BAYES CLASSIFICATION FOR ASYNCHRONOUS EVENT-BASED CAMERAS

Lionel Fillatre

I3S Laboratory - UMR 7271 - University Nice Sophia Antipolis - CNRS
CS 40121 - 06903 Sophia Antipolis CEDEX, France

ABSTRACT

Asynchronous event-based cameras use time encoding to code the pixel intensity values. A time encoding of an input pattern generates a random stream of asynchronous events. An event is defined as a pair containing a timestamp and the variation sign of the input signal since the last emitted event. The goal of this paper is the recognition of the input pattern among a set of several known possibilities from the observation of the event stream. This paper proposes a statistical model of the random event stream based on the physical model of the event-based camera. It also calculates the optimal Bayes classifier which recognizes the input pattern. The numerical complexity of the classifier is rather low. The Bayes risk, which measures the performance of the classifier, is numerically evaluated on simulated data. It is compared to the mean number of events, which entails the power consumption of the camera, exploited to take the decision.

***Index Terms*—** Time encoding, Statistical classification, Event-based camera, Bayes risk.

1. INTRODUCTION

Asynchronous event-based cameras, also called neuromorphic cameras, are beginning to provide a paradigm shift in the current approaches to address image and video processing [1–5]. Contrary to conventional frame-based image acquisition and processing technologies, event-based dynamic vision sensors provide a novel and efficient way for encoding light and its temporal variations by registering and transmitting only the changes at the exact time at which they occur [4,6,7]. Event-based cameras sense and encode the spatial locations (addresses) and times of changes in light intensity at the pixel level [1]. Asynchronous, continuous-time signal coding and processing based on nonuniform sampling has shown (e.g., [8] and [9]) that such a sampling approach allows to recover some information about the sampled signal. Consequently, event-based cameras have a very high temporal resolution exceeding the speed of most conventional frame based cameras and permit to save power. This novel paradigm of visual data acquisition calls for a new methodology in order to efficiently process the sparse event-based image information without sacrificing its beneficial characteristics [10, 11].

Event-based sensors allow a radical new variety of processing [1, 12, 13]. Asynchronous event-based cameras use a time encoding [4, 14, 15] of the pixel log-intensities $x_k(t)$ for $t > 0$ and all $k = 1, \dots, K$ where K is the number of pixels. Hence, the input time-varying scene is represented as a random sequences, called the event stream, of events $(t_{k,j}, y_{k,j})$ where $(t_{k,j})$, $j \in \mathbb{N}$ is a sequence of strictly increasing times and $y_{k,j} \in \{-1, 1\}$ is the polarity of the event. Many models [16] exist for encoding a signal in the time domain. This paper deals with the level crossing sampling which remains one of the most used techniques, especially with event-based sensors [11]. Most efforts in the literature have been devoted towards the design of reconstruction algorithms and filtering methods [10, 11, 14, 17–19] for bandlimited signals. One of the main lack of existing works is the absence of results about statistical models and classification of event streams.

The contributions of this paper are twofold. First, the event stream is modeled as a sequence a random events. The randomness of the event stream is due to the noise naturally present inside the neuromorphic sensor and coming from the photoreceptors and electronic circuits. Second, we propose an optimal Bayes classifier which processes the random event streams in order to identify an input pattern. The paper is organized as follows. Section 2 describes the asynchronous event-based camera and proposes a statistical model of the event streams. Section 3 derives the optimal Bayes classifier to classify some patterns from random event streams. Section 4 proposes a numerical study of the classifier. Finally, Section 5 concludes this paper.

2. EVENT-BASED CAMERA

2.1. Event-Based Sampling

The goal of this section is to model the event stream produced by the event-based sampling of an input time-varying scene in order to classify the pattern occurring inside the scene. The pattern classification process is described in Fig. 1. An event is defined as a couple composed of a timestamp and a polarity at a given pixel coordinate. The polarity is a sample related to the luminance signal variation captured at the level of a sin-

This work was supported by French Provence Alpes Côte d'Azur project COBRA.

