EVENT CAMERA CALIBRATION OF PER-PIXEL BIASED CONTRAST THRESHOLD

AUTHOR ACCEPTED VERSION*

Ziwei Wang
Systems Theory and Robotics Group
Australian National University
ACT, 2601, Australia
ziwei.wang1@anu.edu.au

Yonhon Ng
Systems Theory and Robotics Group
Australian National University
ACT, 2601, Australia
yonhon.ng@anu.edu.au

Pieter van Goor
Systems Theory and Robotics Group
Australian National University
ACT, 2601, Australia
pieter.vangoor@anu.edu.au

Robert Mahony
Systems Theory and Robotics Group
Australian National University
ACT, 2601, Australia
robert.mahony@anu.edu.au

December 18, 2020

ABSTRACT

Event cameras output asynchronous events to represent intensity changes with a high temporal resolution, even under extreme lighting conditions. Currently, most of the existing works use a single contrast threshold to estimate the intensity change of all pixels. However, complex circuit bias and manufacturing imperfections cause biased pixels and mismatch contrast threshold among pixels, which may lead to undesirable outputs. In this paper, we propose a new event camera model and two calibration approaches which cover event-only cameras and hybrid image-event cameras. When intensity images are simultaneously provided along with events, we also propose an efficient online method to calibrate event cameras that adapts to time-varying event rates. We demonstrate the advantages of our proposed methods compared to the state-of-the-art on several different event camera datasets.

1 Introduction

An event camera is a bio-inspired silicon retina that can provide high pixel bandwidth, high dynamic range and low latency measurement of image intensity changes [15, 21]. In contrast to the traditional cameras, which generate intensity images at a fixed frequency, event cameras are based on the Dynamic Vision Sensor (DVS), which allows them to generate asynchronous events (ON or OFF signals) to represent changes of intensity. These events are generated by pixel-level brightness changes, which contain timestamps at microsecond resolution with a High Dynamic Range (HDR) [27, 31]. The Dynamic and active pixel vision sensor (DAVIS) incorporates the DVS and a synchronous frame-based active pixel sensor (APS) [3], which also generates intensity images simultaneously. Another higher resolution (480×360) event camera that can provide both grayscale intensity image along with events is the HVGA ATIS sensor [16].

Due to the properties of an event camera, it is particularly suitable for applications where fast robotic vision is required. For example, an event camera was used in simultaneous localization and mapping (SLAM) [7, 11, 12, 28], feature detection [27, 24], feature tracking [9, 1], visual odometry [17, 13], optical flow [2], and image reconstruction [20, 19, 24].

In DVS, contrast threshold defines the logarithmic intensity change that will trigger an event. It is commonly assumed that the contrast threshold is a constant (typically 10%) for all pixels [24, 22, 11, 12]. However, due to a number of factors like event generation speed, intensity change, circuit noise, and manufacturing imperfections, the actual contrast threshold at each pixel may be different from the desired contrast threshold. In addition, undesirable junction leakage in the differencing amplifier and biased circuit cause noisy background positive events that are not correlated to the actual visual input [13, 30, 5].

As previously reported, the effective contrast threshold can vary between 10-15% [25, 15, 3, 14]. Algorithms that correctly model the variability of the inter-pixel contrast threshold and bias can achieve an improvement in the accuracy of their output.

This paper presents novel calibration methods to correctly recover the different per-pixel contrast threshold for an event camera. The contribution of this paper are:

- Analyse the source of contrast threshold variation at circuit level and introduce four types of pixels based on their different behaviours compared to the commonly used model.
- Introduce a new event generation model that can represent inter-pixel biases and contrast thresholds.
- Propose both online and offline linear regression methods for event-only cameras (*e.g.*, DVS128 cameras [15], DVS-Gen2 cameras [26]) and event cameras with intensity images (*e.g.*, DAVIS and HVGA ATIS sensor).

