https://doi.org/10.2352/EI.2024.36.7.ISS-289

Joint Parameter Estimation for Event-based Vision Sensor Characterization

Xiaozheng Mou, Rui Jiang, Wei Zhang, Menghan Guo, Bo Mu, Andreas Suess;
OMNIVISION, Santa Clara CA/USA

Abstract

This paper proposes a pixel-wise parameter estimation framework for Event-based Vision Sensor (EVS) characterization. Using an ordinary differential equation (ODE) based pixel latency model and an autoregressive Monte-Carlo noise model, we first identify the representative parameters of EVS. The parameter estimation is then formulated as an optimization problem to minimize the measurement-prediction error for both pixel latency and event firing probability. Finally, the effectiveness and accuracy of the proposed framework are verified by comparison of synthetic and measured event response latency as well as firing probability as function of temporal contrast (so-called S-curves).

Introduction

The Event-based Vision Sensors (EVS) [1–5], sometimes also referred to as Dynamic Vision Sensors (DVS), capture our world from a completely different perspective compared to classical CMOS Image Sensors (CIS). In EVS, each pixel independently detects whether the logarithmic luminance changes beyond defined relative thresholds in an asynchronous manner (Fig. 1a). This enables sparse, low-latency, and low-power image acquisition. This paradigm change requires novel models, characterization techniques as well as model parameter estimation.

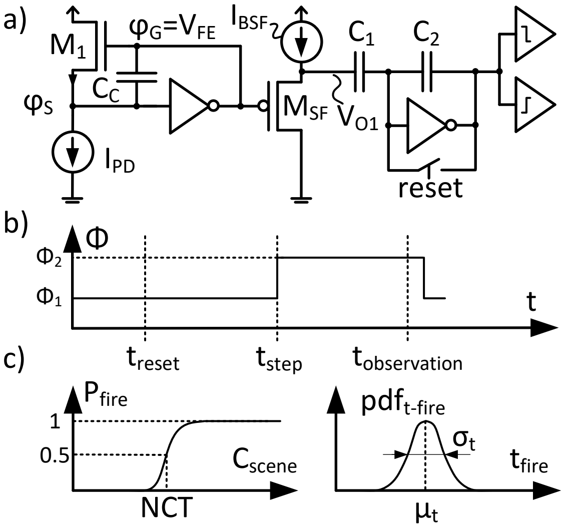

Figure 1. *a) analog EVS pixel circuit schematic, b) temporal contrast measurement timing diagram, and c) fire probability "S-curve" and timestamp distributions [10]. Nominal contrast threshold (NCT) is the scene contrast required to yield* 50 % *trigger probability.*

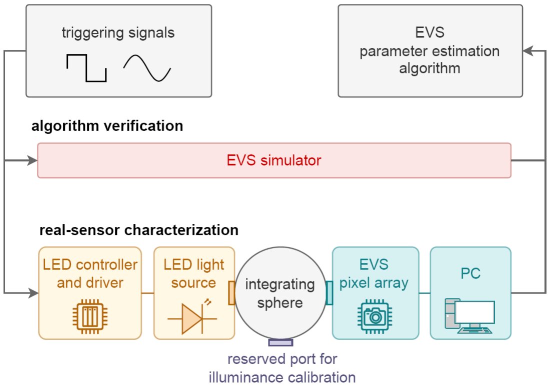

Figure 2. *The proposed setup. The parameter estimation is formulated as an optimization problem to minimize the measurement-prediction error.*

As EVS sensors are novel, so is their characterization. Commonly used characterization methods measure e.g. the average time from a contrast step to an event trigger or the average event trigger probability for a given temporal contrast step (Fig. 1 b & c). These methods, unfortunately, provide no predictive value as the resulting sample averages depend largely on measurement conditions such as sensor contrast threshold, temporal contrast step, or timing aspects such as the time used to determine whether an event was observed. Therefore, a model-based approach is needed. Simulators to emulate event sensors have been published in recent years [6, 7], however, these describe sensors only phenomenologically and do lack careful calibration against simulation and characterization. Only recently, a physically based model has been proposed and reasonable resemblance to simulated and measured observations was demonstrated [8, 10]. This paper builds upon the physical model published in [10]. It analyzes the relevance of the employed model parameters, determines key parameters, and then promotes an optimization-based model parameter estimation method.

