applied optics

Characterization setup for event-based imagers applied to modulated light signal detection

DAMIEN JOUBERT,[1,2,*] MATHIEU HÉBERT,[3] HUBERT KONIK,[3] AND CHRISTOPHE LAVERGNE[1]

[1]*Groupe Renault, DEA-SAR, F-78280, Guyancourt, France*
[2]*Université Clermont Auvergne, F-63000 Clermont-Ferrand, France*
[3]*Univ. Lyon, UJM-Saint-Etienne, CNRS, Institut d' Optique Graduate School, Laboratoire Hubert Curien UMR 5516, F-42023, Saint-Etienne, France*
Corresponding author: joubert.damien@yahoo.fr

Event-based cameras bring new perspectives for perception systems by making them faster, smarter, and less energy-consuming. While they are spreading into many application domains, new algorithms are designed to process the data they provide, and new databases are needed to validate and train them. Simulations are an efficient way to increase databases, as they give direct access to ground truth for applications such as target detection or depth estimation, provided the simulation models used are as close as possible to the physical reality. The model should also be designed generically enough to be applicable to different kinds of event-based imagers. The characterization setup proposed in this paper aims at measuring the main characteristics of the dynamic vision sensor in each pixel under outdoor lighting conditions. A simulation model of the imager's response can be generated using the measured characteristics. These measurements are used to estimate the robustness of an algorithm to detect modulated light signals exploiting event-based data. An improvement is then provided so this algorithm can detect higher frequencies.   © 2019 Optical Society of America

https://doi.org/10.1364/AO.58.001305

1. INTRODUCTION

The tremendous rise of advanced driver assistance systems (ADASs) in the automotive industry and the prospect of autonomous vehicles [1] bring new constraints to cameras. A lot of work must be done to build smarter imagers in order to increase perception performance. Part of the scientific community concerned tries to build systems whose architecture is inspired by biologic perception systems according to the so-called neuromorphical approach [2]. But industrialized perception systems in automotive applications are not yet able to compete with the human perception function [3]. Even being more powerful, for example, fitted with higher computational strength, cameras, and computing units in a car still lag behind human perception. The neuromorphic approach assumes that principles used to build state-of-the-art perception systems have to be redesigned and that biologic systems can bring new relevant principles to build better systems [4,5]. These new systems should not necessarily mimic the human system. Many species are much more efficient in the detection of targets, in particular in dark environments. Each part of the perception chain has to be reinvented: optics, detectors, computing units, actuators, etc. The recent successes achieved in deep learning, for example, thanks to convolutional neural networks (CNNs) [6] are not completely inspired by nature. The use of "neural"

mainly stands for the hierarchical architecture and the use of learned Gabor-like filters, acting as features extractors, similar to cat's eye behavior [7]. Neurons in a CNN have a very different behavior from biologic ones. A key difference between CNN and natural neurons lies in the treatment of temporal information: human neurons are asynchronous, while CNN neurons are synchronous. These new kinds of data can be used to design more efficient vision systems. For example, Ref. [8] proposed an architecture where both synchronous and asynchronous information is used to drive a robot following its target. Moreover, the perception function of autonomous cars, prior to being validated, should be characterized step-by-step according to the strong interactions between scenes, optics, imagers, and algorithms. The validation could be obtained by testing a huge number of scenarios, for example, by driving millions of kilometers. This explains why the size of databases is increasing, but this method is not reasonable, because as systems evolve rapidly, a system would not be characterized and tested before the next generation is available. Simulation could solve this issue by building a functional sensor model, used in rising simulation tools [9] and databases [10–12]. In order to build a simulation model of an artificial retina, several characteristics of the sensor have to be identified and measured through a dedicated setup. The estimation must be done in

each pixel to guarantee a sufficient level of the model's robustness. To illustrate how characterization can help validation, we propose to use it in order to estimate how robust a vision system designed to detect modulated light signals is. This is a current issue in automotive applications because of the LED flickering in some road signs.

We present first how neuromorphic imaging devices actually work, also identifying what are the key parameters in order to build a simplified simulation model (Section 2). Afterward, we describe how we plan to measure these characteristics and how they influence the detection of the modulated light signal (Section 3). Finally, each relevant parameter, as well as the performance of the modulated light detection, is measured (Section 4). Section 5 draws the conclusions.

2. DYNAMIC VISION SENSOR

Even acquiring data from many RADARs, LIDARs, cameras, and other sensors, a human driver is still required for autonomous vehicles. The high redundancy of the information in these systems underlines the fact that the information is not properly obtained from the scene. We will try to understand how a strong part of the scientific community tries to add some intelligence in the early part of the perception to build smarter systems.

A. Neuromorphic Systems

One key characteristic of biological systems is that information is asynchronously carried, while conventional human-made systems are almost all synchronous (processor clock, camera frame rate, etc.). Neuromorphic engineering has already met several successes, since it is now possible to design very large-scale integrated analog chips [13] such as imagers inspired from the human vision system [14–16], artificial cochlea [17], and artificial computing systems inspired by the human brain [18,19], which can be used to merge information from both the artificial retina and cochlea [20]. The human retina seems to be sensitive to temporal radiance shifts [21], a behavior that the neuromorphic community started to copy in the 1990s [15]. State-of-the-art artificial retinas are the Dynamic and Active Pixel Vision Sensor (DAVIS) [22], inspired by the dynamic vision sensor (DVS) [23], and also the asynchronous time-based image sensor (ATIS) [24]. Now new artificial retinas are designed by companies such as Samsung [25]. Other sensors inspired by the DAVIS have been developed and improved, using new manufacturing processes like microlenses in the SDAVIS192 [26], color-sensitive pixels [27] with color filters and new pixel architecture like BackSide Illuminated (BSI) pixels [28]. Both DAVIS and ATIS allow absolute irradiance measurement in each pixel (active pixel system), as well as detection of temporal irradiance variations in the focal plane (dynamic vision system). One pixel belonging to ATIS or DAVIS can be seen as a pair of neighbor pixels, one APS and one DVS. According to the architecture, APS and DVS parts are more or less interdependent. Each DVS pixel asynchronously detects the change of irradiance and transfers its position and the absolute time of detection out of the imager with microsecond precision. Since DVS delivers asynchronous data, the treatment of this information must rely on a dedicated architecture different from the

conventional column and row readout architecture implemented in a global or rolling shutter camera. A special bus using the address event representation (AER) principle [29] has been mainly used for artificial retinas [30], and this architecture is still improving to deal with higher bandwidths [25]. For the DAVIS, two arbiters, one for the rows and one for the columns, allow the pixel to communicate an event, timestamp the information, and then reset the pixel, which is said to be "spiking" an event. The main task for the arbiter is to resolve collisions between events generated at the same time, for example, by making a pixel wait before being allowed to spike. The architecture of the AER arbiter directly manages the time precision. As the spatial resolution of the imager increases, more events are generated at the same time and the time precision may be decreased [31], an issue that requires rethinking the arbiter's architecture [25].

Hereafter, we focus more specifically on the DVS part, which detects asynchronously a temporal variation of light flux collected by its photodiode. The APS part, which collects the flux, measures its absolute value and generates a gray-level value, and has already been characterized and standardized [32]. In the DAVIS, the tension at the pixel's output is converted into a gray-level value through an analog-to-digital convertor (ADC). In the ATIS sensor, the APS part does not use an ADC. It asynchronously waits photodiode tension to cross two thresholds. Each time the tension crosses a threshold, it generates an event [33,34]. The gray level is then computed as the time difference between the two corresponding events. This process is more efficient to measure high irradiance levels as encountered in automotive scenarios, where irradiance ranges between tens of milliluxes and tens of thousands off luxes. Then, the ATIS sensor, at first sight, seems to be the more efficient to fit automotive constraints.

B. DVS Pixel

One of the key achievements in the design of the DVS pixel was presented in [23] and is summarized in Fig. 1. The behavior of the pixel can be described in four steps:

1. The current generated by the photodiode is logarithmically converted into a tension, thanks to a transimpedance assembly, whose architecture is tuned in order to increase the bandwidth [35].

2. An emitter-follower circuit arrangement buffers the tension to keep the first step isolated.

3. A differential amplifier circuit with two capacitances amplifies the difference between the reset tension and the output of the buffer.
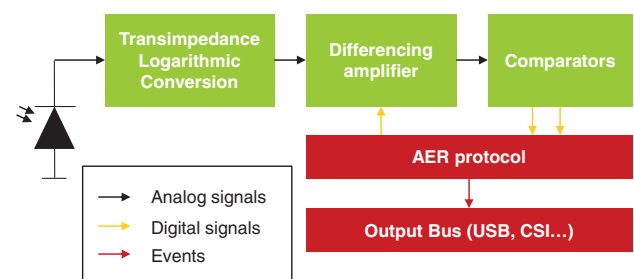

Fig. 1. Main blocks of the DVS architecture.