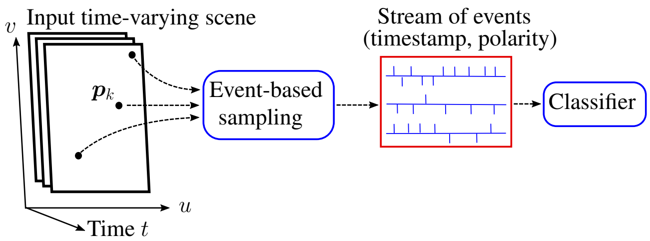

Fig. 1. Pattern classification from the event-based sampling of an input time-varying scene.

gle pixel along with its time of acquisition. Let us introduce the event-based sampling in continuous time. The discrete time model will be modeled in Subsection 2.2. Let $x(t)$ be a continuous time signal we want to sample. The k-th random event is a pair $(t_k, y(t_k))$ where each sample is acquired according to the generic rule

$$x(t_{k+1}) = x(t_k) + \varepsilon_k + \Delta\, y(t_{k+1}) \tag{1}$$

where Δ is the known sampling step and k is a positive integer. This means that each sample $x(t_{k+1})$ is determined by adding or subtracting a constant positive term Δ, up to the noise ε_k, to the previous sample $x(t_k)$. The noise ε_k is difficult to model directly but it is implicitly taken into account in the statistical distribution of the events given in the following.

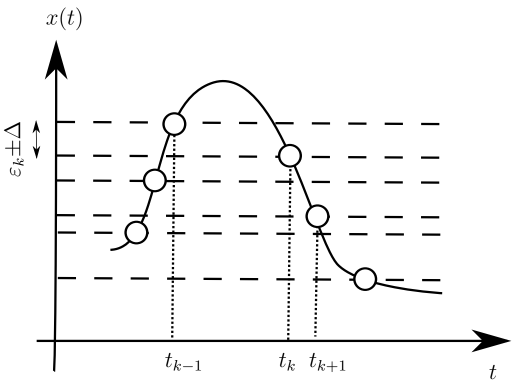

Fig. 2. Random level crossing sampling of $x(t)$.

Fig. 2 shows an example of signal and resulting sampling points in time. This is a random level crossing sampling which is very close to the conventional level crossing sampling, also called send-to-delta or Lebesgue sampling [20], except that the difference between two samples is random.

2.2. Random Event Model

A typical event-based sensor, as the Dynamic Vision Sensor [4], models the transient responses of the retina [21]. A neuromorphic vision system is composed of sensors which are sensitive to the light contrast of the input scene. Each pixel memorizes the log-intensity of the input after each event and emits a new event when the input is significantly different from the memorized log-intensity. More formally, let us assume that the input pattern is characterized by a spatiotemporal intensity function $(u, v, t) \mapsto s(u, v, t)$ where (u, v) belongs to a compact $\mathcal{D} \subset \mathbb{R}^2$ and $t \in [0, T]$ where T is the duration of the acquisition in microseconds. The set of pixels $\boldsymbol{p}_k = (u_k, v_k)$ corresponds to some sampling points of $\mathcal{D}$ for $k = 1, \ldots, K$. Each pixel $\boldsymbol{p}_k$ is associated to a photoreceptor which is sensitive to the log-intensity $x_k(t) = \ln I_k(t)$ of the input scene where $I_k(t) > 0$ is the input photocurrent.

The neuromorphic vision sensor works in discrete time with a typical time resolution of $1\ \mu s$. Let $\{t_{k,1}, \ldots, t_{k,n_k}\}$ be the instants of $[0, T]$ where the event-based sensor emits the events associated to pixel $\boldsymbol{p}_k$. An event produced by pixel $\boldsymbol{p}_k$ is a couple $\boldsymbol{e}_{k,j} = (t_{k,j}, y_{k,j})$ where $t_{k,j}$ is the time of the j-th event and $y_{k,j}$ is its polarity, say -1 (OFF event) or $+1$ (ON event). By convention, $t_{k,0} = 0$ for all pixels k. At discrete time $t \geq t_{k,j}$, the variation of the log-response with respect to the memorized value at time $t_{k,j}$ is given by

$$\begin{aligned} \delta_k(t, t_{k,j}) &= x_k(t) - x_k(t_{k,j}) + \xi_k(t) \\ &= \dot{x}_k(t, t_{k,j}) + \xi_k(t) \end{aligned} \tag{2}$$