Figure 2 illustrates the proposed experimental characterization setup EVS used for parameter estimation. Our sensor is modeled based on an ordinary differential equation (ODE) pixel latency model and an autoregressive Monte Carlo noise model [10]. The model was derived using common circuit modeling techniques. Comparisons between measurements, simulations, and model demonstrated good resemblance and thus the validity of the approach [10].

In this work, we deduce that the most significant pixel model parameters are the temporal contrast threshold (C_{TH}), the coupling capacitance (C_{C}), and source follower bias current (I_{BSF}).

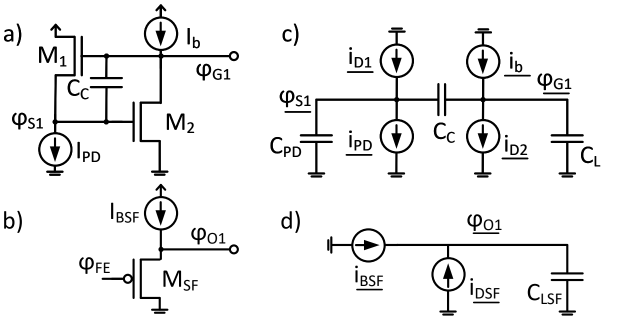

Figure 3. *Simplified log-amplifier schematic a) and SF schematic b) and their respective small signal circuits in c) and d), respectively.*

Other model parameters are assumed to match our prior expectations from design. Our goal in characterization is then to find the pixel-wise parameter values that can best align the model with the real sensor data for both event response latency and event firing probability. We formulate this model parameter estimation as an optimization problem and then minimize the loss function based on the relative error between the model prediction and the actual measurement weighted in the L2 norm. We jointly account for pixel latency and event firing probability. We validate the quality of the parameter extraction method using the simulation model as this provides a ground truth. Finally, we apply the method to real measurements.

Circuit Modelling

Assuming that all transistors operate in saturation and weak inversion and that the inverting amplifier in the feedback loop of the logarithmic amplifier has an idealized affine-linear behavior, and neglecting its additional pole, we can model the large signal response through simple ordinary differential equations [10]. The logarithmic amplifier follows:

$$\frac{\mathrm{d}\Delta V_{\mathrm{FE}}(t)}{\mathrm{d}t} + \frac{A \cdot I_{\mathrm{D1}}(t_0)}{[1+A] \cdot C_{\mathrm{C}}} \cdot \exp\left(\frac{\zeta + A}{A \cdot \zeta \cdot V_{\mathrm{T}}} \cdot \Delta V_{\mathrm{FE}}(t)\right) = \frac{A}{[1+A] \cdot C_C} \cdot I_{\mathrm{PD}}, \quad (1)$$

where ΔV_{FE} describes the relative signal change at the output of the logarithmic amplifier with respect to the signal at the last sampled reference luminance level. A is the open-loop gain of the inverting amplifier, I_{D1} is the current through the transistor whose source is connected to the photodiode operating the linear-to-logarithmic conversion. I_{PD} is the photocurrent, C_{C} is the feedback capacitance between the input and output of the log-amplifier, ζ is the transistor subthreshold slope parameter and V_{T} is the thermal voltage.

The source-follower is designed to avoid kick-back from the switched capacitor circuit performing the difference detection. It is modeled as:

$$\frac{\mathrm{d}\Delta V_{\mathrm{O1}}(t)}{\mathrm{d}t} = \frac{1}{C_{\mathrm{LSF}}} \cdot \Bigg[I_{\mathrm{BSF}} - I_{\mathrm{SF}}(t_0) \cdot \exp\left(\frac{\zeta \cdot \Delta V_{\mathrm{O1}}(t)}{\zeta \cdot V_{\mathrm{T}}} - \frac{\Delta V_{\mathrm{FE}}(t)}{\zeta \cdot V_{\mathrm{T}}}\right)\Bigg], \quad (2)$$

with ΔV_{O1} being the relative output signal change, I_{SF} is the current through the source-follower transistor and I_{BSF} is its bias current. C_{LSF} is the load seen by the source-follower.