4. Two comparators threshold the amplifier's output to detect positive or negative variations.

The tension V_{dif}, at the output of the differential amplifier, is related to the variation of current I_{ph} through a transimpedance assembly, thanks to the relation

$$V_{\text{dif}}(t + \delta_t) = -\frac{C_2}{C_1} \frac{U_T \kappa_{sf}}{\kappa_{fb}} \ln(I_{ph}(t + \delta_t)), \qquad \textbf{(1)}$$

where U_T is the thermal tension, defined by $U_T = \frac{kT}{q}$, where k is Boltzman's constant, T the absolute temperature in Kelvin, and q the electron's elementary charge. V_{dif} is then compared to $V_{b,\text{on}}$ and $V_{b,\text{off}}$, the two offset tensions of the comparators allowing the detection of positive and negative variations. The correlated photodiode current thresholds according to the parameters of the pixel are given by [23]. When a threshold is reached, an event is generated. If the positive (respectively, negative) threshold is crossed, a positive (respectively, negative) event will be generated. The event is produced by the AER bus, whose architecture can vary according to the generation of an artificial retina sensor or, more generally, the neuromorphic device [2]. Consider the DVS128 retina [23]: when a threshold is reached, a request signal is sent to a first-row arbiter. Different strategies can be implemented to process the request, and most of the time the first arrived signal is treated. When the row requested to output the spike is allowed, the event is time-stamped, and a new arbiter process is applied to the columns. When the column is allowed to spike, the event corresponding to the last pixel, corresponding to one couple of an allowed row and column, is sent to the AER bus with the four needed information elements: row, column, polarity, and timestamp. The differential amplifier is also reset to allow new detection. This global process is fully described in [36]. Some retinas, as described in [24], add another transistor to reset the output of the differential amplifier. The AER arbiter reset transistor M_r works in parallel with M_s, the reset transistor of the AER protocol, and M_r allows the user to disable some part of the imager if needed. This functionality will be useful in Section 3, when the response of only one group of pixel behavior will be studied.

C. DVS Pixel Parameters

The global latency to emit an event from detection to transmission in the AER bus depends on the latency of the pixel, whose bandwidth has already been studied in [16,23,24,37]. It also depends on the time the arbiter solves AER access request and the delay to reset the pixel. This latter is commonly called the *refractory period*. Thereafter, δ_{pix} will denote the pixel latency, δ_{AER} the AER bus latency, and δ_{ref} the refractory period.

The differential amplifier, as well as the two comparators of the reset AER handshake protocol, depends on the physical characteristics of the transistors. These characteristics vary from one pixel to another one due to the manufacturing of the imaging system [23]. In Eq. (1), the tension before the thresholds, as well as the threshold operation, depends on these characteristics; therefore, the sensitivity varies from one pixel to another one.

The response of a pixel, characterized by its set of physical parameters, is a function of the temporal flux contrast collected by its photodiode. The current crossing the photodiode is

sensitive to the dark noise, thereby on the background irradiance. This noise mostly bounds the dynamic range in case of low irradiance, because the pixel sensitivity is strongly disturbed when the photodiode current is dominated by the dark noise. The global response of the pixel is also influenced by its latency. Considering a given light stimulus generating an event, if this stimulus is submitted several times to the same pixel, the measured latency varies. The latency dispersion may be characterized by its standard deviation, most commonly called the *jitter* [23]. In the same way as sensitivity distribution, jitter and latency are not uniform over the imager's plane. The nonuniformity latency will also be measured. After detecting an irradiance change, the pixel is not able to react again during a short delay, called the refractory period. Then, because of the temporal fluctuations of the noise, and because of the global activity of the imager, some pixels can generate positive or negative events, even when in the absence of temporal flux contrast is collected. These parasitic events constitute a temporal background noise [23,38] (BGN), which downgrades the useful stream of events used for detection and classification of the observed light signal. For example, if a pixel spikes because of the BGN and shortly after is presented a temporal flux contrast, its refractory period will prevent it from spiking. The BGN downgrades the quantification of the light signal in the event domain; thus an analog measure such as signal-to-noise (SNR) can be defined [37]. Considering all of these characteristics, the following indicators will be used to characterize asynchronous dynamic vision sensors:

- the latency of the pixel δ_{pix} in µs and its temporal σ_{pix} and spatial Σ_{pix} deviations in percentage;
- the latency of the AER protocol δ_{aer} and its temporal σ_{aer} and spatial Σ_{aer} deviation;
- the sensitivity in contrast detection c and its spatial deviation, expressed in %, also called the field pattern noise, FPN;
- the refractory period T_r;
- the BGN.

In this paper, we propose to measure these characteristics for each DVS pixel in order to build a simulation model and to improve our understanding of how modulated light signal detection can be achieved. The measurement is mainly inspired by previous works.

D. DVS Characterization

Several methods have already been proposed to measure these DVS characteristics. Recently, a deep study concerning the impact of the temperature on BGN and sensitivity has been reported in [38]. In this work, it is explained how leakage current in the reset transistor generates parasitic activity. The noise is characterized using the frequency of noise events on each pixel, which can be used to estimate the parasitic photocurrent quantum efficiency. It is also shown how the DVS pixel has to be modified in order to keep the same sensitivity independently of temperature. The definition of noise in DVS sensors is also tackled when it comes to defining and characterizing the sensor's sensitivity. The work by [37] provides a good review of how sensitivity has been characterized so far. Mechanical characterization setups [23,39,40] have been used to measure

the sensitivity, its deviation, and how it influences the SNR of the sensor. For example, in [23], a white disk turning around its center with a black line printed on one radius stimulates every pixel at each revolution with a high contrast. In that way, the same stimulus triggers pixels, and repeating the operation enables sensitivity field pattern noise (FPN) measurement. The repeatability and reproducibility of moving mechanical setups has been improved in the work by [16] with a thin-film transistor monitor, but the difficulty of precisely measuring the photometry of the experiment discouraged us from building a mechanical-based characterization setup. More recent works include the use of integrating sphere [28,37,38,41], as advised by standards to characterize gray-level imager [32]. This setup gives access to the photometry of the light stimulus [42], whose temporal characteristics can also be known precisely (around a few nanoseconds) if the light source is composed with LEDs [23,37,43]. The temporal profile of the light signal depends how its sensitivity is characterized. In [37], sensitivity is correlated with the SNR measure. The temporal profile is a sinusoidal wave, and the peak-to-peak value (called TCON) is quantified by the DVS sensor and therefore depends on the sensitivity threshold:

$$\theta_{\mathrm{ON}} = \frac{\ln(\mathrm{TCON_{ON}})}{s_{\mathrm{ON}} - n_{\mathrm{ON}}}, \tag{2}$$

where s_{ON} (n_{ON}) is the number of events generated by the signal (noise). As the light signal and its derivative in the temporal domain are not constant and continuous over time, the kind of stimulus covers a large number of real-world conditions. However, since the tension in the comparator V_{dif} depends on the constant time of the transimpedance assembly, sensitivity will depend on the frequency of the sinusoidal wave. In the work of [43], the sensitivity is seen as the probability of generating an event according to the contrast. The setup proposed, however, also uses LEDs directly in front of the imager's surface, which makes it hard for us to estimate how much sensitivity uniformity depends on the light signal. The FPN of the sensitivity across pixels is defined as the root mean square of the measured contrast threshold. Finally, jitter and latency are often measured by observing the response of the imager to a high-contrast light signal [23,37] by computing the standard deviation of the latency and counting the number of events generated during the sequence. DVS sensor characterization is still an open question, and new metrics are provided to better model these sensors such as modulated transfer function. A summary of some characterization works is proposed in Table 1.

Here, we propose a full characterization protocol, including the measurement of main parameters of the sensor for each pixel, as well as the ability of the sensor to detect a modulated square wave light signal when using a given detection algorithm. Each measurement will depend on ambient temperature and offset irradiance, which increases the space of characterization parameters dramatically. The following measurements are performed under laboratory temperatures (around 20°C). The quantum efficiency is not measured in this work, and the size of the transistor, the fill factor, and its complexity in terms of number of transistors have all been provided by the manufacturer.

3. METHOD

In order to build a DVS model independent of the retina's architecture, one must characterize each pixel in real-world conditions, which is permitted by the characterization setup we propose. The temporal profile of the light stimulus must be precisely known, which is almost impossible with mechanical actuators. We chose to interact only with the light source. This section explains how our constraints obliged us to design the light stimulus, and how we used it to measure sensitivity, latency, and other main characteristics of complementary metal-oxide semiconductor artificial retinas. Particular attention is paid to correctly exploiting the stream of events after each stimulus by trying to separate useful events from the noise.

A. Light Stimulus

Artificial retinas are sensitive to temporal variations of irradiance. Therefore it is important to master the light stimulus used for characterization.