for all $j \geq 1$ where $\xi_k(t)$ is a noise. The non-random difference $\dot{x}_k(t, t_{k,j}) = x_k(t) - x_k(t_{k,j})$ is called the temporal contrast after the j-th event. The random temporal contrast $\delta_k(t, t_{k,j})$ is then compared to two non-random thresholds Δ^+ and Δ^- satisfying $\Delta^- < \Delta^+$. It is assumed that $\Delta^+ = \Delta = -\Delta^-$ and that these thresholds are nominally the same for all the pixels. If the difference exceeds the upper threshold Δ^+, the sensor emits an ON event; if it is below the lower threshold Δ^-, the camera emits an OFF event; if the value is between the two thresholds, there is no event. After emitting the $j+1$ event, the input value $x_k(t_{k,j+1})$ is saved in the neuromorphic sensor in order to compute $\delta_k(t, t_{k,j+1})$ for the next event. Strictly speaking, the thresholds Δ^- and Δ^+ are random since they depend on the electronic system. To simplify the model, the randomness of Δ^- and Δ^+ are not considered in this paper. The refractory period of each pixel and the background ON activity described in [4] are ignored.

Experiments in [4, 7, 22, 23] shown that $\xi_k(t)$ is an aggregate of many sources of noise: photonic noise, contrast threshold, uniformity of response, etc. These experiments evaluate contrast sensitivity by measuring the event response probability as a function of increasing contrast at identical initial illuminance. In an ideal noise-free world, this would result in a step (0% to 100% probability with infinite slope) at a given threshold contrast. In reality, noise turns the ideal step into an"S"-shaped curve (see Fig. 14 in [7]). Hence, the noise $\xi_k(t)$ can be approximated by a Gaussian white noise such that $\{\xi_k(t),\ 1 \leq k \leq K,\ 1 \leq t \leq T\}$ is a family of independent and identically distributed variables satisfying $\xi_k(t) \sim \mathcal{N}(0, \sigma^2)$ where $\sigma > 0$ is the standard-deviation of the noise. It is assumed that the sensor is well calibrated, so σ is known.

2.3. Statistical Event Stream Model

Let $p_{k,j+1}(t,-1)$, resp. $p_{k,j+1}(t,+1)$, be the probability that the $j+1$ event is an OFF event, resp. an ON event, emitted at time t by pixel $\boldsymbol{p}_k$. The probability of non-emitting a $j+1$ event at time t at pixel $\boldsymbol{p}_k$ is $p_{k,j+1}(t,0)$. Let us describe the probability distribution of the random event stream for pixel $\boldsymbol{p}_k$. The event stream of pixel $\boldsymbol{p}_k$ is a vector, denoted S_k, of n_k couples $(t_{k,j}, y_{k,j})$:

$$S_k = \{(t_{k,j}, y_{k,j}),\ 1 \leq j \leq n_k\}.$$

The probability of the stream S_k is given by

$$\Pr(S_k) = B_{k,T} \prod_{q=1}^{q=n_k} A_{k,q} \tag{3}$$

where $A_{k,q}$ is the probability of the sequence of events starting from time $t_{k,q-1}+1$ and finishing to time $t_{k,q}$ given by

$$A_{k,q} = p_{k,q}(t_{k,q}, y_{k,q}) \prod_{i=t_{k,q-1}+1}^{t_{k,q}-1} p_{k,q}(i,0) \tag{4}$$

and $B_{k,T}$ is the probability of non-emitting events after the last event emitted at time t_{k,n_k}:

$$B_{k,T} = \prod_{\ell=t_{k,n_k}+1}^{T} p_{k,n_k+1}(\ell,0). \tag{5}$$

These equations underline that the absence of events in the stream is also a source of information. Indeed, the silence between two events, which corresponds to the product over index i in (4), depends on the two events limiting this silence period. The event stream for all the pixels is

$$S_1^K = \{S_k,\ k=1,\ldots,K\}.$$

Let $\mathcal{S}$ be the set of all event streams over $[1,T]$.

Since the noise has a Gaussian distribution, the probabilities of emitting and non-emitting an event are given by:

