# Performance comparison of area-scan and event-based camera

Alkhazur Manakov[1], Helmut Herrmann[2], and Bernd Jähne[1,2]

[1] HCI at IWR, Heidelberg University,
Berliner Straße 43, 69120 Heidelberg
[2] AEON Verlag and Studio GmbH
Alter Rückinger Weg 31, 63452 Hanau

**Abstract** Event-based sensors asynchronously measure pixel brightness changes, and output a stream of events that encode the time, location and sign of the brightness changes instead of capturing images at a fixed rate, as conventional area-scan sensors do. Advantages of event-based sensors include: high temporal resolution and dynamic range (120 dB), low power consumption, and a compressed output stream. Comparison methodology between the two types of sensors is not available, therefore choosing between event-based and conventional area-scan camera for a given machine vision application is a challenge. We extended the dynamic range for the irradiance of the test equipment to 120 dB, characterized the performance in relation to irradiance (photon/(pixel s)), emulated event-based sensor functionality with conventional area-scan sensor and thus enabled a comparison. Normal EMVA 1288 standard measurement suffice for emulation, provided the irradiance series is dense enough. Area-scan cameras which meet the linear model of the EMVA 1288 standard, require no measurements, because it is possible to compute the emulated performance analytically. The comparison covers several area-scan cameras and three event-based cameras with different sensors.

**Keywords** Sensor characterisation, event-based, area-scan, performance, comparison

DOI: 10.58895/ksp/1000174496-9 erschienen in:
**Forum Bildverarbeitung 2024**
DOI: 10.58895/ksp/1000174496 | https://www.ksp.kit.edu/site/books/m/10.58895/ksp/1000174496/

## 1 Introduction

State-of-the-art image sensors suffer from limitations imposed by their frame-based operation. The sensors acquire the visual information as a series of "snapshots" recorded at a predetermined frame rate. Biology does not know the concept of a frame. Biological vision systems outperform the best state-of-the-art artificial vision devices. Frames are not the most efficient form of encoding visual information. Firstly, the world, the source of the visual information, unlike frames, works asynchronously and in continuous time. Classical machine vision approach faces a dilemma loosing information between the frames or choosing high frame-rate. The latter requires more complex acquisition and processing hardware, with large bandwidth connection between them. Secondly, each recorded frame conveys the information from all pixels, regardless of whether this information, or a part of it, has changed since the last frame had been acquired. Two frames adjacent, dynamic contents of the scene, contain redundant information. Acquisition and handling of these dispensable data consume valuable resources and translate into high transmission power dissipation, increased channel bandwidth requirements, increased memory size, and processing power demands. An engineering solution inspired by the biological pixel-individual, frame-free approach may be more efficient than a traditional one.

The most advanced bioinspired vision sensors today [1] follow the natural, event-driven, frame-free approach, capturing transient events in the visual scene. Pixel analogue electronics stores a reference brightness level, and continuously compares it to the current brightness level. If the difference in brightness exceeds a threshold, that pixel resets its reference level and generates an event: a discrete packet that contains the pixel address, timestamp and polarity (increase or decrease) of a brightness change. Some sensors of these type do instantaneous measurement of the illumination level [2]. These type of sensors are called *even-based sensors*.

Choosing between cameras equipped with event-based and conventional area-scan sensors for a given machine vision application is a challenge, since the comparison methodology between the two types of sensors is not available. A first step in this direction was performed by Manakov and Jähne [3], who established the main concepts of event-based sensors in extend to EMVA1288 characterization standard. Manakov et al [4] proposed the setup, data acquisition procedure and first measurement results. They also propose propose key performance indicators: event-delay and an analogous to signal-to-noise ratio defined for conventional area-scan cameras.

In this work we show the first direct performance comparison of event-based cameras with traditional are-scan cameras. The measuring equipment covers an extended irradiance range of 120 dB in order to cover the dynamic range of event-based cameras and also HDR area-scan cameras. In addition, we modified the characterization procedure to measure the performance not in relation to exposure (photon/pixel) as in the EMVA 1288 standard but to irradiance (photon/(pixel s)), because event-based cameras cannot be characterized by an exposure time. These changes enabled emulation the functionality of event-based sensor with conventional area-scan sensor. We demonstrate that for area-scan cameras which meet the simple linear model of the EMVA 1288 standard, no measurements are required, since their performance can be computed analytically. Thus, there is an additional advantage: it is possible to compute the best possible performance of an ideal area-scan camera with a quantum efficiency of one and no dark noise. Measurements of area-scan cameras require only normal EMVA 1288 measurements, provided the irradiance series is dense enough so that these measurements can be used to determine with which probability an intensity change can be detected, given a fixed exposure time with the corresponding frame rate.

