Characterization of Event-Based Image Sensors in Extent of the EMVA 1288 Standard

Alkhazur Manakov[1] and Bernd Jähne[2]

[1] Imago Technologies GmbH
Strassheimer Str. 45, 61169 Friedberg
[2] HCI at IWR, Heidelberg University,
Berliner Straße 43, 69120 Heidelberg

Abstract Recent years have seen a steady trend towards faster image sensors with higher resolution. It is well known that images and to a larger extent image sequence contain a lot of redundant information. An areas-scan image sensor, which is not sampled with a constant pixel and frame rate, but which outputs information only if something happens is therefore an interesting alternative. Such sensors are known as event-based or neuromorphic image sensors. Currently, there are several types of event-based image sensors on the market, but no universal concepts available to characterize these image sensors. In this work, we propose the characterisation concepts for neuromorphic sensors in extent of the EMVA standard 1288.

Keywords Sensor characterisation, event-based, neuromorphic

1 Introduction

In the recent years, state-of-the-art image sensors have seen a steady trend towards higher resolution and speed. The trend is driven by the need for faster and higher resolution vision systems in automotive, industrial and other fields. Despite of a significant progress made in the last decades, modern artificial vision systems are still much less effective and robust in solving real-world tasks than their biological counterparts. Even small insects outperform the most

DOI: 10.58895/ksp/1000124383-1 erschienen in:
Forum Bildverarbeitung 2020
DOI: 10.5445/KSP/1000124383 | https://www.ksp.kit.edu/site/books/m/10.58895/ksp/1000124383/

powerful vision systems in such routine tasks as, for instance, real-time perception.

One of the limitations of the human-engineered vision systems is imposed by the image sensors and their principle of operation. Conventional sensors acquire the visual data in form of a series of images, recorded at discrete points of time. Visual data is sampled at a predetermined temporal intervals (frame rate) without any relation to the dynamics of the scene. On top of that, every image contains the data of all the pixels independently from whether this information, or part of it, has been recorded in previous images. This inflates the data rate unnecessarily and fast changes might be missed.

The alternative is the biologically inspired sensors: the dynamic vision sensors that implement the event-driven, frame-free approach. They are often referred to as "event-based" sensors due to their principle of operation. This family of sensors capture and is driven by the transient events in the visual scene, unlike conventional image sensors, that work with artificially created timing and control signals [1]. The latter implies that the control over the acquisition is transferred to single pixel, that handles its own information individually. The output of this sensor is compressed at the sensor level, thus optimizing data transfer, storage, and processing.

Characterisation of the conventional image sensors is a well known problem. The concepts, methodology and techniques for these sensors have been analysed, structured and resulted into the EMVA 1288 characterization standard [2]. These concepts cannot be applied to the event-based sensors. Characterization of event-based sensor is of a great importance, since it provides the means to compare them between each-other, and, most importantly, to conventional image sensors. In this paper, we address the problem of application-oriented characterisation of event-based sensors, establishing a link to EMVA 1288 standard, proposing characterisation techniques and presenting the first results. Dynamic vision sensor shows no response to static images. Therefore, new characterization concepts and procedures need to be developed, which take into account temporal aspect and can be applied to this type of sensors. At the same time, we would like to keep the possibility to compare the performance of the conventional image sensors with and the event-based ones.

2 Related work

2.1 Neuromorphic sensors

Biological retinas have many desirable characteristics, which are lacking in conventional image sensors, thus inspiring and driving the design of neuromorpic vision devices. Many of these advantageous characteristics have been modeled and implemented on silicon. Early development of such devices originated from the biological sciences community. The main purpose of these chips was to provide means for demonstration of neurobiological models and theories. Real-world applications were never the main focus. Therefore, very few of the sensors have been used in practical applications. Circuit complexity, large silicon area, low fill factors, or high noise levels and other factors prevented realistic applications [3], [4]. Recently, the development of practical vision sensors based on biological principles gained an increasing amount of attention and effort.

There is a family of event-based sensors, that encode illuminance in the time domain, namely in the rate of spike "events". The pixels of these devices do not autonomously react to visual events in the scene. Thus, despite of having some advantages against conventional sensors, they fail to achieve redundancy suppression or latency reduction [4]. Large fixed pattern noise, complexity of the digital frame grabber and the big advantage of brighter pixel over the darker ones in allocating the communication bus make "octopus retina" devices [5] impractical for conventional imaging. The pixels of so called "time-to-first spike" imager [6], [7], [8] generate only one spike per frame, static parts of the scene generate spikes at the same time saturating the readout bus.