1. Light Signal

LEDs are sometimes used to generate the light stimulus [23,43] in the characterization of artificial retinas because the latency between the command current and the response in flux of an LED, around a few nanoseconds, can be considered as instantaneous in our applications. Authors of [43] directly put LEDs in front of the imager, while authors of [38,41] use an integrating sphere. This latter is conventionally used to characterize gray-level measurement imagers and their FPN [32]. It seems suitable to use this instrument with asynchronous imagers. Because of the dynamically temporal response of the camera, the light signal is not constant over time. The use of an integrating sphere does not dramatically modify the light signal between its entrance and output. The time constant of our sphere is around 16 ns, according to the relationship given by the manufacturer [42]:

$$\tau = -\frac{2}{3} \frac{D_s}{c} \frac{1}{\ln(\rho_{\mathrm{moy}})}, \tag{3}$$

where D_s is the diameter of the the integrating sphere, $c = 3.10^6$ the celerity of the light in m/s, and ρ_{moy} the mean

Table 1.   DVS Sensor Characterization According to Experimental Setup (m for Mechanical, s for Integrating Sphere), to Tested Environment Conditions (i for Irradiance, t for Temperature), and to Measured Deviation Between Pixels f for FPN

Parameter	This Work	DVS128 [23]	SDAVIS192 [37]	ATIS [24,43]	Nozaki [38]
BGN	Event frequency (**i**, **f**, **s**)	No	SNR (**i**, **s**)	Event frequency	Leak activity (**i**)
Sensitivity	S curve (**s**, **i**, **f**)	TCON (**m**, **f**)	TCON (**s**, **i**, **f**)	S curve (**i**, **f**)	TCON (**u**, **t**)
Latency and jitter	Yes (**i**, **f**)	Yes (**i**)	Yes	Yes (**i**)	No
QE	No	No	Yes	No	Yes

spectral reflectance of the sphere's internal coating. The main property of the integrating sphere is to produce a uniform irradiance on its sphere surface, and to scatter in every point the same radiance in every direction (therefore, towards the imager). Our goal is to present the same stimulus on each pixel of the imager in order to compare their responses. Thus, the irradiance must be as uniform as possible at the imager's plane. For conventional gray-level imagers, a 3% mean square error (MSE) in irradiance is accepted [32], but the acceptable error still needs to be determined for automotive applications. The temporal contrast C could be defined as a function of the current of the photodiode when changing from I_{ph1} to I_{ph2} [24], but as we cannot access these currents, we chose to define C using the corresponding irradiance at the surface of the photoreceptor; then

$$C = \frac{E_2 - E_1}{E_1}, \tag{4}$$

where the irradiance at the surface of the photoreceptor varies from E_1 to E_2. Instead of using a sinusoidal [37] or triangular [38] wave, a two-state signal is generated in our experimental setup, even if this kind of temporal light profile is uncommon in natural scenes. Two different light sources are used, each with its own driver. The LED source used to generate E_1 is called the *high LED*, and the LED source charged to generate E_2 is called the *offset LED*. The relation between the temporal contrast and the probability of spiking is a step function centered on the contrast threshold. In reality, as explained by the work of [43], the response is rather an S-like shape, often due to the noise, called an S curve. The sensitivity of a pixel depends on the scene irradiance [43], and the FPN is not uniform over irradiance's dynamic range. The deviation is higher around high irradiances, and the FPN seems to decrease under these conditions, so the uniformity of the light stimulus should be as low as possible to avoid adding an extra deviation in the computation of the sensitivity's FPN. The target is a 1% uniformity error in irradiance. The setup is illustrated in Fig. 2.

In this configuration, the irradiance error in comparison with the irradiance at the intersection between the output axis and the yellow parallel plane (i.e., the imager plane) is given by the relationship [44]:

$$\text{error}(P) = \frac{1}{2\sin^2(\alpha)}(1 - \gamma(\alpha)) - 1, \tag{5}$$

where

$$\gamma(\alpha) = \frac{1 - \sin^2(\alpha)}{\sqrt{\sin^4(\theta) + 2\sin^2(\theta)\cos^2(\theta)(1 - \tan^2(\alpha)) + 1}}, \tag{6}$$

where θ is the half-angle under which the output of the sphere watches the point P, and α is the half-angle under which the center of the imager watches the aperture of the sphere. In the experiments, the distance between the imager and the aperture of the sphere is $d = 75$ mm $\pm$ 1 mm. The irradiance uniformity error, computed according to Eq. (5), was found to be less than 1%.

The irradiance, computed on the plane at the meeting point P with the optical axis of the sphere, at a distance d of the sphere, is given by the following equation:

$$E(P) = \frac{F}{4\pi R^2} \frac{\rho}{1 - \rho(1 - f)} \frac{1}{1 + \left(\frac{2d}{D}\right)^2}, \tag{7}$$

where R is the radius of the sphere, D, the diameter of the output port, ρ, the average spectral reflectance of the sphere coating, f, the ratio of the cumulative sum of aperture areas in the sphere and the sphere whole area, and F, the flux of the light source.

According to Eq. (7), the range of the flux F and the distance d needed to obtain the expected target range of the irradiance dynamic can be computed. Let us now see how to generate the supply current of the LEDs.

2. Current Signal

In order to obtain low irradiance levels, we add density filters between the LEDs and the sphere. Obtaining high irradiances is more difficult because the LEDs radiating high flux also dissipate heat. Temperature deviations cause two main drawbacks: a spectral shift and a risk of damaging the LED. In our case, a light source (SIWH-SC211-WIR200) from Intelligent LED Solutions (ILS) is used, which emits 440 lm when supplied by a 300 mA current. The current generator is an IPSC4 Smartek strobe controller, well suited for generating high currents at high frequencies with different duty cycles. Three density filters of density 10, 1, and 0.1 were used to characterize the imager under four irradiance levels, from luxes to kiloluxes. The Smartek strobe controller can only generate a signal between zero and a given current under a given tension. Interactions between high LED and offset LED are summarized in Fig. 3. Since the flux emitted by one set of LEDs is proportional to its current, the contrast can be directly expressed as a function of the two supplying currents. For example, if the current of the
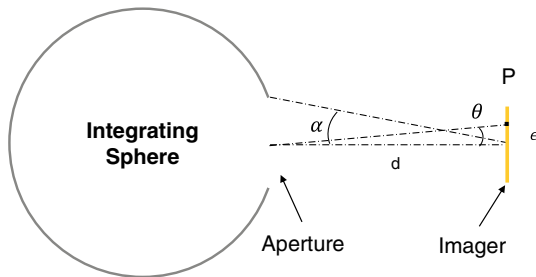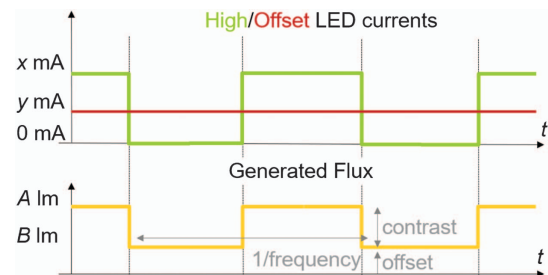

Fig. 2. Experimental setup used to compute the irradiance of a point located in a plane parallel and distant by d from the aperture of the sphere, at a distance ϵ from its optical axis.

Fig. 3. Chronograph of the electric signals supplying the LEDs in order to generate irradiance contrasts.

offset LED is 200 mA, the ratio of contrast is 0.5%/mA as a function of the high LED's current. As explained in Section 2, the AER bus takes some time to treat the response of all the pixels, a time that we want to estimate in order to optimize the stimulus frequency. According to our measurements, the time between the stimulus and the output of the last correlated event sent to the AER bus is 8 ms when the whole imager is flashed, which represents 72,960 pixels. With this flow of 9 Mevents/s, we can accept a stimulus frequency of 50 Hz.

As we are now able to control the light stimulus by modifying LED currents, we must check if the signal stays constant over the experiment to guarantee experiment repeatability.

3. Light Stimulus Measurement

In order to know precisely the flux emitted by the two sets of LEDs, and therefore the irradiance in the imager plane, a Gershun's tube is used to collect the flux. This latter is placed near the imager, facing the aperture of the integrating sphere, and is connected to a calibrated spectrometer with an optical fiber. The flux collected only comes from the sphere, and its value is proportional to the flux emitted by the LEDs with a given spectral ratio. The integration time of the spectrophotometer is adjusted according to the irradiance to obtain the best SNR. In our experiment, the integration time reached hundreds of milliseconds for low irradiances, which was too long to observe the response of the retina in front of a 50 Hz frequency stimulus. We therefore decided to perform spectral measurements of the LEDs under continuous supply and not under pulsed supply. We can assume that the spectral distribution under pulsed supply is similar to the one under continuous supply, provided the LED junction temperature is constant, which is the case as the alternative current mode generates less heat when the current duty cycle is low. We consider that there is no latency between the startup of the current signal in the LED and the startup of the emission of the light. However, a latency comes from the Smartek strobe controller. This latter is measured and is around 4 µs for a 200 mA supply and 8 µs for a 350 mA supply. As the latency depends on the current provided, it is measured each time the current of the LEDs is modified and removed to the measured latency of the pixel.

In order to guarantee that the offset flux is constant over the experimental process, we measure it in each test. Periodically, the offset and high fluxes are measured and currents adapted with a 3% tolerance. This corrector is also used when the ratio to convert current into flux is not the same between the two sets of LEDs. It adjusts the two current values in order to get identical ratios. High LEDs flash at a fixed frequency f during 1 min.