The rest of this paper is organised as follow: we briefly introduce the prior works on estimating the contrast threshold in Section 2. In Section 3, we describe the traditional event generation model and our new model. We also propose four types of special pixels in this section. Section 4 describes our proposed methods. Then, Section 5 shows intensity image reconstruction results to verify our online and offline calibration methods. Section 6 concludes the paper.

2 Related Work

In the early design of event-based camera software *jAER*, a filter was implemented to remove the background noisy positive events caused by transistor junction leakage or noise [8]. For each pixel, if no event happens at the neighbouring patch recently, the event that is detected at the pixel is potentially noise that does not correlated to actual intensity changes. The filter directly removes all isolated events without the support of their 8 neighbouring pixels within the previous period T, which largely removes the background noisy events [8]. The DAVIS camera and related software we used in our experiments have already suppressed the noise discussed in the paper, however, it still leaves distinct pixels that require special attention.

Some researchers worked on improving the quality of the output of event cameras by increasing the threshold sensitivity, which would increase the intensity resolution and alleviate the quantization effects in fine texture reconstruction and recognition [5, 30, 25, 19]. For example, Serrano *et al.* increased the threshold sensitivity from 10% to 1.5% by designing a novel transimpedance pre-amplification [25]. Yang *et al.* also proposed an in-pixel asynchronous delta modulator (ADM) to replace the self-timed reset (STR) in event cameras, which increases the threshold sensitivity to 1%. However, these designs still cannot solve the problem of threshold mismatch and bias [30]. Recently, Pan *et al.* obtained sharper reconstruction images by combining motion blurred intensity image and events. Their work computed an optimal (single) contrast threshold across every pixel with an energy function minimization method.

Our work is most related to [5], because they accounted for the variance of contrast threshold between pixels. To improve the real-time video decompression performance, Brandli *et al.* estimated the unbalanced ON and OFF threshold and the mismatched threshold for each pixel by computing the direct integration error of events between two intensity image frames. The algorithm avoids cumulating the threshold error by continuously resetting ON and OFF threshold values for each pixel. However, the method still has drawbacks: 1) The model lacks theoretical justification, 2) The method heavily relies on intensity images. Therefore the method cannot calibrate event cameras which can only generate events (*e.g.*, DVS128 cameras [15], DVS-Gen2 cameras [26]).

3 Problem Formulation

3.1 Event Generation Model and Notation

Consider a continuous image intensity function

$$I : \mathbb{N}^2 \times \mathbb{R} \to \mathbb{R}^+ \\ (x, y), t \mapsto I((x, y), t), \tag{1}$$

where the inputs (x, y) are some pixel coordinates, and t is the time of measurement. Define the continuous log image intensity function as $L := \log \circ I$. Events are modelled as $e_i^{\boldsymbol{p}}$ at given pixel coordinates $\boldsymbol{p} = (\boldsymbol{p}_x, \boldsymbol{p}_y)^T$ consisting of a timestamp $t_i^{\boldsymbol{p}}$ and a polarity $\sigma_i^{\boldsymbol{p}}$ such that

$$e_i^{\boldsymbol{p}} = (t_i^{\boldsymbol{p}}, \sigma_i^{\boldsymbol{p}}). \tag{2}$$

The polarity $\sigma_i^{\boldsymbol{p}}$ of an event $e_i^{\boldsymbol{p}}$ is either $+1$ or -1, and we model the triggering of an event by