It can be shown that the logarithmic amplifier has an analytical solution to a step-response in photocurrent $I_{\text{photo-0}} \rightarrow I_{\text{photo-1}}$:

$$\Delta v = \frac{\Delta V_{\mathrm{FE}}}{V_{\mathrm{T}} \cdot \frac{A \cdot \zeta}{\zeta + A}} = \ln\left(\frac{I_{\text{photo-1}}}{I_{\text{photo-0}}} \cdot \frac{1}{1 + C_{\text{scene}} \cdot \exp[-[t - t_0]/\tau]}\right), \quad (3)$$

with the time-constant $\tau = \frac{1+A}{A} \cdot \frac{A \cdot \zeta}{\zeta + A} \cdot \frac{V_{\mathrm{T}} \cdot C_{\mathrm{C}}}{I_{\text{photo-1}}}$. Assuming that the bandwidth limitation of the log-amplifier is more significant than the source-follower, an analytical response can be found as:

$$\Delta V_{\mathrm{O1}}(t) \approx \frac{1}{\zeta} \cdot \Delta V_{\mathrm{FE}}(t) \quad (4)$$

Figure 3 depicts the small signal equivalent circuits used to derive the noise model. For small photocurrents corresponding to low-light conditions, the shot noise of the photocurrent and transistors dominates over e.g. flicker noise assuming the operating range of interest is not at very low frequencies. We thus approximate the noise to follow shot noise processes resulting in the following autocorrelation functions $R_{\mathrm{FE,\,FE}}$, and $R_{\mathrm{O1,\,O1}}$ which are injected at the input and output of the source-follower circuit [10]:

$$R_{\mathrm{FE,\,FE}}(\Delta t) = \zeta \cdot \frac{k_{\mathrm{B}} \cdot T}{C_{\mathrm{C}}} \cdot e^{-\frac{|\Delta t|}{C_{\mathrm{C}}/g_{\mathrm{mG1}}}} + \zeta \cdot \frac{k_{\mathrm{B}} \cdot T}{C_{\mathrm{PD}} \cdot C_{\mathrm{C}} \cdot \frac{C_{\mathrm{C}}+C_{\mathrm{L}}}{[C_{\mathrm{C}}+C_{\mathrm{PD}}]^2}} \cdot e^{-\frac{|\Delta t|}{C_{\mathrm{PD}} \cdot \frac{C_{\mathrm{C}}+C_{\mathrm{L}}}{C_{\mathrm{C}}} / g_{\mathrm{mG2}}}} \quad (5)$$

$$R_{\mathrm{O1,\,O1}}(\Delta t) = \frac{k_{\mathrm{B}} \cdot T}{C_{\mathrm{LSF}}} \cdot e^{-\frac{|\Delta t|}{C_{\mathrm{LSF}}/g_{\mathrm{mG-SF}}}} \quad (6)$$

Here, k_{B} is the Boltzmann constant, C_{PD} is the photodiode capacitance and C_{L} is the load capacitance seen by the inverting amplifier. $g_{\mathrm{mG1}} = \frac{I_{\mathrm{D1}}}{\zeta \cdot V_{\mathrm{T}}}$, $g_{\mathrm{mG2}} = \frac{I_{\mathrm{b}}}{\zeta \cdot V_{\mathrm{T}}}$, and $g_{\text{mG-SF}} = \frac{I_{\mathrm{BSF}}}{\zeta \cdot V_{\mathrm{T}}}$ are the small-signal gains of the log-amplifier transistor, the transistor of the inverting amplifier driven with bias I_{b} and the source follower. In our event vision simulator, these autocorrelation functions are used to create Monte Carlo trials using autoregressive processes. This allows the synthesis of training data for algorithm development from ground truth high-speed video data. Furthermore, this enables the evaluation of parameter extraction methodologies as conversely to actual measurements, the underlying model parameters are known apriori.