Now that the light stimulus is specified, let us begin the characterization of the retina. We will investigate protocols to measure some of different parameters of the retinas illustrated in the introduction. In the following part, we will, most of the time, only consider positive events. The same experiments could be done for negative events by using a similar characterization method as for positive events. The goal of this work is to measure the performance of a DVS sensor provided its parameters (thus, no deeper experiments) have been realized to study the influence of one or several settings.

B. Sensitivity

In order to characterize the sensitivity and its distribution over the imager, we compute S curves for each pixel as proposed [24,43]. Considering a given stimulus contrast, the probability that a pixel outputs an event is not spatially and temporally uniform. The spatial nonuniformity results from the sensitivity's FPN, while the temporal nonuniformity results from temporal noise in the pixel's electronics. The spatial deviation in sensitivity reaches a maximum when the probability to output an event is 0.5. Hence, sensitivity FPN is defined as the distribution of the needed contrast to produce an event with a 0.5 probability. Probability 0.5 corresponds to the worst case for sensitivity dispersion: in the extreme cases where the contrast is very high, near 100%, or very low, near 0%, all the pixels are expected to respond in the same way; therefore, no deviation is observed over the region of interest (ROI). However, there is an issue in the measurement of the sensitivity FPN according to this definition, because the contrast value for which a probability of 0.5 for spiking an event is different for each pixel, and it would be very long to establish it for all the pixels individually. We therefore prefer to measure the sensitivity FPN for fixed contrast values, then to interpolate in order to obtain a continuous S curve and find the contrast needed to spike with a 0.5 probability for each pixel. In practice, we apply n times the same stimulus to pixels belonging to the selected ROI, at a given frequency f. For each stimulus, we record whether a pixel spikes or not by defining $P_i(e|C)$, the probability that a pixel spikes at least once after the ith stimulus when the contrast is C. The probability that a pixel spikes under a $C\%$ irradiance contrast is

$$P(e|C) = \frac{1}{n}\sum_{i=1}^{n} P_i(e|C). \qquad (8)$$

To measure the sensitivity, we increase the contrast, most of the time around the estimated sensitivity, between 40% and 80%. Thus, considering a square ROI, whose size is (s_x, s_y), if for each pixel $p_{x,y}$ belonging to this ROI, with $x \in [1, s_x]$ and $y \in [1, s_y]$, it is possible to interpolate the sensitivity measurement to compute the contrast $C_{p_{x,y}}$ so that $p(e|C_{p_{x,y}}) = 0.5$. This average contrast in the ROI is

$$\overline{C_{p_{x,y}}} = \frac{1}{s_x \cdot s_y}\sum_{i=1}^{s_x}\sum_{j=1}^{s_y} C_{p_{i,j}}. \qquad (9)$$

Then, considering given irradiance offset and environment temperature, the FPN is defined as

$$\text{FPN}_{\text{ROI}} = \sqrt{\frac{1}{s_x \cdot s_y}\sum_{i=1}^{s_x}\sum_{j=1}^{s_y}(C_{p_{i,j}} - \overline{C_{p_{x,y}}})}. \qquad (10)$$

It should be noted that the sensitivity of the pixel is theoretically independent of the offset irradiance. However, when the irradiance increases in the focal plane, the noise and the temperature of the array also increase. Photons interact with analog transistors and generate noisy electrons, which decreases the global sensitivity of the camera. Sometimes a pixel does not spike because the noise made it spike just before the light change. Since the output of the sensor is noisy, a strategy must

be set to separate noise from event corresponding to a light change.

Generating a temporal contrast of irradiance on the pixel photodiode modifies the tension at the output of the differential amplifier V_{dif} by a relative theoretical value given by Eq. (1). If V_{dif} is closed to one of the two thresholds, V_{dif} could cross the nearest threshold by itself because of the noise due to the transistors or the photodiode. Hence, according to V_{dif}, different responses can result from the same stimulus. In order to minimize the influence of these artifacts, we repeat the experiment many times (a few hundreds of times), and average the S curves resulting from the different pixels of the ROI. When the statistics of the response in sensitivity are computed, pixels just spiking because of the noise are excluded. As we know which pixel spikes and when, thanks to its timestamp, we can discard pixels spiking in a time window before the stimulus. The time window length is equal to the sum of the refractory period and the maximum latency. A second temporal window is used after the stimulus, whose length must be carefully selected: if it is too long, too much noise will be included in the average S curve. We therefore propose to analyze the activity of the ROI after a flash in order to determine the most appropriate length for this temporal window.

By definition, an event is defined as a quadruplet (x, y, t, p), where (x, y) is the position of the pixel in the array, t the timestamp according to camera's clock, and p the polarity of irradiance temporal change. Considering that we only focus on a positive event, events will only be represented by a triplet (x, y, t). At the ith stimulus of contrast C from the offset irradiance o, $E_{i,C,o}(x, y)$ is the set of events generated. This set can be seen as a set of timestamps and thus represented as $E_{i,C,o}(x, y) = (t_1, \ldots, t_n)$, where $n = \#E_{i,C,o}(x, y)$. Using an ideal retina, $n = 1$ for each pixel. But this is not the case in reality because of the noise and the dead pixels. We express the activity in events generated per microsecond. Figure 4 shows four mean activity curves in response to four different contrasts.
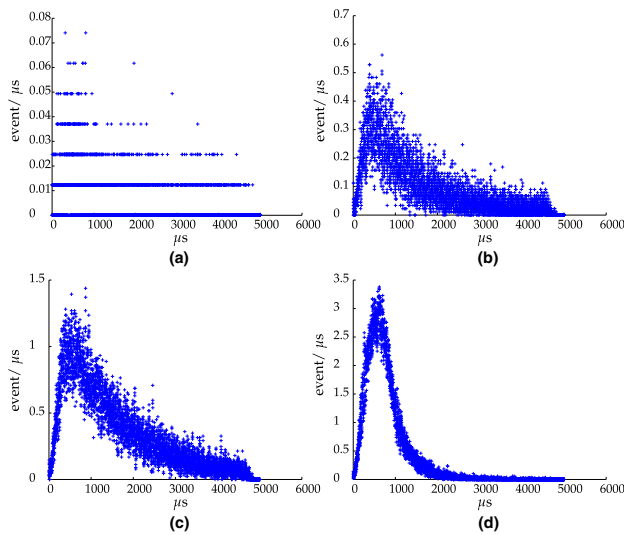

Fig. 4. Activity for four different contrasts under a 617 lux offset. (a) Contrast of 48%; (b) contrast of 52%; (c) contrast of 56%; (d) contrast of 60%.

Activity curves reveal some pixels spiking with a very low latency (less than 10 μs). These correspond to pixels spiking because of the noise. A higher contrast decreases the noisiness of the activity curve. We have to select the temporal window that takes all the signal and minimizes the noise. There are several ways to set the temporal window after the stimulus: the activity of the AER bus can be monitored as in Fig. 4, but a compromise must be found between noise and signal. We define two more representations, which are the position in the stream of events depending on the timestamp, or computing the temporal histogram of the time of first spike (TFS). The TFS for a given stimulus (i, C, o) and a pixel (x, y) is defined as

$$\text{TFS}_{i,C,o}(x, y) = \min_t(E_{i,C,o}(x, y)). \quad \textbf{(11)}$$

The TFS histogram is defined as the function that associates the time after the stimulus and the number of pixels whose first event is received at this time. Considering each time t belonging to a given temporal window after the stimulus W,

$$\text{hist}(t) = \#\{(x, y, t) \in P \times W | \text{TFS}_{i,C,o}(x, y) = t\}. \quad \textbf{(12)}$$

No significant differences were observed using TFS histograms or AER activity. However, we have tried to use the curves relating the timestamp of an event and its position in the event stream after the flash. This consists in adding a new dimension to the event vector, which becomes: (x, y, t, p, ψ), where ψ corresponds to pixel position in the queue in the AER bus. The initial position in the queue begins just after the stimulus. There are three parts in this representation: the first part is not noisy: events leave the imager as soon as possible and the slope of the curve directly depends on the bandwidth of the AER handshake protocol. The second part is more noisy: the bus is less saturated, and it corresponds to slower pixels reacting to the flash among other noisy pixels. The last part results to the noise of the imager, whose frequency is easily known (see further sections). This curve is measured for each flash and the mean of these curves is analyzed. The length of the temporal window used to separate the useful signal and the noise corresponds to the time when the mean curve starts to be noisy. This noisy state is met when the deviation of ordinates crosses a fixed threshold. In practice, methods used to separate noise events from the true signal, which are based on the activity or the histogram of the first spike, generate longer temporal windows. For the method based on the position, it is the contrary. We also observed experimentally that when many pixels are stimulated at the same time, the jitter increases. This phenomenon enlarges the size of the temporal window of the true signal. Then the method based on the activity was used in practice, but the drawback is that when more events are taken into account, then the imager is more sensitive.