$$\begin{aligned}
p_{k,j+1}(t,0) &= \Pr(\Delta^- \leq \delta_k(t,t_{k,j}) \leq \Delta^+) = \\
&\Phi\Big(\frac{\Delta^+ - \dot{x}_k(t,t_{k,j})}{\sigma}\Big) - \Phi\Big(\frac{\Delta^- - \dot{x}_k(t,t_{k,j})}{\sigma}\Big), && (6)\\
p_{k,j+1}(t,-1) &= \Pr(\delta_k(t,t_{k,j}) \leq \Delta^-) \\
&= \Phi\Big(\frac{\Delta^- - \dot{x}_k(t,t_{k,j})}{\sigma}\Big), && (7)\\
p_{k,j+1}(t,+1) &= \Pr(\delta_k(t,t_{k,j}) \geq \Delta^+) \\
&= 1 - \Phi\Big(\frac{\Delta^+ - \dot{x}_k(t,t_{k,j})}{\sigma}\Big), && (8)
\end{aligned}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard univariate normal distribution. We have $p_{k,j+1}(t,0) + p_{k,j+1}(t,-1) + p_{k,j+1}(t,+1) = 1$ for all $t \geq t_{k,j}$. It is important to note that the probability of the $j+1$-th event is only conditioned by the j-th event, not by the previous events.

3. OPTIMAL BAYES CLASSIFIER

The observation model shows that the event stream model (3) depends on the input signal, also called the input pattern,

$$X = \{x_k(t), k=1,\ldots,K,\ 1 \leq t \leq T\}.$$

Let us assume that X belongs to a set $\mathcal{P} = \{X_1,\ldots,X_m\}$ of m known input patterns. Let us suppose that the neuromorphic sensor is observing the pattern X_θ. Let $\Pr_\theta(S_1^K)$ the probability of random stream S_1^K when the input pattern is X_θ. It is assumed that each pattern occurs with the same probability, i.e., the prior probability is $q_\theta = 1/m$ for each pattern X_θ. A classifier is a function $\phi : \mathcal{S} \mapsto \{1,\ldots,m\}$ such that pattern X_θ is chosen when $\phi(S_1^K) = \theta$. Under the above mentioned assumptions, it makes sense to seek the Bayes classifier ϕ^* which minimizes the Bayes risk $R(\phi)$:

$$R(\phi) = \frac{1}{m}\sum_{\theta=1}^{m}(1-\alpha_\theta(\phi)), \tag{9}$$

where $\alpha_\theta(\phi) = \Pr_\theta(\phi(S_1^K) = \theta)$ is the probability of correct recognition of X_θ. Hence, as described in [24], the Bayes classifier ϕ^* is given by

$$\begin{aligned}
\phi^*(S_1^K) &= \arg\max_{1\leq\theta\leq m} \Pr_\theta(S_1^K) \\
&= \arg\max_{1\leq\theta\leq m} \sum_{k=1}^{K} \ln \Pr_\theta(S_k) && (10)
\end{aligned}$$

where $\Pr_\theta(S_k)$ is given in (3) in the case $X = X_\theta$. The derivation of (10) is based on the statistical independence of the pixels. It is important to note that the randomness of the pixels is due to the noise $\xi_k(t)$, not to the input pattern structure X.

4. NUMERICAL EXPERIMENTS

This section illustrates the performance of the Bayes classifier on simulated data. We consider three rotating input geometric patterns as shown in Fig. 3: a disk (with label $\theta = 1$), a triangle ($\theta = 2$) and a square ($\theta = 3$). The geometric pattern is rotating over a constant level background.

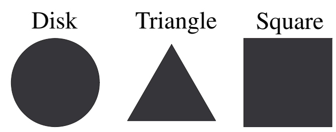

Fig. 3. Geometric patterns which are rotating over a white background to produce the event stream.

The simulation lasts 1 ms, hence $T = 1000\ \mu s$. Each image is composed of 64×64 pixels. The pixel value of the geometric pattern is 1 whereas each background pixel takes the

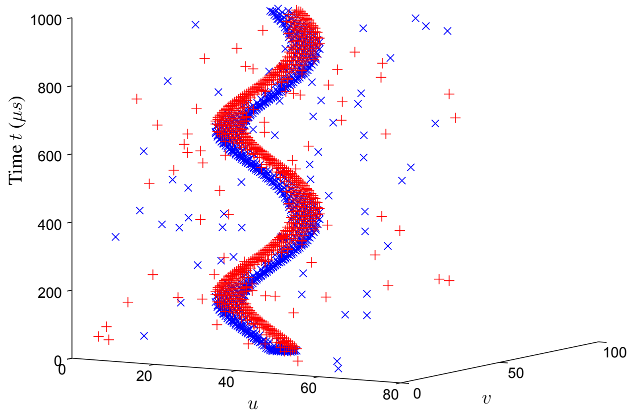

Fig. 4. Event stream produced by the rotating disk. The blue cross, resp. red plus, signs represent ON, resp. OFF, events.