$$\sigma_i^{\boldsymbol{p}} = \begin{cases} +1, & \Delta L^{\boldsymbol{p}}(t_i^{\boldsymbol{p}}) \geq c^{\boldsymbol{p}}(t_i^{\boldsymbol{p}}) + b^{\boldsymbol{p}}(t_i^{\boldsymbol{p}}), \\ -1, & \Delta L^{\boldsymbol{p}}(t_i^{\boldsymbol{p}}) \leq -c^{\boldsymbol{p}}(t_i^{\boldsymbol{p}}) + b^{\boldsymbol{p}}(t_i^{\boldsymbol{p}}), \end{cases} \tag{3}$$
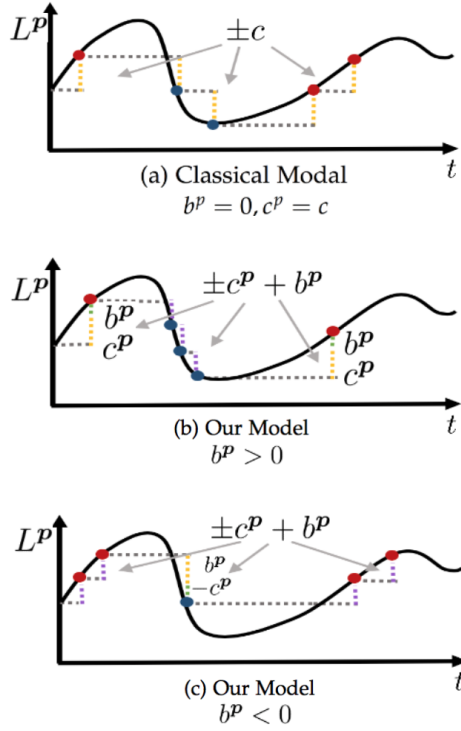
Figure 1: Event Generation Model. The classical model was used in [10, 27, 24].

where c^p is the (time-varying) pixel-wise threshold value, b^p is the (time-varying) pixel-wise bias value, and $\Delta L(t_i^p)$ is defined by

$$\Delta L^p(t_i^p) = L^p(t_i^p) - L^p(t_{i-1}^p). \tag{4}$$

Equation (2) shows the proposed model for discrete time event signals, which allows for different biases and threshold values to apply to each pixel. A new event $e_i^p = (t_i^p, \sigma_i^p)$ with positive polarity ($\sigma_i^p = +1$) is triggered when the log intensity changes $\Delta L^p(t_i^p)$ is larger than the biased contrast threshold at timestamp t_i, and vice versa.

3.2 Special Pixels

Most of the previous works model the contrast threshold as being consistent across all pixels in the event camera and ignore biased pixels altogether [24, 22, 11, 12]. This can be seen as a special case of our more general model in (5) and can be written as

$$\sigma_i^p = \begin{cases} +1, & \Delta L^p(t_i^p) \geq c, \\ -1, & \Delta L^p(t_i^p) \leq -c, \end{cases} \tag{5}$$

where $\Delta L^p(t_i^p)$ is defined as in (4).

If we ignore the mismatch in contrast threshold and the circuit bias, the classical model of event generation in Figure 1 will generate four types of special pixels, which will affect the performance of image reconstruction and video compression [6, 23, 5], event-based feature tracking [13, 1]. Special pixels are shown in Figure 2.

- *Hot Pixels* are pixels that have lower threshold values than the assumed value. This means that both positive and negative events are generated more often than expected.
- *Cold Pixels* are pixels that have higher threshold values than the assumed value. This means that both positive and negative events are generated less often than expected.
- *Warm Pixels* are pixels that have negative bias values. This means positive events are triggered more often, and negative events are triggered less often than expected.
- *Cool Pixels* are pixels that have positive bias values. This means negative events are triggered more often, and positive events are triggered less often than expected.

	Hot Pixel $c^{\boldsymbol{p}} < 10\%$	Cold Pixel $c^{\boldsymbol{p}} > 10\%$	Warm Pixel $b^{\boldsymbol{p}} < 0$	Cool Pixel $b^{\boldsymbol{p}} > 0$
Positive Event Generation Model	Postivie Event	No Event	Postivie Event	No Event
Negative Event Generation Model	Negative Event	No Event	No Event	Negative Event

Figure 2: Four Types of Special Pixels

4   Proposed Methods

4.1   Offline Calibration: Hybrid Case (Event + Intensity)

Because each event could cause reset bias in the reference log intensity level [30], the direct integration with constant contrast threshold as in (5) cannot perfectly represent the actual intensity change. For DAVIS event cameras, which also produce intensity images as well as event data, we introduce a calibration method that uses both data sources.