C. Latency and Jitter

In order to measure the latency and the jitter, the delay between the stimulus and the events generated must be precisely known. As the timestamping is made by the arbiter protocol, the performance of this latter in the treatment of collisions influences the latency directly. In previous latency measurements, such as [23], it is not said how many pixels are triggered. The total latency to emit an event δ_{tr} is the sum of the latency δ_{pix} of the pixel and the latency δ_{AER} of the AER protocol.

To measure it, the clock of the camera (used to timestamp the event) must be synchronized with the stimulus generated by the trigger of the Smartek strobe controller. The synchronization is done by adding an event corresponding to the detection of a rising (or a falling) edge of one of the field-programmable gate array (FPGA) general purpose input output (GPIO) belonging to the camera. This FPGA is used to implement the AER protocol and to timestamp the events. This additional event is sent with maximum priority into the AER bus to measure the absolute time of the stimulus with good precision. Once the setup is synchronized with the camera clock, it is important to correct the constant error in latency due to the strobe controller, which generates an output signal a few nanoseconds after the trigger signal. After correction, the global latency δ_{tr} is measured. When a single pixel spikes, the latency of the AER bus is negligible, since it represents a few nanoseconds compared to the latency of the pixels, whose value is around tens of microseconds. The AER process latency is not negligible anymore if the stimulated ROI contains a lot of pixels. This is why the ROI previously used for characterizing the sensitivity cannot be used for characterizing the latency. Moreover, the contrast used to measure latency must be high enough to be sure that each pixel spikes. A 200% temporal irradiance contrast is applied in our experiments.

ATIS designers propose an estimation of the latency [24]. It is shown that when the irradiance level is low, pixel bandwidth is limited by the first part of the pixel, which transforms the photocurrent logarithmically into a voltage, whereas when the irradiance level is high, the latency depends on the bandwidth of the differential amplifier and the comparators. In order to measure the pixel mean latency and its standard deviation, the imager is stimulated in various ROIs covering the whole imager area. The size selected for the ROIs is chosen to be as small as possible in order to avoid perturbations due to the latency of the AER arbiter. The ideal operation would be to illuminate each pixel individually, but this would take too much time. A trade-off is found measuring the impact of the arbiter by increasing the size of an ROI and measuring the mean latency of one pixel shared by different ROIs.

Let us consider a pixel $p_{x,y}$ with $x \in [1, s_x]$ and $y \in [1, s_y]$. $\overline{\delta_{pix}(p_{x,y})}$ is the average latency of pixel $p_{x,y}$ and $\overline{\sigma_{pix}(p_{x,y})}$ its jitter. According to Fig. 5, if the ROI is a 15 × 15 square, then the measured latency and the jitter are not influenced by the number of pixels stimulated. The average latency of the pixel can be written
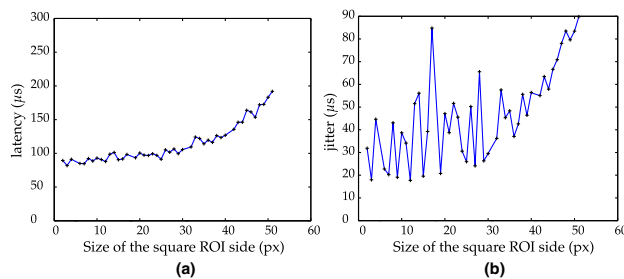

Fig. 5. Latency and jitter of one pixel belonging to an ROI whose size increases. (a) Latency; (b) jitter.

$$\overline{L}(p_{x,y}) = \overline{\delta_{pix}(p_{x,y})} + a\overline{\delta_{AER}} \simeq \overline{\delta_{pix}(p_{x,y})}, \quad (13)$$

and its standard deviation,

$$\overline{J}(p_{x,y}) = \overline{\sigma_{pix}(p_{x,y})} + a\overline{\sigma_{AER}} \simeq \overline{\sigma_{pix}(p_{x,y})}. \quad (14)$$

The latency FPN is measured by computing the spatial deviation of the latency.

D. BGN

When a constant irradiance is applied to the pixel array, parasitic events are generated as deeply analyzed in [23,38]. They result from different temporal noises, and their number tends to increase when the comparator thresholds decrease [37]. These events could be considered as imager BGN. This noise is not spatially uniform over the pixel array. The noise map is described by a frequency map. The measurement of this map is done as follows: during a given time T, a constant irradiance is presented in front of the array and the event stream is recorded for each pixel (x, y). The number of events with polarity pol, denoted as $b_{x,y}^{pol}$, is measured. The BGN is then defined as an event activity characterized by its frequency. Thus, considering a fixed threshold s,

$$\text{BGN}_{x,y}^{pol}(s) = \frac{b_{x,y}^{pol}}{T}. \quad (15)$$

The BGN depends on the imager settings (biases) and the irradiance level. Its measurement provides an idea of the quality of the imager's manufacturing and how its dynamic range is compromised by noise under low or high levels of irradiance.

E. Modulated Light Signal Detection

Since contrast sensitivity, latency, and BGN measuring protocols are now defined, we can investigate how the same experimental setup can be used with a vision system. This latter is composed of a DVS vision sensor, provided with some settings, and an algorithm that analyzes event data to extract pixels receiving modulated light. DVS sensors own the ability to quickly detect light changes [23]. Thus, if the light stimulus is modulated at given frequency f, the DVS pixel will sample the signal and a frequency analysis in the event domain must help to detect the modulation. This application of DVS imagers has already been explored by [45] in order to track a drone equipped with modulated light sources and to estimate its position in the environment. The algorithm proposed detects event patterns in the event stream and measures the period between two following patterns. As the signal is modulated at a given frequency, the pattern "positive event following negative event" and its opposite should be observed with a period equal to the period of the light signal. It is supposed that the distribution of the measured period follows a Gaussian law. Then, each time a period Δ is captured between two patterns in a pixel (x, y), it updates the evidence map pixel value with the relation

$$Ev(x, y)+ = N\left(\frac{1}{\Delta} - f, \sigma^2\right), \quad (16)$$

where f is modulated light frequency and σ the standard deviation of the distribution. The evidence map Ev is then a one-channel image containing the probability that each pixel detects the light signal.

According to the contrast of the light signal, it is known if the sensor will be able to detect the step of the light signal.

After a detected change of illuminance, parameters of the distribution of the timestamp, which follows a normal law whose standard deviation is the jitter [23], are also known. The pixel is not able to detect another change right after it spikes during the refractory period. If δ_{pix} is pixel latency, σ_{pix} the jitter, and T_r the refractory period, then if we consider a square wave light stimulus with a T period, then the minimum period value, T, is given by Eq. (17):

$$T_{\min} = 2(\delta_{\mathrm{pix}} + 3\sigma_{\mathrm{pix}} + T_r). \tag{17}$$

As the timestamp distribution is modeled as a normal law, it is considered that it is not possible to get an event after $\delta_{\mathrm{pix}} + 3\sigma_{\mathrm{pix}}$. In this particular case, the duty cycle of the light signal is equal to 50%. Let suppose that the goal is to generate two different light signals, with the same frequency but with two different duty cycles. In order to separate the two distributions generated by events due to two duty cycles, $6\sigma_{\mathrm{pix}}$ should separate the two rising edges. Thus, considering a square wave light signal, the number of possible duty cycles N is given by

$$N = \lfloor \frac{T - T_{\min}}{6\sigma_{\mathrm{pix}}} \rfloor. \tag{18}$$

For example, if the latency of the pixel is 100 µs, the jitter 30 µs, and the refractory period 10 µs, then the trade-off between the number of duty cycles and the frequency of the signal is given by Fig. 6.

Thanks to Fig. 6, it can be estimated if a light signal modulated at a frequency f with a contrast c can generate events using a DVS sensor whose jitter, latency, and refractory period have been measured, as described in the previous section. This does not give any information on the final vision system performance: it will also depend on the algorithm. We suppose that the best algorithms are those which try to measure period value between two event patterns. Let us define the absolute error Er between the measured and the true light signal period. If two event patterns are following in the time, the constraints of jitter make it impossible to precisely estimate the time between these two events. In order to deal with it, a threshold
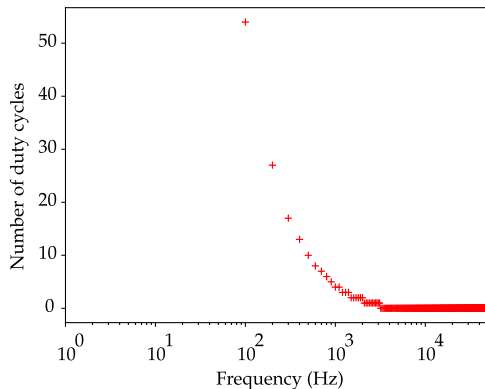
of the accepted error in the measurement must be set. Let s be the accepted absolute error. Due to the jitter, the incertitude in the measurement of the period is $3\sqrt{(2)}\sigma_{\mathrm{pix}}$. Then, a reasonable value of s is 6σ. To measure the robustness of the algorithm to detect the modulation in the event stream provided by the sensor, the absolute error (Er) normalized distribution through its cumulative distribution function is used. In this latter, the probability under which the signal is detected with an accepted absolute error s, which can be specified by the user according to his application, is given by $P(\mathrm{Er} < s)$. The limit of this function, corresponding to an ideal retina-algorithm set, is a step function and means that $P(\mathrm{Er} <= 0) = 1$, i.e., the measured period is the period of the light signal. Because of the incertitude, it is irrelevant to take s less than $3\sqrt{(2)}\sigma_{\mathrm{pix}}$. In the next characterization work, the duty cycle value used is 50%.