## 2 Event-based sensor characterization basics

Sensitivity to small temporal contrasts, the response relation to the event-based sensor settings and its uniformity across the array are crucial performance parameters for the asynchronous, event-driven sensors. The minimum detectable temporal contrast or simply *noise equivalent contrast* is barely detectable by an event-based pixel step change of the irradiation level. Noise equivalent contrast sensitivity corresponds to the signal-to-noise ratio property of a conventional image sensor.

The simplest way of experimentally determining the irradiation contrast $\Delta E$ necessary for generating one event for given mean irradiance level $E$ and event threshold settings is gradually increase the stimulus step until an event is generated. In an ideal noise-free world, minimal found stimulus amplitude always results in an event when applied. In the real world conditions, the very same pixel will react differently to the same stimulus due to its, possibly different, initial condition, electronic noise, etc. Therefore, for event-based sensor characterization it has been proposed to operate with "event probability" instead [2,3]. It is defined for a given as ratio between the number of event responses $M$ and the number of applied stimuli $N$, while background irradiance level and all the sensor settings remain unchanged.

$$p = \frac{M}{N} \tag{1}$$

Plotting the "event probability" vs. stimulus amplitude, in an ideal noise-free world, would yield a step function. In reality, such curve would have an "S"-shape, and is therefore named *S-curve*. Analysis of an S-curve provides crucial information about the performance of event-based sensor at the background irradiation levels and temporal contrasts the S-curve was acquired for. The contrast at 50% event probability point of an S-curve is the barely sensible contrast, similar to conventional area-scan cameras [5]. The slope at this point of an S-curve indicates the amount of noise. High slopes make S-curve closer to a step function, the influence of noise is small, low slopes indicate significant influence of noise on event probability. Vertical offset of an S-curve, as shown in Fig.1 for irradiation levels around $10^2$ and $10^3$, indicate the presence of events in absence of temporal contrast. In the next section we describe the S-curve acquisition procedure in detail, present the results for three different event-based sensors and introduce a metric, which enables area-scan and event-based sensor comparison.

## 3 Measures S-curves and change detectability

### 3.1 S-curves acquisition

The acquisition of S-curves presented in this section has been done on an EMVA1288 Standard conform setup, which consists of an integrating sphere, 4 LED modules, filter wheel with neutral density filters and a calibrated photo-diode. The LED modules are electronically controlled to generate background irradiation level and generate irradiation impulse with a controlled length. The neutral density filters allow to extend the dynamic range of the system, reaching very low irradiation levels and sample the irradiation space densely. The calibrated photo-diode provides the reference for the background irradiation levels $E$ and the impulse amplitude $\Delta E$. The acquisition of an S-curve is conducted with fixed sensor settings, namely biases and event-thresholds. The acquisition is performed for various background irradiation levels, varied by more than 6 orders of magnitude using neutral density filters. Each background irradiation level yields an S-curve. The sensor is stimulated by many impulses of various amplitudes are acquired for each background irradiation level. Thus, every sample of an S-curve corresponds to a pair (background irradiation; impulse amplitude). The measurement for each pair is repeated several hundred times for computing per-pixel event-probability. Three different event-based sensors were used in our measurements:

- Prophesee, gen. 3.1 (resolution 640x480, pixel 15μm x 15μm)
- Prophesee, gen. 4.1 (resolution 1280x720, pixel 4.86μm x 4.86μm)
- DAVIS 346, (resolution 346x260, 18.5μm x 18.5μm)

All the measurements were conducted with factory sensor settings, default bias values. There are 16 S-curves acquired, one for each neutral density filter, which determine the background irradiation level. The irradiation impulse amplitude is set by controlling LED module current. There are 128 different impulse logarithmically scaled amplitudes used for stimulating the sensors. Both Prophesee sensors were measured without a lens, limiting a region of active pixels to an area of 64x64 pixels around the center of the sensor. This was done in order to make sure, that the bandwidth of the sensor is not overloaded. Davis 346 sensor was measured with optics, which allowed to irradiate a small portion of the sensor. Davis the sensor does not provide the possibility to deactivate the pixels, therefore without the optics the bandwidth of the sensor is overloaded. The acquired S-curves are presented in figures 1 and 2.