Joint Parameter Estimation

From Eq. 1 it can be seen that for $A \gg \zeta$ changes in A and C_{C} are indistinguishable, so we assume A to follow the expectation from design and focus on estimating C_{C}. Assuming a settled starting condition $I_{\mathrm{SF}}(t_0) = I_{\mathrm{BSF}}$ such that there is no displacement current flowing through C_{LSF}, one can see from Eq. 2 that the impact of I_{BSF} and C_{LSF} are indistinguishable. Thus we focus on estimating I_{BSF}. We empirically determined that the impact of

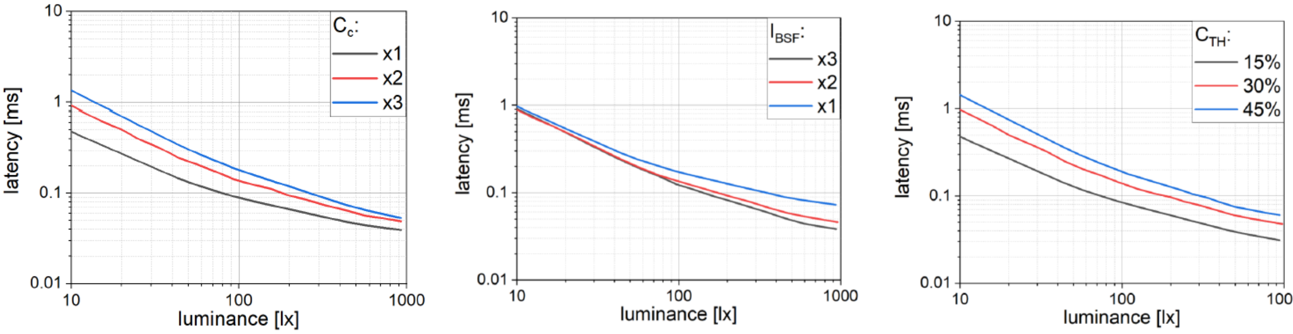

Figure 4. Pixel response latency vs. luminance as function of C_C, I_{BSF} and C_{TH}.

ζ is minor resulting in C_{TH}, C_{C}, and I_{BSF} being the most significant parameters which we aim to estimate. Figure 4 shows latency variability as function of these three key parameters.

Given the effective 1.5 bit output of an EVS pixel (*'increase'*, *'decrease'*, or *'no event'*), we can essentially only observe triggering probability distributions or time-stamp distributions as well as moments of these distributions. Reference [10] focused on utilizing average time stamps to derive model parameters. Assuming that the log amplifier is the latency bottleneck, the analytical step-response solutions can be rearranged to determine the noise-free time-stamp for a temporal contrast step:

$$t_{\text{event}} = \tau \cdot \ln\left(\frac{C_{\text{scene}} \cdot [1 + C_\infty]}{C_{\text{scene}} - C_\infty}\right), \tag{7}$$

with $C_{\text{scene}} = \frac{I_{\text{photo-1}} - I_{\text{photo-0}}}{I_{\text{photo-0}}}$ and $C_\infty = \exp\left[\frac{V_{\text{threshold}}}{V_{\text{T}} \cdot G}\right] - 1$, the comparator threshold $V_{\text{threshold}}$, and the difference detector gain G [10]. It was shown in [10] that at sufficient temporal contrast, this noise-free time-stamp is close to the noise-affected time-stamp. From Eq. 7, it can be seen that in order to derive τ and C_∞ a combination of contrast changes C_{scene} as well as reference level changes $I_{\text{photo-0}}$ are required to yield independent measurements.

Conversely to [10] in this work we not only utilize the trigger latency but also the event trigger probability to extract model parameters. This helps to balance the matching of extracted parameters for both key observables. However, this requires the model to consider the stochastic properties of the EVS pixel which we model using Monte-Carlo trials.