Since the biased event-based sensor DVS does not affect the APS circuit which generates intensity images, the corresponding intensity images provide reliable and independent intensity references [5]. Therefore for any pixel $\boldsymbol{p}$, given two log intensity images $L^{\boldsymbol{p}}(T_k)$ and $L^{\boldsymbol{p}}(T_{k+d})$ recorded at times T_k and T_{k+d} respectively, the following approximation holds

$$\sum_{t_i^{\boldsymbol{p}} \in (T_k, T_{k+d}]} \begin{bmatrix} \sigma_i^{\boldsymbol{p}} & |\sigma_i^{\boldsymbol{p}}| \end{bmatrix} \begin{bmatrix} c^{\boldsymbol{p}}(t_i^{\boldsymbol{p}}) \\ b^{\boldsymbol{p}}(t_i^{\boldsymbol{p}}) \end{bmatrix} = \Delta_k^{k+d} L^{\boldsymbol{p}}, \tag{6}$$

where $\Delta_k^{k+d} L^{\boldsymbol{p}} = L^{\boldsymbol{p}}(T_k) - L^{\boldsymbol{p}}(T_{k+d})$.

Since this approximation holds for any choice of k and d, we may stack these equations into the following overdetermined linear system. For any pixel coordinates $\boldsymbol{p}$, and any choice of k, d and n,

$$Ax = z, \tag{7}$$

$$\begin{bmatrix} \sum_{t_i^{\boldsymbol{p}} \in (T_k, T_{k+d}]} \begin{bmatrix} \sigma_i^{\boldsymbol{p}} & |\sigma_i^{\boldsymbol{p}}| \end{bmatrix} \\ \vdots \\ \sum_{t_i^{\boldsymbol{p}} \in (T_{k+(n-1)d}, T_{k+nd}]} \begin{bmatrix} \sigma_i^{\boldsymbol{p}} & |\sigma_i^{\boldsymbol{p}}| \end{bmatrix} \end{bmatrix} \begin{bmatrix} c^{\boldsymbol{p}} \\ b^{\boldsymbol{p}} \end{bmatrix} = \begin{bmatrix} \Delta_k^{k+d} L^{\boldsymbol{p}} \\ \vdots \\ \Delta_{k+(n-1)d}^{k+nd} L^{\boldsymbol{p}} \end{bmatrix}, \tag{8}$$

where $c^{\boldsymbol{p}}$ and $b^{\boldsymbol{p}}$ are assumed to be approximately constant (slowly time-varying) over the time period between T_k and T_{k+nd}.

An ordinary least-squares approach provides a solution to the overdetermined system in (8). In our offline calibration model *OffEI*, we estimate bias and contrast threshold for each pixel using raw images from the first frame k to the last frame $k + nd$ of a dataset. The number of sampling frames d may be adjusted based on the performance of the direct integration.

4.2 Online Calibration: Hybrid Case (Event + Intensity)

The self-timed reset (STR) event encoding mechanism in event cameras generates new events when the amplified change of log intensity is larger than the threshold of a comparator, which could lose events because of event queueing, comparators comparison and other factors [4, 15, 30]. [5] also shows that different scenes, lighting conditions and object moving speeds may lead to different contrast threshold values. Hence, we introduce an online calibration method to compute dynamic bias and threshold to adjust changing event rate and environment.
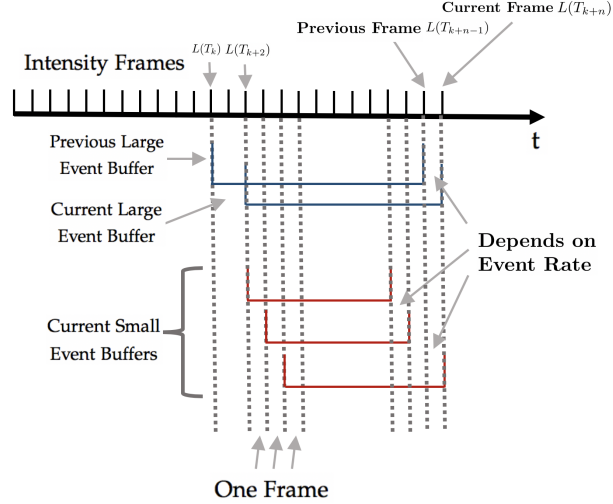