### 3.2 S-curves analysis

All the three sensors demonstrate high dynamic range, aver 6 orders of magnitude, see Fig. 1. The S-curves of both Prophesee sensors in the lowest irradiation range are flat, namely the pixels did not produce any event for this background-impulse pairs. Davis 346 is more sensitive at this irradiation levels and produces events with over 60% and probability. Prophesee generation 3.1 sensor starts producing events at irradiation levels of 10 photons per second, unlike the Prophesee, generation 4.1 sensor. Sensors with larger light-sensitive part of the pixels have higher sensitivity at low background irradiation levels.

Vertical displacement the S-curves of the Prophesee sensors in the irradiation range from 20 to 100 photons per second is caused by the noise events sensors produce without stimulation. In case of the Davis 346 sensor the effect is less prominent. The slope of S-curves around 50% event probability, which can be observed from the distance between the samples on an S-curve, grows prominently and steadily from lower irradiation range to higher for Prophesee, generation 3.1 sensor. This means that for higher irradiation levels the influence of noise becomes less significant as S-curves' shape gets more similar to step function. Davis 346's S-curve slopes also grow not as fast and steady as in case of Prophesee, generation 3.1. The slope growth in case of Prophesee generation 4.1 sensor barely noticeable. The latter indicates, noise influence on the temporal contrast detection performance of the sensor is low.

### 3.3 Change detectability

S-curve is a useful performance indicator of an event-based sensor, but enable the comparison of contrast detection performance between area-scan and event-based sensors. Therefore, *change detectability* $\theta$ is introduced. It enables quantitative characterization of event-based image sensors and enabling their comparison to area-scan sensors. It is defined as the mean irradiance $E$ at 50% event probability, $E_{50\%}$, divided by a measure for the "width" of the S-curve, which indicates the amount of noise mixed in with the signal. As a measure for the width of the S-curve, the inverse slope at 50% event probability is taken. This

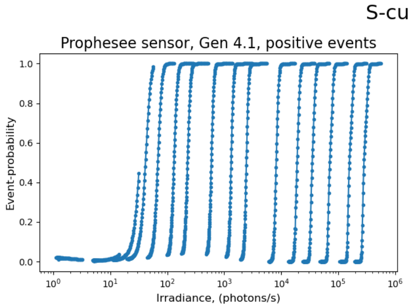


**Figure 1:** S-curves. Left - Prophesee sensor, generation 4.1; Right - Prophesee sensor, generation 3.1.

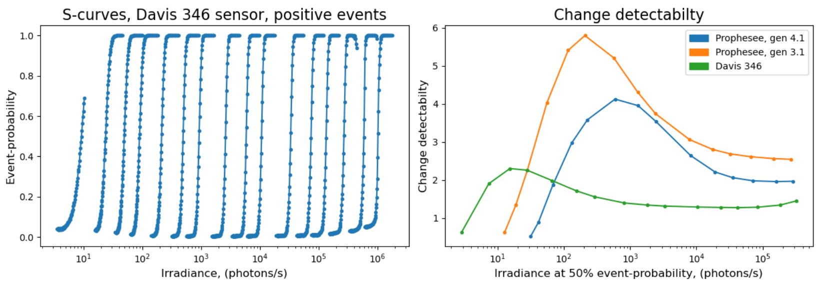


**Figure 2:** Left - S-curves for the Davis 346 sensor; Right - Change detectability comparison for the Prophesee and Davis sensors.

results in the following definition:

$$\theta = E_{50\%} \frac{\mathrm{d}S(E)}{\mathrm{d}E}. \quad (2)$$

The higher $\theta$ is, the lower contrast is required to detect an event.

Change detectability for the three sensors under test was computed and presented in Fig 2. The contrast detection performance of all the sensors in lower irradiation levels is low, but grows with the background irradiation. The peak of change detectability for all the three sensors coincides with the maximum of event noise in absence of stimuli, as if the noise would help the sensor reaching the 50% event-probability threshold. Further growth of the background irradiation levels leads to gradual decrease of the change detectability. That is, for these irradiation levels higher contrasts/impulse amplitudes are required to generate an event. In the highest irradiation levels the metric becomes constant.

## 4 Event-based sensor emulation

In this section we emulate an event-based sensor with an area-scan sensor having a linear response. Namely, we theoretically investigate the event-probability response of an area-scan sensor, which is used for temporal contrast detection.

### 4.1 Basic approach

In order to detect an event, two frames must be taken after each other. Thus the maximum frame rate of an area-based image sensor determines the rate and temporal resolution with which events can be detected. An event can be detected if the difference in the gray values is larger than a given threshold $\tau$. The proper stetting of the threshold depends on the temporal noise. If the threshold is set too low, events will also be generated if there is no generated if there is no gray value change. It is therefore required to compute the probability density function (pdf) of the difference signal with a given noise level.