Now that all the parameters considered as relevant for the retina's characterization have been defined, their measurement protocol established, and the protocol to measure our vision system performance is defined, the measurement is realized. Event-based sensor characteristics will be compared with those provided by the literature.

4. EXPERIMENTS

This section presents results obtained with one DVS sensor, whose architecture is similar to ATIS [24]. The settings are not designed for a typical application and are a good compromise between sensitivity, latency, and noise. The following measurements illustrate some key characteristics of event-based imagers.

A. Sensitivity Measurement and Its Associated FPN

The sensitivity is measured by applying the protocol described in Section 3.B with four measured irradiance offsets: 7, 72, 1048, and 8686 lux. We observed that the imager sensitivity decreases when the irradiance increases, confirming the experiments of [43]. The used ROI is a 100×100 square centered on the center of the retina. The S curve is computed for each pixel, and the set of S curves obtained for the 10 k pixels in the ROI is averaged. Considering each irradiance offset, the contrast is incrementally increased until 99% of the pixels react. The frequency of the stimulus is 10 Hz, and each acquisition lasts 1 min in order to generate around 600 flashes. Then, $P(e|C)$ is computed for each pixel and contrast value. The temperature of
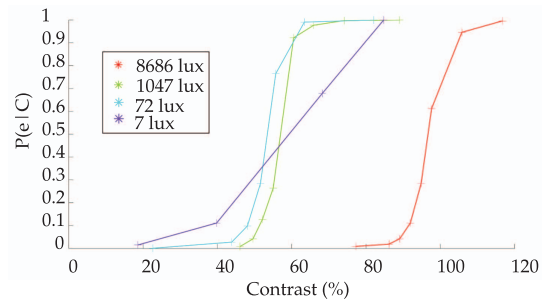

Fig. 6. Number of duty cycles as a function of light signal frequency.

Fig. 7. Average probability to have spikes in an ROI of 100 pixels × 100 pixels as a function of the contrast for various offset irradiance levels.
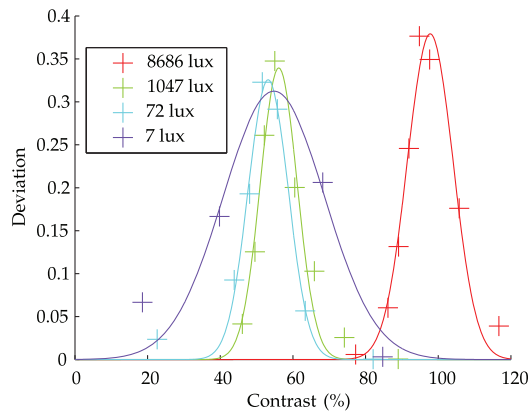
Fig. 8. Deviations of S curves as a function of the contrast and offset's irradiance.

the test chamber is around 20°C. The plots in Fig. 7 confirm the principles raised by [43] regarding the evolution of imager contrast sensitivity according to the irradiance offset. Moreover, for low irradiance levels, it is impossible to separate events due to the signal or the noise, owing to the important noise level resulting in. The main consequence is that the S curve is smoother than expected compared to the S curve measured with high irradiance levels.

Figure 8 illustrates the influence of the noise on the global S curve interpretation. For low irradiance levels (e.g., 7 lux), some pixels have a high probability of spiking, even when the contrast is low (near 20%), but the deviation in the probability estimation in Fig. 8 is much higher than FPN; therefore, the S curve is very noisy. For high irradiance levels (e.g., 8 lux), each pixel does not spike with low contrast levels, while for higher contrast levels, some pixels have a low probability of spiking. Indeed, for low irradiance levels, the noise generates events when the contrast is low, because the threshold could be crossed by dark noise. Considering both high and low irradiances, we cannot expect low FPN, as all pixels do not react in the same way. The authors of Ref. [43] measured a 1.9% FPN in the best conditions and also showed a drop of FPN in low irradiances, as observed in Table 2. It is not possible to provide a relevant comparison between the characterization of works, on the first hand, because the sensor's parameters are not provided, and, on the other hand, because the manufacturing process and pixel design may vary. A deeper study must be conducted to provide a better setting to a given application. Finally, some pixels always spike and can be seen as dead pixels. They are called hot pixels [37] and are taken into account into various computations. It is still an open question how precise statistics of an imager must be close to physical reality for simulation and validation purposes. Therefore, it has been decided to keep them.

Table 2.   FPN Obtained for an ROI of 100 Pixels × 100 Pixels and for Various Offset Irradiances

Irradiance (lux)	7	72	1047	8686
FPN (%)	10.28	4.142	4.013	6.134

B. Latency and Jitter Measurement and Its Associated FPN

Latency measurements are only realized with one irradiance offset, which is about 580 lux. The results are similar to the ones presented in [26,34], with an average latency of 111 μs and a jitter of 40 μs. The FPN of the latency is 15 μs, which represents 15% of the mean latency. Compared to the sensitivity FPN curve according to the irradiance offset, the expected sensitivity FPN under 580 lux is around 4% in contrast, which represents 8% of the mean contrast sensitivity. Then the relative FPN in latency is higher than the interpolated relative FPN in sensitivity at 580 lux. The difference is mainly due to involved transistors, which are not the same. We measured and found that both latency and jitter distributions are rather uniform on the tested retina. The two AER arbiters are not symmetric; then, if the influence of the arbiter process were more important, some lines and columns would be slower to output the information. This result confirms the approximations given by Eq. (14).

C. BGN

Considering several irradiance offsets, the BGN is computed for both positive and negative events. We also observed a significant different behavior between positive and negative noise, as in [38]. This noise also depends on the refractory period, i.e., the delay when the pixel cannot react after spiking. The longer the refractory period, the more noiseless is the imager, but it is also less sensitive and slower to detect irradiance changes. Image settings used in these measurements are the same as used before. The statistics curves shown in Fig. 9 reveal that the noise decreases for both positive and negative events when the irradiance offset increases until a certain value (around 400 lux), while it only increases for positive events, but keep decreasing for negative events after this value. Under low irradiances, the dark noise of the electronic overcomes the useful signal and generated events. With high irradiances, the photons generate parasitic electrons, which increase the tension V_{dif} in Eq. (1) after the differential amplifier, thus generating unexpected events. To better understand the noise's origin in DVS pixels, experiments under a thermal chamber are necessary [38].

The noise caused by high irradiance levels is periodic, and its period increases with the irradiance level. As all pixels are illuminated in the same way, they all react to the noise, which is more uniform when the irradiance level is high. When the latter is low, the spatial deviation of the noise is important, as illustrated in Fig. 10.
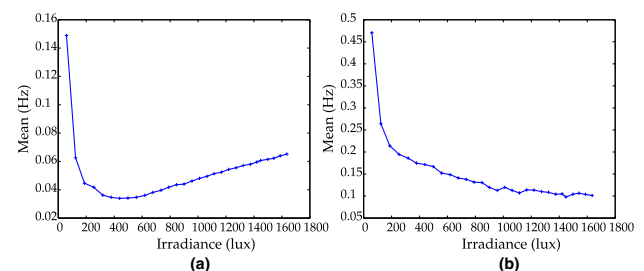

Fig. 9. Average spatial BGN as a function of the irradiance for the positive and negative BGNs. (a) Positive BGN; (b) negative BGN.
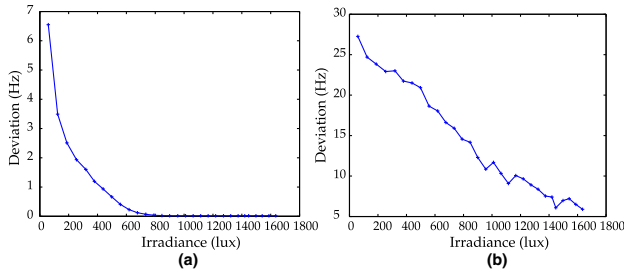
Fig. 10. Spatial deviation of the BGN as a function of the irradiance for the positive and negative BGNs. (a) Positive BGN; (b) negative BGN.

Fig. 11. Percentage of the relative error of the measurement of signal period.

Considering that the main characteristics of the light signal have been measured, it can be characterized how a DVS sensor is able to detect a modulated light signal.