value 0 (it is a binary pattern). The geometric pattern makes 2 complete rounds during the observation period. The rotation is really fast but we could slow down the angular velocity and increase T to be more realistic. We prefer to use a small T in order to save computation effort, i.e., in order to minimize the number of samples to process. The numerical complexity of the Bayes classifier is rather low but the number of samples, especially during the event generation, could be huge since the time resolution is 1 μs. The event stream S_1^K is computed as modeled in Subsection 2.2. A typical event stream for the rotating disk is shown in Fig. 4 with $\Delta^+ = -\Delta^- = 0.5 = \Delta$ and $\sigma = 0.1$. We can see the rotating disk and many noisy events occurring arbitrarily in space and time.

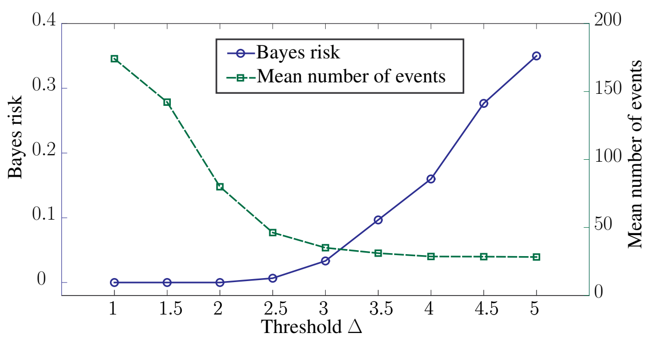

Fig. 5. Bayes risk and mean number of events as a function of the threshold Δ.

The Bayes risk is numerically evaluated with a Monte Carlo simulation. For each pattern, we generate 100 random streams of 100 μs following the model describing in Subsection 2.2 for 9 values of Δ starting from 1 to 5 with a step increment 0.5. The standard deviation satisfies $\sigma = \Delta/4$. The simulation duration is decreased to 100 μs because the performance of the classifier is too high (the Bayes risk is close to 0) when there is a large number of events. The Bayes classifier ϕ^* of Section 3 is computed for each random stream by exploiting all the events in the stream, hence it exploits all the pixels during the full simulation duration. The error probabilities $1 - \alpha_\theta(\phi^*)$ are estimated for each pattern $\theta \in \{1, 2, 3\}$. The error probabilities are estimated by dividing the number of classification errors over the total number of tested random streams. The value of the Bayes risk is shown in Fig. 5. As expected, the Bayes risk is increasing as Δ is increasing. As a matter of fact, when the threshold Δ is increasing, the number of events is decreasing, hence the sensibility of the event-based sensor to the pattern is decreasing.

For each value of Δ, we also compute the mean number of events extracted from the full family of random streams including all the patterns and all the pixels. The mean number of events for a stream duration of 100 μs is plot in Fig. 5. As underlined previously, we clearly see that the mean number of events is decreasing as the threshold is increasing. So there is a natural balance between the Bayes risk and the number of events. Saving energy to process the data involves to increase the threshold of the sensor and, consequently, to reduce the classification performance.

5. CONCLUSION

This paper proposes a statistical model of random event streams and the Bayes classifier derived from this model. The resulting classifier has a more complex structure than conventional classifier exploiting frame-based video. However, this is compensated by the fact that event-based cameras are saving power. Future work will apply the Bayes classifier to real data.

REFERENCES

[1] S.C. Liu, T. Delbruck, G. Indiveri, A. Whatley, and R. Douglas, *Event-Based Neuromorphic Systems*, John Wiley and Sons Ltd., UK, 2014.

[2] C. Brandli, R. Berner, MinHao Yang, Shih-Chii Liu, and T. Delbruck, "A 240 × 180 130 db 3 μs latency global shutter spatiotemporal vision sensor," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 10, pp. 2333–2341, 2014.

[3] MinHao Yang, Shih-Chii Liu, and T. Delbruck, "Comparison of spike encoding schemes in asynchronous vision sensors: Modeling and design," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2014, pp. 2632–2635.

[4] P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128 × 128 120 dB 15 μs Latency Asynchronous Temporal Contrast Vision Sensor," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 2, pp. 566–576, 2008.

[5] C. Posch, T. Serrano-Gotarredona, B. Linares-Barranco, and T. Delbruck, "Retinomorphic event-

based vision sensors: Bioinspired cameras with spiking output," *Proceedings of the IEEE*, vol. 102, no. 10, pp. 1470–1484, Oct 2014.