In this work we used grid search optimization per pixel:

$$C^*_{\text{C-}i,j}, I^*_{\text{BSF-}i,j}, C^*_{\text{TH-}i,j} = \underset{C_{\text{C}}, I_{\text{BSF}}, C_{\text{TH}}}{\arg\min}\; J_{i,j}, \tag{8}$$

with the per-pixel cost function $J_{i,j}$:

$$\begin{aligned} J_{i,j} = \Bigg\{ & w \sum_{\forall k} \left\| \frac{\tau^k_{i,j} - \tau_{\text{model}}(C_{\text{C}}, I_{\text{BSF}}, C_{\text{TH}})}{\tau^k_{i,j}} \right\|^2 \\ & + (1-w) \sum_{\forall k} \left\| \frac{NCT^k_{i,j} - NCT_{\text{model}}(C_{\text{C}}, I_{\text{BSF}}, C_{\text{TH}})}{NCT^k_{i,j}} \right\|^2 \Bigg\}. \end{aligned} \tag{9}$$

w denotes the weight balancing the contributions to the loss function between pixel response latency and NCT. We empirically found that $w = 0.5$ yields good results for our range of interest. τ_{model} and NCT_{model} are the mathematical models for latency and NCT estimated through Monte-Carlo trials. k are all measurements mutually distinct in scene contrast and/or nominal luminance level and i, j are the pixel coordinates. The cost function uses normalized latency and NCT in order to achieve good performance for a large dynamic range.

Method Validation

To validate the effectiveness of the proposed cost function, we employ camera simulation for which we have apriori knowledge of the ground truth parameters. Figure 5 shows the loss curves in which for each curve only one of three parameters is

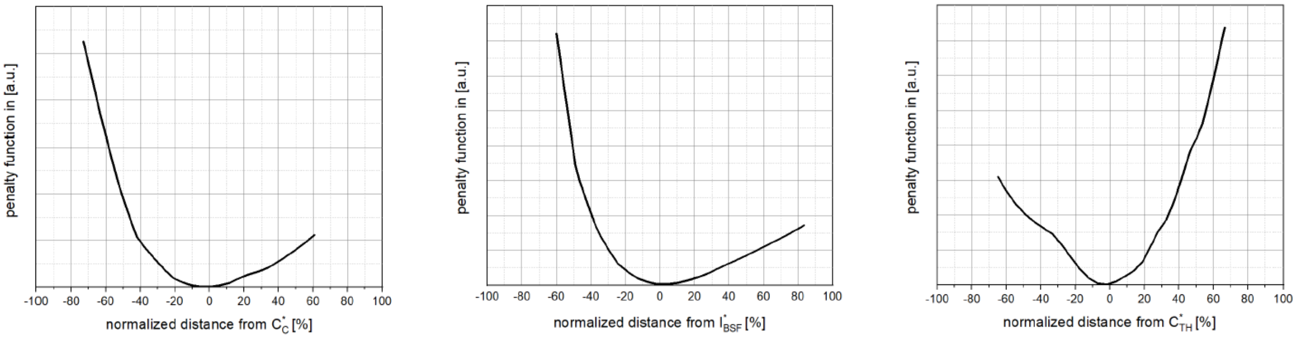

Figure 5. Loss curves of the proposed objective function with one free parameter and the other two fixed.

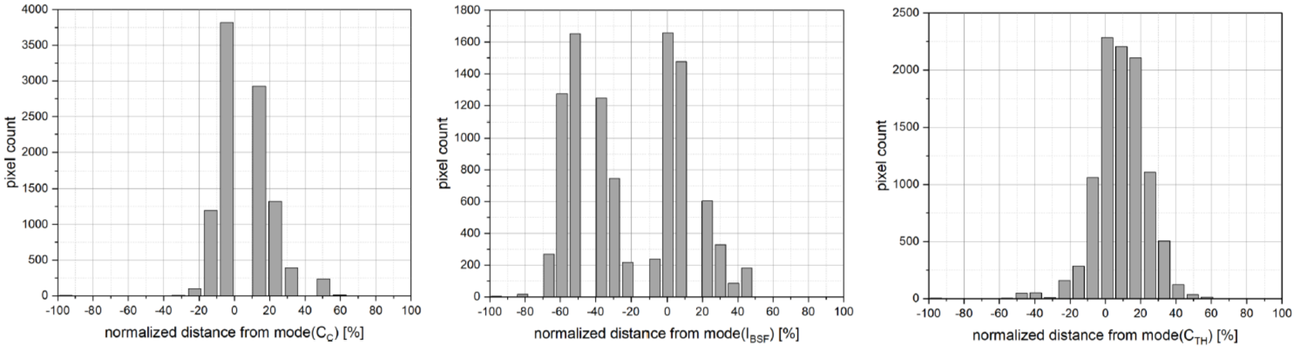

Figure 6. *Pixel-wise distributions of estimated parameters C_C, I_{BSF}, C_{TH}. The parameters here are normalized to the expected value.*