In dynamic vision sensor pixels are autonomous in detecting light changes in the scene. The gain in terms of temporal resolution with respect to conventional image sensors is dramatic. In addition, such parameters like the dynamic range of the scene greatly profit from the this approach. This type of sensor is very well suited for many machine vision applications including high-speed motion detection and analysis, object tracking, and shape recognition.

The pixel model proposed by Lichtsteiner et al. [9] simulates simplified three-layer retina (Figure 2.1). The circuit consists of a photo-

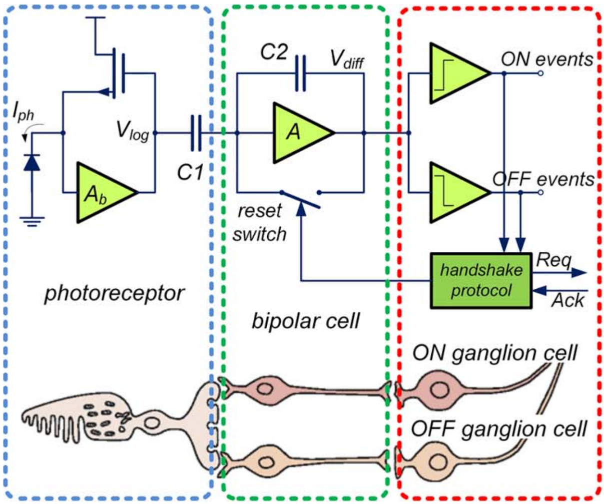

Figure 2.1: Simplified model of a human retina and corresponding event-based pixel circuitry. V_{log} tracking the photocurrent through the photo-receptor. The bipolar cell circuit responds with spike events V_{diff} of different polarity to positive and negative changes of the photocurrent. The ganglion cell circuit monitors the bipolar cells part and transports the spikes to the next processing stage.

receptor front-end, a differencing switched-capacitor amplifier and a comparator-based event generator. The photo-receptor responds logarithmically to irradiance, thus implementing a gain control mechanism that is sensitive to temporal contrast or relative change. The circuitry of the pixel allows to tune for the sensitivity of smaller or larger light changes in the scene. For instance, making the pixel biased to detect brighter-to-darker changes or vice versa. The parameters controlling the setting of the circuitry are called "biases". The pixels independently and asynchronously react to relative changes in intensity, producing sparse, frame-free, event-based output. Upon detection of the relative light intensity change the pixels communicate their state (ON or OFF) to the readout circuitry. The readout and the encoding circuitry encode the coordinated of the pixel, the state and the microsecond resolution time-stamp into an eventpacket. These packets or events can be gathered and analyzed by a visual inspection application.

The relative change events and gray-level image frames are two orthogonal representations of a visual scene. An event carries information about local relative changes, hence encodes all dynamic contents, yet there is no static parts of the scene. The conventional frame-based image is an absolute intensity map at a given point in time. Dynamic information is carried in form of blur. In principle, it is impossible to recreate change events from image frame nor can gray-level images be recreated from the events.

The most recent developments of sensor designs allow to combine the acquisition of static and dynamic information of the scene. Asynchronous time-based image sensor [1], [10] features fully autonomous pixels, that combine a change detector and a conditional exposure measurement circuit. The exposure measurement is initiated when an event is detected. Namely, the measurement starts immediately after the irradiance change of a certain magnitude has been detected by the respective pixel. Another recent approach to combine dynamic and static information into a single pixel is the so-called dynamic and active pixel vision sensor [11]. This pixel combines conventional frame-based sampling of intensity with asynchronous detection of log intensity changes. The advantages of combining the traditional and event-based vision comes at the cost of the capturing redundant output.

2.2 Conventional sensor characterization

Characterization of the conventional image sensors is a well known procedure. There are a number of techniques proposed for characterizing the property of certain sensor. The EMVA standard 1288 measures the mean (μ_y) and variance (σ_y^2) of the digital output signal as a function of the the pixel exposure in photons from dark to saturation [12]. With these measurements and a linear camera model it is possible to determine the signal-to-noise ratio SNR as a function of the exposure per pixel in photons μ_p, neglecting the quantization noise:

$$\mathrm{SNR}(\mu_p) = \frac{\mu_y}{\sigma_y} = \frac{\eta\mu_p}{\sqrt{\sigma_d^2 + \eta\mu_p}}. \tag{2.1}$$

For a linear sensor, SNR depends on the quantum efficiency η and the temporal variance of the dark signal σ_d^2. For a non-linear sensor, the input SNR is the most important quality parameter. It can be computed from the measured output SNR and the slope of the characteristic curve [13]:

$$\mathrm{SNR}_{\mathbf{in}}(\mu_p) = \frac{\mu_p}{\sigma_p} = \frac{\mu_p}{\sigma_y}\frac{\partial\mu_y}{\partial\mu_p} = \frac{\mu_p}{\mu_y}\frac{\partial\mu_y}{\partial\mu_p}\mathrm{SNR}_{\mathrm{out}}. \tag{2.2}$$

3 Event sensor characterization

These procedures are not applicable for the event-based sensors, because the latter are insensitive to static irradiation. Posch et al. [10] have initially addressed the problem of event-based sensor characterization. In their work, they have proposed a test method that allows simultaneously evaluating the main performance parameters and check how well the predictions from theoretical considerations agree with the performance of the sensor. We adapt the ideas proposed by Posch at al. [1] for the application-oriented characterization of the event-based sensors, in context of the EMVA 1288 characterization standard , described in the last section.