D. Modulated Light Signal Detection

The same experimental setup is used: a given square wave light stimulus is generated with a frequency f, a relative contrast c, and an offset off, and stimulates an ROI of the imager. The size of the ROI can be changed during experiments, as it has been observed in Section 3.C that it directly influences the latency and the jitter of the sensor, and then the signal detection. After being stimulated for 1 min, a set of events is recorded for each pixel $E_{x,y}$. The detection algorithm used is inspired by [45]: a positive to negative or negative to positive event pattern has to be detected. The discrete convolution in the event domain between a filter $F = \{f_0, ..., f_j\}$, describing a set of filter events, with an event stream recorded $E = \{e_0, ..., e_i\}$ is as follows:

$$F * E(n) = \sum_k f_k e_{n-k}. \qquad (19)$$

As this operation corresponds to a convolution centered in the nth event, it will also be written that $F * E(n) = F * E(e_n)$. In order to detect a positive to negative pattern, we can use the filter F_{pos}:

$$F_{\text{pos}} = \{(0, 0, 0, 1)^T, (0, 0, 0, -1)^T\}, \qquad (20)$$

and its opposite F_{neg} to detect negative to positive patterns. If the convolution result of $E_{x,y}$ with one of these filters is equal to -2 in position γ, then it would mean that an event pattern is detected in the event e_γ. Using two filters, two sets of events that detect an event pattern are obtained: $E_{x,y}^{\text{pos}} = \{e \backslash F_{\text{pos}} * E_{x,y}(e) = -2\}$ and $E_{x,y}^{\text{neg}} = \{e \backslash F_{\text{neg}} * E_{x,y}(e) = -2\}$. The set of measured periods $\Delta_{x,y}^{\text{pos}}$ and $\Delta_{x,y}^{\text{neg}}$ is finally obtained computing timestamp differences between events in $E_{x,y}^{\text{pos}}$ and $E_{x,y}^{\text{neg}}$. The goal of our study is to estimate the probability that pixels can detect a signal modulated at a given frequency f. The first metrics used to measure the overall performance of the detection is the relative error between the expected period and both $\Delta_{x,y}^{\text{pos}}$ and $\Delta_{x,y}^{\text{neg}}$. As our experimental setup gives us the precise timing between the light signal and recorded events, the error is computed for four different frequencies: 500, 1000, 2000, and 3000 Hz:

$$E_r = \sum_k \frac{|\Delta_{x,y}^{\text{pos}}(k) - T|}{2T} + \sum_k \frac{|\Delta_{x,y}^{\text{neg}}(k) - T|}{2T}. \qquad (21)$$
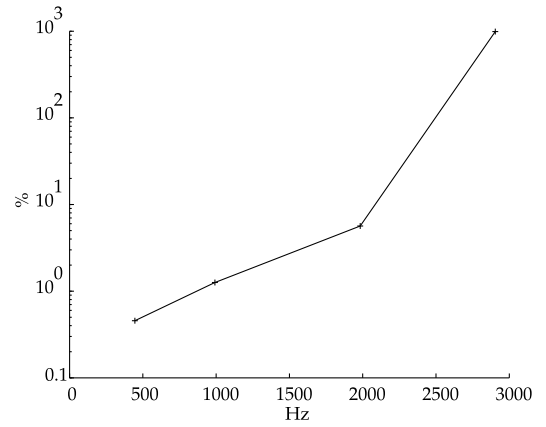
Figure 11 reveals that when the frequency of the light signal increases, the measured relative error also increases and is never null. This confirms that event-based detection algorithms should take into account the detection error of the sensor, and that a threshold must be set. In the work of [45], a similar strategy consists of applying a Gaussian function around the desired period. A new approach, whose threshold is based on the estimation of the probability, is proposed to detect the modulated signal. The imager is stimulated during a given time, and the relative error made to detect a signal's periods is measured in order to approximate the error distribution. We aim at estimating $P(E_r|f)$ through its cumulative distribution function. This curve can be interpreted as the probability of the pixel to detect the signal if we accept a given error E_{th}: $P(X < E_{\text{th}}|f)$, where X is, the signal is detected. As a sensor's response in latency varies, we cannot get rid of this error, and a threshold must be set as close as possible to $E_{\text{th}} = 0$ to ensure that the detection of the signal is robust.

If a 10% error is accepted in the measurement of the period, then the probability of detecting the signal is superior to 0.9 for 500, 1000, and 2000 Hz, but is about 0.7–3000 Hz. This represent an error of 50, 100, 200, and 300 μs, which is enough as the measured jitter j_{pix} is around 30 μs, and as the absolute error in the measurement cannot be less than $6j_{\text{pix}}$. In Fig. 12, the blue curve corresponds to the distribution function of $P(E_{\text{rel}}|f)$ in one pixel, and the observed steps correspond to measurements of multiples of the period. This supposes that sometimes a pixel misses an edge of the signal. In order to counter that effect, we propose to use neighboring pixels. Indeed, in function of the targeted application, a modification on the lens can be made to ensure that the projection of the light source on the focal plane crosses several pixels. In that way, if a pixel misses detecting a change, maybe one of its neighboring pixels detected it. Let us suppose that we know $\Delta_{x,y}^{\text{pos}}$ and $\Delta_{x,y}^{\text{neg}}$ for each pixel of the imager (we can take the last measurement). The corrected measurement, $\widehat{\Delta_{x,y}^{\text{pos}}} = \Delta_{m,n}^{\text{pos}}$, satisfies the constraint $\forall (i, j) \in [x - s, x + s] \times [y - s, y + s], \|\Delta_{m,n}^{\text{pos}} - T\| < \|\Delta_{i,j}^{\text{pos}} - T\|$.
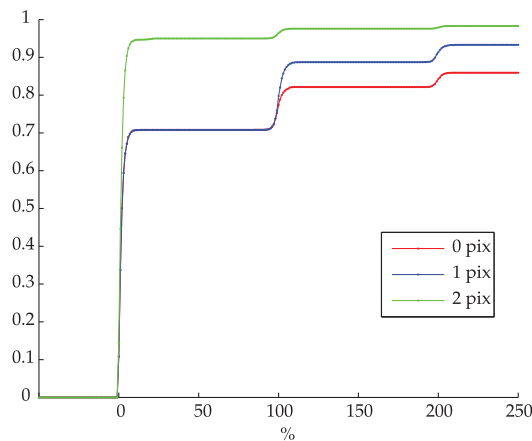
Fig. 12. Distribution function of the corrected relative error for a signal modulated at 3000 Hz with different neighborhoods (in pixels).

The results are displayed in Fig. 12, considering a light signal modulated at 3000 Hz, and an algorithm using three *s* values: 0, 1, and 2. The larger the neighborhood, the better is the probability of detecting the signal. This modest improvement in the method proposed in [45] allows us to reduce the error made by this algorithm to detect a signal modulated at 3 kHz with a 50% duty cycle.

The detection of the light signal is now characterized as a function of some sensor, light signal, and algorithm properties. Further experiments could be done to increase the space of parameters, such as measurements in a thermal chamber.

5. CONCLUSION

This paper proposes an experimental setup in order to characterize DVS sensors and how event-based vision systems can detect square wave modulated light signals. Several characteristics of the sensor were identified and measured, without asking the manufacturer to provide them exhaustively. Measurement of characteristics such as the noise, the contrast sensitivity and the latency, and their spatial and temporal statistics, is now possible on the whole area of the imager. This enables the development of a simulation model for CMOS retina aimed at increasing the databases necessary to validate a vision algorithm for these new imagers. The setup can also be easily adapted to test new generations of DVS sensors. Moreover, this setup is used to measure the detection robustness of a vision system to detect modulated light. A new strategy is presented to increase the frequency of the modulation while maintaining a high level of the detection performance. Event-based sensors are easier to characterize than conventional cameras, as their parameters are not adapted to lighting conditions as they could be for frame-based sensors. This also permits better characterization and monitoring of the performance of the vision system, for example, designed so as to detect a modulated light signal.