swept from the optimality condition. In general, we have no guarantee of convexity. Thus we empirically validated convergence in the parameter range of our interest. Furthermore, global optimization techniques help to avoid getting stuck in a local optima. We do employ grid search. It can be observed that the slope of the cost function can be very small. Thus noise can have a considerable impact on the parameter estimation and a sufficient amount of Monte-Carlo samples ought to be used to mitigate this issue. Again this can be empirically validated using camera simulation for a given expected parameter range of interest.

Experimental Results

The proposed parameter extraction methodology (Eq. 9) is now applied to measurements using the sensor published in [9]. A region of interest (ROI) of 100×100 in the center of the image plane is selected for characterization. In order to yield independent measurements, we generate various temporal contrast steps as well as varying reference luminance levels L_{ref}:

$$C_{\text{scene}} \in \begin{cases} \{[100\,\%, 110\,\%, 120\,\%, ..., 200\,\%\} & \text{for pixel latency} \\ [2\,\%, 4\,\%, 6\,\%, ..., 40\,\%] & \text{for NCT} \end{cases} \tag{10}$$

and

$$L_{\text{ref}} \in \{15\,\text{lx}, 130\,\text{lx}, 310\,\text{lx}, 700\,\text{lx}, 910\,\text{lx}\}. \tag{11}$$

Figure 6 shows the estimated parameter distributions. C_C and C_{TH} roughly resemble normal distributions, whereas I_{BSF} exhibits large variability and shows an almost bimodal distribution. This is because I_{BSF} has a limited impact in the L_{ref} range. The latency of the source follower only dominates the overall latency at $\gtrapprox 1000\,\text{lx}$ [9].

The extracted model parameters were used to plot expected latency curves and sample trigger probability curves vs. temporal scene contrast. The resemblance of the model and measurements is depicted in Figure 7 for selected pixel. For low luminance levels, the latence model slightly overestimates the latency. This can be improved either by adjusting the cost function to place more emphasis on this region or by introducing more degrees of freedom than the simplified single-pole approximation given by the ordinary differential equation model. The firing probability curves also show reasonable resemblance vs. measurements

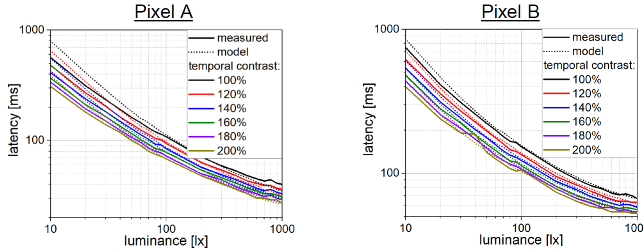

Figure 7. *Pixel latency comparison between the measurement and our model using the estimated parameters under different illuminances and temporal contrast for two sampled pixels.*

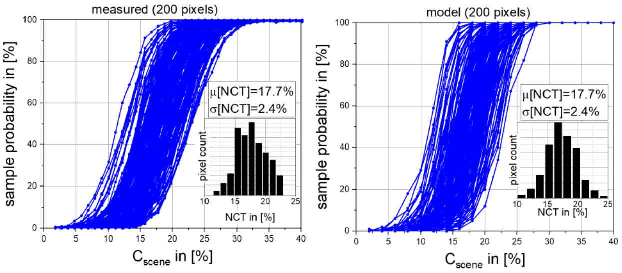

Figure 8. *S-curves comparison between the measurement and our model using the estimated parameters for 200 sampled pixels.*