The goal is to compute the event probability and the resulting S-curves analytically. Therefore it was decided to use a normal distribution. Photon shot noise is Poisson distributed, but the normal distribution is a sufficiently good approximation. Standard industrial image sensors have saturation capacities in the order of 10,000 electrons. The normal distribution is already a good approximation for mean values of just 30 electrons [6]. In the following computations we neglect nonuniformities. This is justified because differences of two only slightly different gray values are subtracted from each other so that the stationary inhomogeneity is canceled out.

It is assumed that the standard deviation of the temporal noise changes with the gray value without assuming a special dependency.

With this flexible approach, it is possible to emulate any area sensor. Therefore two random variables with the distributions $N(\mu_1, \sigma_1)$ and $N(\mu_2, \sigma_2)$ must be subtracted. This results in

$$N(\mu_n, \sigma_n) = \frac{1}{\sqrt{2\pi}\sigma_n} \cdot e^{-\frac{(g-\mu_n)^2}{2\sigma_n^2}} \quad \text{with} \quad n = 1,2 \tag{3}$$

The distribution of the difference signal $\Delta g = g_2 - g_1$ is given by the convolution of the two distribution and also normally distributed with a mean $\Delta\mu = \mu_2 - \mu_1$ and with added variances ($\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$):

$$N_\Delta(\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(\Delta g - \Delta\mu)^2}{2\sigma}} \tag{4}$$

## 4.2 Computation of the S-curve

As it is implemented in the event-based sensor a non-zero threshold $\tau$ is defined. In order to detect an event at a pixel, the difference $\Delta g$ must be larger than $\tau$ in order for the sensor to generate an event. This means that $N_\Delta(\mu, \sigma)$ must be integrated from $\tau$ to $\infty$ resulting in

$$S(\tau, \Delta\mu, \sigma) = \frac{1}{2}\text{erf}(\frac{\tau - \Delta\mu}{\sqrt{2}\sigma}). \tag{5}$$

The Gaussian error function has S-curve shape. There is a consequence which follow from Equation 5: if $\tau = \Delta\mu$, then the event-probability $S(\tau, \Delta\mu, \sigma) = 1/2$. The slope of the S-curve $S(\tau, \Delta\mu, \sigma)$ is given by

$$\frac{\mathrm{d}S(\tau, \Delta\mu, \sigma)}{\mathrm{d}\Delta\mu} = -\frac{e^{-\frac{(\tau-\Delta\mu)^2}{4\sigma}}}{2\sqrt{\pi}\sigma}. \tag{6}$$

The slope of $S(\tau, \Delta\mu, \sigma)$ is a non-linear function. In order to find its maximum we compute the second order derivative of $S(\tau, \Delta\mu, \sigma)$.

$$\frac{\mathrm{d}^2 S(\tau, \Delta\mu, \sigma)}{\mathrm{d}\Delta\mu^2} = -\frac{(\tau - \Delta\mu) \cdot e^{-\frac{(\tau-\Delta\mu)^2}{4\sigma^2}}}{2\sqrt{\pi}\sigma^3}. \tag{7}$$

The right side of the Equation 7 is equal to zero an the point where $\Delta\mu = \tau$. Therefore, the S-curve's maximum values is at $\Delta\mu = \tau$, where the value of the event probability is $1/2$. The slope is independent of the chosen threshold $\tau$ and is equal to $1/(2\sqrt{\pi}\sigma)$.

### 4.3 Signal-to-noise ratio (SNR) and chance detection

Characterization of the conventional image sensors is a well known procedure. With respect to temporal noise the essential parameter is the signal-to-noise ratio, or short SNR as a function of the exposure per pixel in photons $N_p$:

$$\mathrm{SNR}(N_p) = \frac{\mu}{\sigma}, \tag{8}$$

where $\mu$ and $\sigma$ are the mean and standard deviation of the digital output signal.