REFERENCES

1. C. D. Harper, C. T. Hendrickson, and C. Samaras, "Cost and benefit estimates of partially-automated vehicle collision avoidance technologies," Accident Anal. Prev. **95**, 104–115 (2016).
2. S.-C. Liu, T. Delbruck, G. Indiveri, R. Douglas, and A. Whatley, *Event-Based Neuromorphic Systems* (Wiley, 2015).
3. S. Zhang, R. Benenson, M. Omran, J. Hosang, and B. Schiele, "How far are we from solving pedestrian detection?" arXiv:1602.01237 (2016).
4. A. Amir, B. Taba, D. Berg, T. Melano, J. McKinstry, C. Di Nolfo, T. Nayak, A. Andreopoulos, G. Garreau, M. Mendoza, and J. Kusnitz, "A low power, fully event-based gesture recognition system," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017), pp. 7243–7252.
5. J. A. Pérez-Carrasco, B. Zhao, C. Serrano, B. Acha, T. Serrano-Gotarredona, S. Chen, and B. Linares-Barranco, "Mapping from frame-driven to frame-free event-driven vision systems by low-rate rate coding and coincidence processing: application to feedforward ConvNets," IEEE Trans. Pattern Anal. Mach. Intell. **35**, 2706–2719 (2013).
6. Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," Neural Comput. **1**, 541–551 (1989).
7. D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," J. Physiol. **160**, 106–154 (1962).
8. D. P. Moeys, F. Corradi, E. Kerr, P. Vance, G. Das, D. Neil, D. Kerr, and T. Delbrück, "Steering a predator robot using a mixed frame/event-driven convolutional neural network," in *Second International Conference on Event-based Control, Communication, and Signal Processing (EBCCSP)* (IEEE, 2016), pp. 1–8.
9. A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: an open urban driving simulator," in *Proceedings of the 1st Annual Conference on Robot Learning* (2017), pp. 1–16.
10. A. Gaidon, Q. Wang, Y. Cabon, and E. Vig, "Virtual worlds as proxy for multi-object tracking analysis," arXiv:1605.06457 (2016).
11. F. Yu, W. Xian, Y. Chen, F. Liu, M. Liao, V. Madhavan, and T. Darrell, "BDD100k: a diverse driving video database with scalable annotation tooling," arXiv:1805.04687 (2018).
12. M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 3213–3223.
13. S.-C. Liu, J. Kramer, G. Indiveri, T. Delbrück, and R. Douglas, *Analog VLSI: Circuits and Principles* (MIT, 2002).
14. P.-F. Ruedi, P. Heim, F. Kaess, E. Grenet, F. Heitger, P.-Y. Burgi, S. Gyger, and P. Nussbaum, "A 128 × 128 pixel 120-db dynamic-range vision-sensor chip for image contrast and orientation extraction," IEEE J. Solid-State Circuits **38**, 2325–2333 (2003).
15. T. Delbrück and C. Mead, "An electronic photoreceptor sensitive to small changes in intensity," in *Advances in Neural Information Processing Systems* (1989), pp. 720–727.
16. J. A. Leñero-Bardallo, T. Serrano-Gotarredona, and B. Linares-Barranco, "A 3.6 s latency asynchronous frame-free event-driven dynamic-vision-sensor," IEEE J. Solid-State Circuits **46**, 1443–1455 (2011).
17. V. Chan, S.-C. Liu, and A. van Schaik, "AER EAR: a matched silicon cochlea pair with address event representation interface," IEEE Trans. Circuits Syst. I, Reg. Papers **54**, 48–59 (2007).
18. F. Akopyan, J. Sawada, A. Cassidy, R. Alvarez-Icaza, J. Arthur, P. Merolla, N. Imam, Y. Nakamura, P. Datta, G.-J. Nam, and B. Taba, "Truenorth: design and tool flow of a 65 mw 1 million neuron programmable neurosynaptic chip," IEEE Trans. Computer-Aided Design Integr. Circuits Syst. **34**, 1537–1557 (2015).
19. S. B. Furber, D. R. Lester, L. A. Plana, J. D. Garside, E. Painkras, S. Temple, and A. D. Brown, "Overview of the spinnaker system architecture," IEEE Trans. Comput. **62**, 2454–2467 (2013).
20. V. Y.-S. Chan, C. T. Jin, and A. van Schaik, "Neuromorphic audio-visual sensor fusion on a sound-localising robot," Front. Neurosci. **6**, 21 (2012).

21. R. H. Masland, "The neuronal organization of the retina," Neuron **76**, 266–280 (2012).

22. C. Brandli, R. Berner, M. Yang, S.-C. Liu, and T. Delbruck, "A 240×180 130 db 3 µs latency global shutter spatiotemporal vision sensor," IEEE J. Solid-State Circuits **49**, 2333–2341 (2014).

23. P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128×128 120 db 15µs latency asynchronous temporal contrast vision sensor," IEEE J. Solid-State Circuits **43**, 566–576 (2008).

24. C. Posch, D. Matolin, and R. Wohlgenannt, "A QVGA 143 dB dynamic range frame-free PWM image sensor with lossless pixel-level video compression and time-domain CDS," IEEE J. Solid-State Circuits **46**, 259–275 (2011).

25. B. Son, Y. Suh, S. Kim, H. Jung, J.-S. Kim, C. Shin, K. Park, K. Lee, J. Park, J. Woo, and J. Roh, "4.1 a 640 × 480 dynamic vision sensor with a 9 µm pixel and 300 meps address-event representation," in *IEEE International Solid-State Circuits Conference (ISSCC)* (IEEE, 2017), pp. 66–67.

26. D. P. Moeys, C. Li, J. N. Martel, S. Bamford, L. Longinotti, V. Motsnyi, D. S. S. Bello, and T. Delbruck, "Color temporal contrast sensitivity in dynamic vision sensors," in *IEEE International Symposium on Circuits and Systems* (2017), pp. 1–4.

27. C. Li, C. Brandli, R. Berner, H. Liu, M. Yang, S.-C. Liu, and T. Delbruck, "Design of an RGBW color VGA rolling and global shutter dynamic and active-pixel vision sensor," in *IEEE International Symposium on Circuits and Systems (ISCAS)* (IEEE, 2015), pp. 718–721.

28. G. Taverni, D. P. Moeys, C. Li, C. Cavaco, V. Motsnyi, D. S. S. Bello, and T. Delbruck, "Front and back illuminated dynamic and active pixel vision sensors comparison," IEEE Trans. Circuits Syst. II, Exp. Briefs **65**, 677–681 (2018).

29. K. A. Boahen, "Point-to-point connectivity between neuromorphic chips using address events," IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process. **47**, 416–434 (2000).

30. R. Serrano-Gotarredona, M. Oster, P. Lichtsteiner, A. Linares-Barranco, R. Paz-Vicente, F. Gomez-Rodriguez, H. K. Riis, T. Delbruck, S.-C. Liu, S. Zahnd, and A. M. Whatley, "AER building blocks for multi-layer multi-chip neuromorphic vision systems," in *Advances in Neural Information Processing Systems* (2006), pp. 1217–1224.

31. M. Yang, S.-C. Liu, and T. Delbruck, "Analysis of encoding degradation in spiking sensors due to spike delay variation," IEEE Trans. Circuits Syst. I, Reg. Papers **64**, 145–155 (2017).

32. "Emva1228 standard for characterization of image sensors and cameras release 3.1" (2016), https://www.emva.org/wp-content/uploads/EMVA1288-3.1a.pdf.

33. X. Guo, X. Qi, and J. G. Harris, "A time-to-first-spike CMOS image sensor," IEEE Sens. J. **7**, 1165–1175 (2007).

34. C. Posch, D. Matolin, and R. Wohlgenannt, "An asynchronous time-based image sensor," in *IEEE International Symposium on Circuits and Systems* (IEEE, 2008), pp. 2130–2133.

35. T. Delbruck and C. A. Mead, "Adaptive photoreceptor with wide dynamic range," in *Proceedings of IEEE International Symposium on Circuits and Systems-ISCAS'94* (IEEE, 1994), Vol. **4**, pp. 339–342.

36. K. A. Boahen, "A burst-mode word-serial address-event link-i: transmitter design," IEEE Trans. Circuits Syst. I, Reg. Papers **51**, 1269–1280 (2004).

37. D. P. Moeys, F. Corradi, C. Li, S. A. Bamford, L. Longinotti, F. F. Voigt, S. Berry, G. Taverni, F. Helmchen, and T. Delbruck, "A sensitive dynamic and active pixel vision sensor for color or neural imaging applications," IEEE Trans. Biomed. Circuits Syst. **12**, 123–136 (2018).

38. Y. Nozaki and T. Delbruck, "Temperature and parasitic photocurrent effects in dynamic vision sensors," IEEE Trans. Electron Dev. **64**, 3239–3245 (2017).

39. X. Berthelon, G. Chenegros, T. Finateu, S. Ieng, and R. Benosman, "Effects of cooling on the SNR and contrast detection of a low-light event-based camera," IEEE Trans. Biomed. Circuits Syst. **12**, 1467–1474 (2018).

40. M. Yang, S.-C. Liu, and T. Delbruck, "A dynamic vision sensor with 1% temporal contrast sensitivity and in-pixel asynchronous delta modulator for event encoding," IEEE J. Solid-State Circuits **50**, 2149–2160 (2015).

41. D. P. Moeys, "Analog and digital implementations of retinal processing for robot navigation systems," in *Ph.D. dissertation* (ETH Zurich, 2016).

42. Labsphere, Integrating Sphere Theory and Applications (2014). https://www.labsphere.com/site/assets/files/2551/integrating_sphere_theory_apps_tech_guide.pdf.

43. C. Posch and D. Matolin, "Sensitivity and uniformity of a 0.18 µm CMOS temporal contrast pixel array," in *IEEE International Symposium of Circuits and Systems (ISCAS)* (IEEE, 2011), pp. 1572–1575.

44. P. D. Foote, *Illumination from a Radiating Disk* (Government Printing Office, 1916).

45. A. Censi, J. Strubel, C. Brandli, T. Delbruck, and D. Scaramuzza, "Low-latency localization by active LED markers tracking using a dynamic vision sensor," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (IEEE, 2013), pp. 891–898.