[6] J.A. Leñero Bardallo, T. Serrano-Gotarredona, and B. Linares-Barranco, "A 3.6 μs Latency Asynchronous Frame-Free Event-Driven Dynamic-Vision-Sensor," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 6, pp. 1443–1455, 2011.

[7] 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 Journal of Solid-State Circuits*, vol. 46, no. 1, pp. 259–275, 2011.

[8] C. Vezyrtzis and Y. Tsividis, "Processing of signals using level-crossing sampling," in *IEEE International Symposium on Circuits and Systems*, May 2009, pp. 2293–2296.

[9] Yannis Tsividis, "Event-driven data acquisition and digital signal processing: A tutorial," *Trans. Cir. Sys.*, vol. 57, no. 8, pp. 577–581, 2010.

[10] Ryad Benosman, Sio-Hoi Ieng, Charles Clercq, Chiara Bartolozzi, and Mandyam Srinivasan, "Asynchronous frameless event-based optical flow.," *Neural Networks*, vol. 27, pp. 32–37, 2012.

[11] Sio-Hoi Ieng, C. Posch, and R. Benosman, "Asynchronous neuromorphic event-driven image filtering," *Proceedings of the IEEE*, vol. 102, no. 10, pp. 1485–1499, Oct 2014.

[12] O. Braddick, "Neural basis of visual perception," in *International Encyclopedia of the Social & Behavioral Sciences*, Neil J. Smelser and Paul B. Baltes, Eds., pp. 16269–16274. Pergamon, Oxford, 2001.

[13] T.Y.W. Choi, P.A. Merolla, J.V. Arthur, K.A. Boahen, and B.E. Shi, "Neuromorphic implementation of orientation hypercolumns," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 52, no. 6, pp. 1049–1060, 2005.

[14] A.A. Lazar and L.T. Toth, "Perfect recovery and sensitivity analysis of time encoded bandlimited signals," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 51, no. 10, pp. 2060–2073, oct. 2004.

[15] C. Posch, "Bio-inspired vision," *Journal of Instrumentation*, vol. 7, no. 1, pp. C01054, 2012.

[16] A. A. Lazar, E. K. Simonyi, and L. T. Toth, "Time encoding of bandlimited signals, an overview," in *Proceedings of the Conference on Telecommunication Systems, Modeling and Analysis*, Nov. 2005.

[17] Hans G. Feichtinger, José Príncipe, José Luis Romero, Alexander Singh Alvarado, and Gino Angelo Velasco, "Approximate reconstruction of bandlimited functions for the integrate and fire sampler," *Advances in Computational Mathematics*, vol. 36, no. 1, pp. 67–78, 2012.

[18] A.A. Lazar and E.A. Pnevmatikakis, "Video time encoding machines," *IEEE Transactions on Neural Networks*, vol. 22, no. 3, pp. 461–473, march 2011.

[19] Khaled Masmoudi, Marc Antonini, and Pierre Kornprobst, "Streaming an image through the eye: The retina seen as a dithered scalable image coder," *Signal Processing: Image Communication*, vol. 28, no. 8, pp. 856–869, 2013.

[20] Marek Miskowicz, "Send-on-delta concept: An event-based data reporting strategy," *Sensors*, vol. 6, no. 1, pp. 49–63, 2006.

[21] Botond Roska, Alyosha Molnar, and Frank S. Werblin, "Parallel Processing in Retinal Ganglion Cells: How Integration of Space-Time Patterns of Excitation and Inhibition Form the Spiking Output," *J Neurophysiol*, vol. 95, no. 6, pp. 3810–3822, 2006.

[22] Christoph Posch and Daniel Matolin, "Sensitivity and uniformity of a 0.18μm cmos temporal contrast pixel array," in *International Symposium on Circuits and Systems*, 2011.

[23] Teresa Serrano-Gotarredona and Bernabé Linares-Barranco, "A 128$\times$128 1.5% Contrast Sensitivity 0.9% FPN 3 μs Latency 4 mW Asynchronous Frame-Free Dynamic Vision Sensor Using Transimpedance Preamplifiers," *J. Solid-State Circuits*, vol. 48, no. 3, pp. 827–838, 2013.

[24] A. A. Borovkov, *Mathematical Statistics*, Gordon and Breach Sciences Publishers, Amsterdam, 1998.