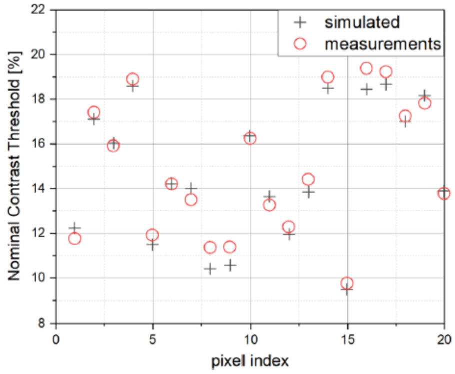

Figure 9. *NCT comparison between the measurement and our model using the estimated parameters for 20 sampled pixels.*

as can be seen in Figure 8 where a population of 200 measured curves is compared to 200 model curves. A slight change in slope of the S-curve can be observed which can be attributed to the simplistic noise model. We compute the NCT values of 20 pixels based on the extracted model parameters and overlay these with their measured values and also find a good resemblance in Figure 9.

Conclusion

This work presents perspectives on parameter extraction of a physical EVS pixel model. Conversely to [10] a joint optimization approach to account for pixel latency and firing probability is employed. The optimization incorporates the stochastic nature of the pixel operation using Monte Carlo trials. The model was validated using the device simulator as the ground truth model parameters are known apriori. The model was then used to extract parameters from measured devices and a comparison of latency and estimated firing probabilities is given for a range of measurement conditions proving that the model has predictive value.

References

[1] R. Benosman, "Event Computer Vision 10 years Assessment: Where We Came From, Where We Are and Where We Are Heading To," in 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2021.

[2] G. Gallego, T. Delbruck, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. Davidson, Joerg Conradt, K. Daniilidis, and D. Scaramuzza, "Event-based vision: A survey," in IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020.

[3] H. E. Ryu, "Industrial DVS Design; Key Features and Applications," in 2019 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR), 2019.

[4] T. Finateu, A. Niwa, D. Matolin, K. Tsuchimoto, A. Mascheroni, E. Reynaud, P. Mostafalu, F. Brady, L. Chotard, F. LeGoff, H. Takahashi, H. Wakabayashi, Y. Oike and C. Posch, "A 1280x720 Back-Illuminated Stacked Temporal Contrast Event-Based Vision Sensor with 4.86um Pixels, 1.066GEPS Readout, Programmable Event-Rate Controller and Compressive Data-Formatting Pipeline," in IEEE International Solid-State Circuits Conference, 2020.

[5] P. Lichtensteiner and T. Delbruck, “A 64x64 AER logarithmic temporal derivative silicon retina,” Research in Microelectronics and Electronics, 2005 PhD, 2005, pp. 202-205

[6] D. Gehrig, M. Gehrig, J. Hidalgo-Carrio, and D. Scaramuzza, "Video to Events: Recycling Video Datasets for Event Cameras," in 2020 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR), 2020.

[7] Y. Hu, S.C. Liu and T. Delbruck, "v2e: From video frames to realistic DVS event camera streams," in 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2021.

[8] X. Mou, K. Feng, A. Yi, S. Wang, H. Chen, X. Hu, M. Guo, S. Chen, A. Suess, "Accurate event simulation using high-speed video", in Proc. IS&T Int. Symp. on Electronic Imaging: Imaging Sensors and Systems, pp. 242-1 - 242-6, 2022.

[9] M. Guo, S. Chen, Z. Gao, W. Yang, P. Bartkovjak, Q. Qin, X. Hu, D. Zhou, Q. Huang, M. Uchiyama, Y. Kudo, S. Fukuoka, C. Xu, H. Ebihara, X. Wang, P. Jiang, B. Jiang, B. Mu, H. Chen, J. Yang, T.J. Dai, A. Suess, "A three-wafer-stacked hybrid 15-MPixel CIS + 1-MPixel EVS With 4.6-GEvent/s readout, in-pixel TDC, and on-chip ISP and ESP function", in IEEE Journal of Solid-State Circuits, vol. 58, no. 11, pp. 2955-2964, Nov. 2023.

[10] A. Suess, M. Guo, R. Jiang, X. Mou, Q. Huang, W. Yang, S. Chen, "Physical modelling and parameter extraction for event based vision sensors", in Proc. IISS Int. Image Sensor Workshop (IISW), pp. R5.5, 2023.