitude of this static power draw is usually negligible compared to the overall power consumption of the whole sensor. As the event detection rate increases, a synchronous event readout scheme becomes more power efficient, because the communication of each single event only requires one clocked information exchange.

The periodic communication initiated by the readout controller also means that the worst-case readout latency of a detected event is at least the duration of the global event sampling interval even at a low event detection rate. However, this worst-case readout latency does not scale with the event detection rate as long as the event detection rate is lower than the maximum readout speed. Because of its high maximum readout speed, a synchronous event readout scheme is able to maintain a constant worst-case readout latency even at a high event detection rate. For example, our DVXplorer camera implements a synchronous event readout scheme with a default global event sampling rate at 5 kHz, which means a worst-case readout latency of at least 200 µs at a low event detection rate. However, with a maximum readout speed at 165 MEPS, the DVXplorer is more than capable of sustaining 30k simultaneous events (produced by 10% of the total 0.3M pixels) in every event frame continuously, while keeping a worst-case readout latency of less than 400 µs (as shown in Figure 6, assuming an event is detected at the beginning of a global event sampling interval and is the last to read out in the event frame).

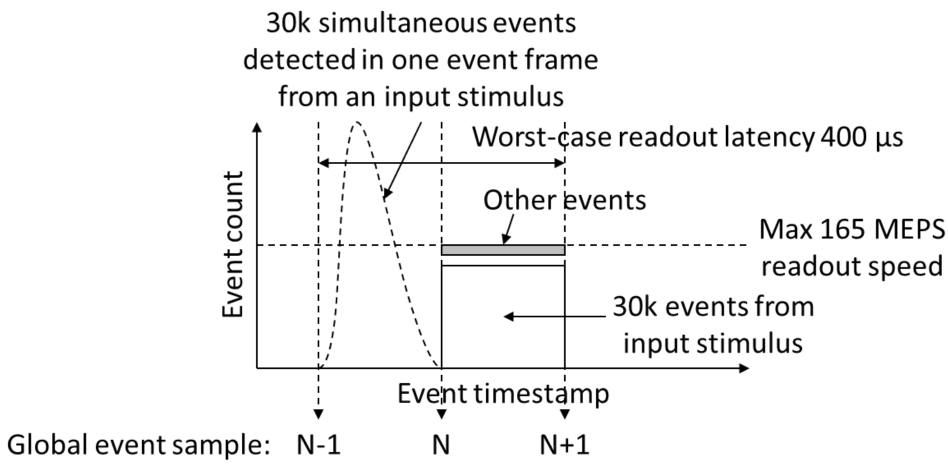

Figure 6 Worst-case readout latency in a synchronous event readout scheme (DVXplorer)

So which event readout scheme is better? The answer to this question depends on the sensor resolution and the use case. An asynchronous event readout scheme can be more suitable for low resolution sensors (e.g. our DAVIS346 sensor with 0.1M pixels), or use cases where the event detection rate is expected to be low. A synchronous event readout scheme is more suitable for high resolution sensors (e.g. our DVXplorer sensor with 0.3M pixels) or use cases where a high event detection rate is expected. Regardless of whether an asynchronous or synchronous readout scheme is used, the event output data is in the same address event representation (AER) format.

How is the latency determined?

The total latency of an event-based sensor is defined as the time elapsed from the moment the ground truth input stimulus occurs in the scene to the moment the corresponding events produced by the ground truth input stimulus are read out.

The total latency consists of the detection latency and the readout latency:

- The **detection latency** a pixel is measured from the moment the projecting ground truth input stimulus occurs to the moment the resulting event is detected inside the pixel.
- The **readout latency** is measured from the moment an event is detected inside a pixel to the moment this event is read out.