3.1 Properties

Sensitivity to small temporal contrasts, the response relation to the ON/OFF-biases 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 as described in Sect. 2.2.

The sensitivity to the event-based sensor to the contrast is controlled by the ON- and OFF-biases. The biases might be set higher for making the sensor insensitive to small temporal contrast in the scene. The relation between the biases and the contrast threshold might be non-linear, depending on the circuitry of the pixel.

In the event-based sensor, the pixels react autonomously and asynchronously to the light transients in the scene. Therefore, the important characteristic of the sensor is *response uniformity*. In other words, how a single-pixel performance translates to the behaviour of the whole array. Due to production imperfection and tolerances the photo-sensors, circuitry will inevitably have variations in how pixel react to the same stimulus. This property of the event-based sensor corresponds to well-known nonuniformity property of the conventional image sensors.

3.2 Measurement procedure

The simplest way of experimentally determining the irradiation contrast $\Delta\mu_p$ necessary for generating one event for given mean irradiance level μ_p and event threshold settings is gradually increase the stimulus step until an event is generated. The stimulus' amplitude must initially be below the response threshold. It should also be fast enough, namely to have the rise time exceeding the bandwidth of the circuit under test. The minimal found stimulus amplitude always results in an event response when applied. In an ideal noise-free world, this would be the case and this method would be applicable.

In the real world conditions, the very same pixel will react differently to the same stimulus. Therefore, Posch et al. [10] propose to operate with "event probability" instead. 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 response thresholds remain unchanged. Plotting the "event probability" vs. stimulus amplitude, in an ideal noise-free world, would result in a step function. In reality, such curve would have an "S"-shape. Fitting the "S"-curve with a Gaussian error function, with the mean at the $M/N = 50\%$ event probability point, indicates the location of the barely sensible contrast.

Thus the irradiation contrast $\Delta\mu_p$ that produces an event probability of 50% corresponds to a temporal standard deviation of one sigma. In this way the input SNR can be measured as the ratio of μ_p and $\Delta\mu_p$ directly as a function of the mean irradiation from dark to saturation of the sensor. A linear response or the measurement of a characteristic curve is not required. This procedure corresponds to the new upcoming release of the EMVA standard 1288 for cameras with a non-linear response (Release 4.0 General, see [13]). The measurements are to be conducted on the entire array (or a selected area-of-interest) to ensure statistically significant conclusions, in the same way as for conventional cameras with the EMVA standard 1288.

Measurement procedure:

1. Choose ranges for background irradiance $[\mu_p; \mu_p^{max}]$ bias levels for ON/OFF events, stimulus amplitude $[\Delta\mu_p; \Delta\mu_p^{max}]$ and respective increments sizes.
2. For every background irradiance level, bias pairs and stimulus in the chosen range perform steps 3 and 4.
3. Apply stimulus and reset the selected pixels N times.
4. Count event responses M compute per-pixel probability P and for every of the selected pixels.

The data acquired this way from the whole sensor or part of it is sufficient to recover the contrast sensitivity, the response uniformity and the contrast threshold dependence on the bias settings.

4 First results

The experimental setup used for the measurements consists of an integrating sphere, background irradiance LED-lamp, contrast generation LED-lamp, the camera under test and the control PC. The data has been acquired and processed according to the procedure described above. The Figure 4.1 presents the event probability P dependence on the stimulus $(\Delta\mu_p/\mu_p)$ for different background irradiation levels. The data is acquired on the area of 128x128 pixels with