The SNR can be measured using the measuring and evaluation techniques described by the EMVA standard 1288 using an irradiation series from dark to saturation [5]. For a simple linear image sensor without any noise changing preprocessing, the SNR can be related to the quantum efficiency $\eta$ and the temporal variance of the dark signal $\sigma_d^2$:

$$\mathrm{SNR}(N_p) = \frac{\eta N_p}{\sqrt{\sigma_d^2 + \eta N_p}}, \quad \mathrm{SNR}_{\mathrm{ideal}}(N_p) = \sqrt{N_p} \tag{9}$$

In case of event-based sensor the SNR cannot be defined the same way. The definition proposed by Manakov et al [4] can be used for comparing event-based sensor between each other, but does not establish the relation to area-scan cameras. This can be established using the definition of *change detectability* in eq. 2, because the slope of the S-curve in known from eq. 6:

$$\theta = \frac{\mu_{50\%}}{2\sqrt{\pi}\sigma} = \frac{\mathrm{SNR}(\mu_p)}{2\sqrt{\pi}}, \tag{10}$$

As follows, the contrast detectability $\theta$ is $2\sqrt{\pi} \approx 3.54$ times smaller than the SNR of an are-scan sensor. For a direct comparison with event-based cameras, not the exposure $N_p$ must be used, but the irradiance $E$. In this way the exposure time $t_{\mathrm{exp}}$ is introduced: $N_p = E t_{\mathrm{exp}}$ and the final result is

$$\theta = \frac{\mathrm{SNR}(E t_{\mathrm{exp}})}{2\sqrt{\pi}}. \tag{11}$$

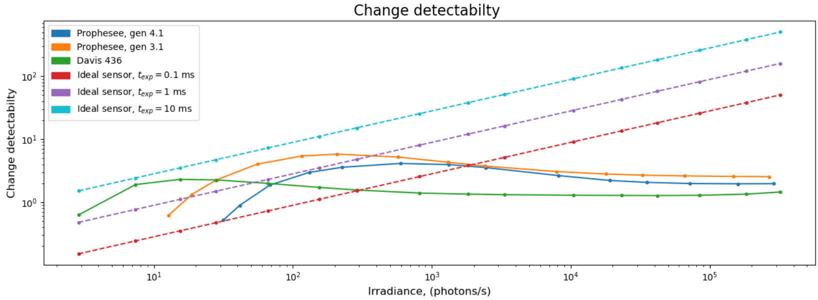


**Figure 3:** Change detectability comparison for event-based and an ideal area-scan sensor.

This can be applied to the SNR of a linear and ideal camera according to eq. 9 and results in

$$\theta(E) = \frac{\eta E t_{\text{exp}}}{2\sqrt{\pi}\sqrt{\sigma_d^2 + \eta E t_{\text{exp}}}}, \quad \theta_{\text{ideal}}(E) = \frac{\sqrt{E t_{\text{exp}}}}{2\sqrt{\pi}} \tag{12}$$

The used exposure time of an area image sensor thus determines which contrast is required to detect an event. In Fig. 3 the change detectability of an ideal area sensor with exposures times of 0.1, 1, and 10 ms is compared with measurements from event-based cameras

## 5 Conclusion and outlook

In this work the change detectability metric was introduced. It enables quantitative characterization of contrast detection performance of event-based cameras. Change detectability metric and the conducted theoretical investigation on the event-probability response of an area-scan sensor, which is used for temporal contrast detection was conducted, establish the comparison link between area-scan and event-based cameras. Moreover, it has been demonstrated that the metric can be calculated for any area-scan sensor with non-linear response, characterized in terms of EMVA 1288 standard. S-curve measurements performed over high dynamic range of irradiance levels was performed for three different event-based sensors. Change detectability for the measurements was calculated and presented together with the S-curve analysis. The theoretical investigations with area-scan image sensors emulating event-based sensors will be complemented in the future by the measurements performed with linear and high dynamic range area-scan sensors.

## References


1. G. Gallego, T. Delbrück, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. J. Davison, J. Conradt, K. Daniilidis, and D. Scaramuzza, "Event-based vision: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 1, pp. 154–180, 2022.
2. C. Posch and D. Matolin, "Sensitivity and uniformity of a 0.18µm CMOS temporal contrast pixel array," in *Circuits and Systems, ISCAS 2011, IEEE International Symposium*, 2011, pp. 1572–1575.
3. A. Manakov and B. Jähne, "Characterization of event-based image sensors in extent of the emva 1288 standard," *Tagungsband "Forum Bildverarbeitung 2020"*, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:269583737
4. A. Manakov, H. Herrmann, and B. Jähne, "Towards integration of event-based cameras into emva 1288 characterization," in *6th European Machine Vision Forum*, Wagingen, 2023.
5. EMVA 1288 Working Group, "EMVA Standard 1288 - standard for characterization of image sensors and cameras, release 4.0," European Machine Vision Association, open standard, 2021.
6. B. Jähne, *Digitale Bildverarbeitung*, 8th ed. Berlin, Heidelberg: Springer, 2024.