Due to the pixel front-end noises, the detection latency varies from time to time and across pixels. Therefore, the detection latency of a sensor usually refers to the average detection latency derived from a sampled group of pixels (as shown in the example in **Figure 3**). The detection latency depends on the scene illumination and the pixel front-end circuit bias configuration. The stronger the scene illumination, the faster the pixel front-end circuit reacts, hence the shorter the detection latency is. Also, if low detection latency is a key requirement of a use case, the pixel front-end circuit bias can be configured to increase reaction speed, at the cost of increased pixel front-end noise.

The readout latency varies depending on the position of the pixel inside the pixel array, the overall event detection rate (or the number of simultaneous events), as well as the event readout scheme. For many use cases where reliable performance is desirable, it is more important to know the worst-case readout latency based on the use case rather than the best-case readout latency.

In most published total latency measurement results of sensors using an asynchronous event readout scheme, only a small group of pixels were stimulated and measured to derive the total latency as the average of the sampled group. Therefore, these results does not reflect the total latency performance when the readout latency is close to the worst-case readout latency. Due to the slow readout speed, the worst-case readout latency of an asynchronous event readout scheme is relatively high in practice. Take our DAVIS346 camera as an example, if the use case expects the scene input stimulus to produce simultaneous events in up to 10% of the pixels, the worst-case readout latency of these events is about 1 ms, or even higher if new events are detected within this 1 ms time (as explained previously in Figure 5). In contrast, because of the fast readout of a synchronous event readout scheme, the worst-case readout latency is usually much smaller than an asynchronous event readout scheme. For example, if the same use case is applied to our DVXplorer camera, producing simultaneous events in 10% of the pixels, the worst-case readout latency is 400 μs (as explained previously in Figure 6).

How are dynamic range and low light performance determined?

Sensor dynamic range is the ratio between the highest and the lowest illumination level under which the sensor functions. A neuromorphic/event-based vision sensor has an intrinsic strength in dynamic range, because it detects logarithmic light intensity changes, and is hence able to cover a wide range of illumination levels. Like any imaging device, the stronger the illumination, the less noisy the event output, until an excessively high illumination level saturates the logarithmic conversion of the pixel front-end circuit, which reduces the temporal contrast sensitivity. At an excessively low illumination level, the signal becomes too diluted by the pixel front-end dark signal, which reduces the temporal contrast sensitivity. In addition, an excessively low illumination level causes excessive pixel front-end noise, which further degrades the actual signal.

The highest real-world natural illumination level is from direct sunlight, at about 100 klux. In most neuromorphic/event-based vision sensor designs, 100 klux is well below the saturation level of the logarithmic conversion of the pixel front-end circuit. Therefore, the highest functional illumination level of most neuromorphic/event-based vision sensors is quoted to be 100 klux. Although it is possible to claim a higher than 100 klux functional illumination level, it is not meaningful because such illumination levels are never reached in natural scenes. As a result, the dynamic range of a neuromorphic/event-based vision sensor is typically more directly determined by the lowest functional illumination level.

The lowest natural illumination level can approach complete darkness. Below a certain illumination level, further decrease in the illumination level results in that less of the output events are signal events produced by the input stimulus and more of the output events are noise events. In the prior literature (at the time this document is written), there is no agreed-upon standard on how many signal and noise events need to be present in the event output for the sensor to be considered functional. Some published lowest functional illumination level results were based on subjective visual impressions. Some other published results did not specify the measurement criteria. For our DVXplorer series sensors, we have adopted a conservative quantitative criterion for the sensor to be considered functional under low illumination – e.g. 99.9% of the pixels must detect an input with a contrast equal to the minimum contrast sensitivity level for the sensor to be considered still functional.

About us

At iniVation we create neuromorphic vision systems. Our bio-inspired intelligent technology offers unprecedented advantages over conventional machine vision systems: ultra-low response latency, low data rates, high dynamic range and ultra-low power consumption.

Founded by the inventors of event-based vision, iniVation combines decades of world-leading R&D experience with a deep network of >300 customers and partners across multiple industrial markets. Our customers include global top-10 companies in automotive, consumer electronics and aerospace.

www.iniVation.com

Thurgauerstrasse 40, 8050 Zurich, Switzerland