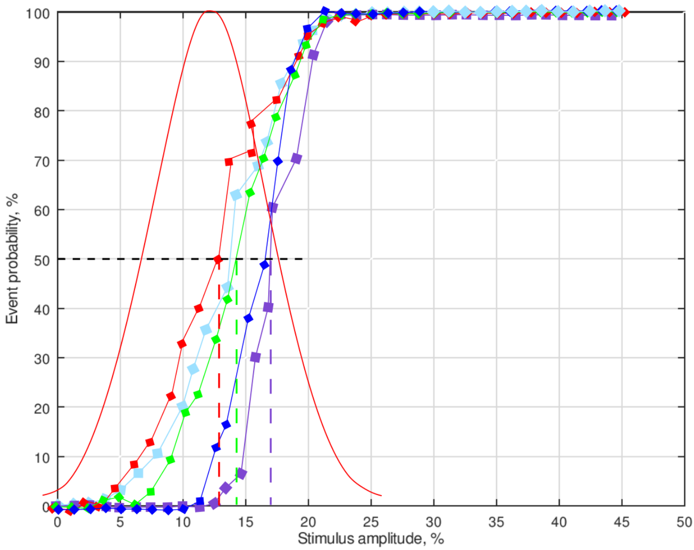

Figure 4.1: Event probability depending on the stimulus amplitude measured at one pixel. The ″S″-curves are acquired for different background irradiation levels. Higher irradiance levels correspond to steeper curves. Gaussian fit to the red ″S″-curve. The black dashed line indicates the even probability point. The vertical dashed lines indicate contrast thresholds for the corresponding curves.

the biases set 100 milliVolts. All the experiments were performed with VisionCam EB featuring Prophesee PPS3MVCD sensor.

The mean point of the Gaussian (50% event probability point) indicated ideal minimum contrast for event generation at this light level and for the chosen bias settings (Figure 4.1). In conventional sensors, this corresponds to a irradiation change of σ_p (Section 2.2). This means that the proposed method is able to measure the input SNR as a function of the irradiation. The response relation to the ON/OFF-biases settings can be extracted from the family of ″S″-curves. The contrast threshold grows with the background irradiance levels as represented by in Figure 4.1. The standard deviation of the fitted Gaussian corresponds to the root-mean-square noise of the pixel.

5 Conclusions

In this work, we have adapted the concepts and methods developed by Posch et al. [10] to the application-oriented characterization of the event-based sensors, in terms of the EMVA 1288 characterizations standard. We have established the link between the properties of conventional and event-based sensors. Preliminary non-calibrated test measurements show that the measurement of the event probability is the correct way to measure the temporal noise of an event-based sensor and that this can be used to measured the SNR as a function of the irradiation. In this way event-based and conventional sensors can be compared directly. The analysis of the nonuniformities of event-based sensors requires further research.

References

1. C. Posch, D. Matolin, and R. Wohlgenannt, "A qvga 143 db dynamic range asynchronous address-event pwm dynamic image sensor with lossless pixel-level video compression," in *IEEE Conference of Solid-State Circuits*, 2010, pp. 400–401.
2. B. Jähne, "EMVA 1288 standard for machine vision – objective specification of vital camera data," *Optik & Photonik*, vol. 5, pp. 53–54, 2010.
3. S. C. Liu and T. Delbruck, "Neuromorphic sensory systems," *Current Opinion Neurobiology*, vol. 20, pp. 288–295, 2010.
4. D. Chen, D. Matolin, A. Bermark, and C. Posch, "Pulse modulation imaging – review and performance analysis," *IEEE Trans. Biomedical Circuits Systems in Cognitive Sciences*, vol. 5, no. 1, pp. 64–82, 2011.
5. E. Culurciello, R. Etienne-Cummings, and K. Boahen, "A biomorphic digital image sensor," *IEEE Journal Solid State Circuits*, vol. 38, no. 2, pp. 281–294, 2011.
6. Q. Luo and J. Harris, "A time-based cmos image sensor," in *IEEE International Symposium Circuits Systems*, vol. 4, 2004, pp. 840–843.

7. X. Qi, X. Guo, and J. Harris, "A time-to-first spike cmos imager," in *IEEE International Symposium Circuits Systems*, vol. 4, 2004.
8. C. Shoushun and A. Bermak, "Arbitrated time-to-first spike cmos image sensor with on-chip histogram equalization," vol. 15, no. 3, pp. 346–357, 2007.
9. P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128 x 128 120db 30mw asynchronous vision sensor that responds to relative intensity change," in *2006 IEEE International Solid State Circuits Conference - Digest of Technical Papers*, 2006, pp. 2060–2069.
10. 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.
11. R. Berner, C. Brandli, M. Yang, S.-C. Liu, and T. Delbruck, "A 240 x 180 10 mw 12 us latency sparse-output vision sensor for mobile applications," in *Symposium on Very Large Scale Integrated Systems*, Kyoto, Japan, 2013, pp. C186–C187.
12. EMVA 1288 Working Group, "EMVA Standard 1288 - standard for characterization of image sensors and cameras, release 3.1," European Machine Vision Association, open standard, 2016.
13. B. Jähne, "Release 4 of the EMVA 1288 standard: Adaption and extension to modern image sensors," *this volume*, 2020.