Figure 3: Event Buffer Model For Online Calibration Method *OnEI*. Each buffer contains a pre-set number of events and buffer length depends on event rate. The figure shows the current large buffer contains enough events and we start updating b^p and c^p. We divide the large buffer into many small buffers for each least square equation in (8). Before update, the oldest information between frame $L(T_k)$ and $L(T_{k+2})$ is removed in the large buffer and the buffer size is decreased by 1 frame.

We use a large event buffer and several small event buffers for linear regression in (8). We continuously add new events between $L(T_{k+n-1})$ and $L(T_{k+n})$ to the large buffer and increase the buffer length by 1 frame when a new intensity image $L(T_{k+n})$ arrives. When the large buffer contains enough events (*i.e.*, the total number of events in the buffer is larger than a pre-set value), we remove the events that occurred between the oldest image frames $L(T_k)$ and $L(T_{k+2})$ to decrease the buffer length by 1 frame. Then, we divide the large buffer into many small buffers. Each small buffer stores a certain amount of events which corresponds to a least square equation. We estimate new b^p and c^p for each pixel using (8). When the large buffer does not contain sufficient events, we will only add new events in the buffer, but not update b^p and c^p or remove oldest events. The time-varying length of two types of event buffer is based on the event rate: shorter buffer for higher event rate and longer buffer for lower event rate. This is illustrated in Figure 3. We assume the bias and contrast threshold are slowly time-varying. Therefore, a low-pass filter is implemented to achieve smoother updates.

4.3 Offline Calibration: Event-only Case

We propose a model to estimate contrast threshold c^p and bias b^p for pure event cameras, *e.g.*, DVS cameras, which can only produce event streams.

Without the intensity image as a reference input, we rely on assumptions about the scene and the resulting event stream:

- The gradient of the mean value of the cumulative sum of the event polarity over time is zero. That is, the intensity value at timestamp T_{k+d} is similar to the intensity value at timestamp T_k.

- The variance of the cumulative sum of the event polarity from its mean value is approximately equivalent across all pixels. That is, the scene is roughly equally textured, such that the excitation of each pixel is similar.

These assumptions apply for long sequences of data, while observing an equally textured scene.

Bias b^p Calibration: Compute b^p to remove the non-zero gradient, such that, for any pixel coordinates p,

$$\mathbb{E}\left[\sum\nolimits_{t_i^p \in (T_k, T_{k+d})} \left(c^p \sigma_i^p + b^p\right)\right] = 0. \qquad (9)$$

5

Contrast Threshold $c^{\boldsymbol{p}}$ Calibration: Compute $c^{\boldsymbol{p}}$ such that, for any two pixel coordinates $\boldsymbol{p}_1, \boldsymbol{p}_2$,

$$
\mathrm{Var}\left[\sum\nolimits_{t_i^{\boldsymbol{p}_1} \in (T_k, T_{k+d}]} \left(c^{\boldsymbol{p}_1} \sigma_i^{\boldsymbol{p}_1} + b^{\boldsymbol{p}_1}\right)\right]
$$
$$
= \mathrm{Var}\left[\sum\nolimits_{t_j^{\boldsymbol{p}_2} \in (T_k, T_{k+d}]} \left(c^{\boldsymbol{p}_2} \sigma_j^{\boldsymbol{p}_2} + b^{\boldsymbol{p}_2}\right)\right]. \tag{10}
$$

The bias and contrast threshold can be computed in three steps. First, the bias for the pixel $\boldsymbol{p}$ can be estimated from the gradient of the linear regression line through the plot of $\sum \sigma_i^{\boldsymbol{p}}$ versus $\sum |\sigma_i^{\boldsymbol{p}}|$. That is, a linear regression of the cumulative sum of event polarities compared to the number of events at a given pixel coordinates $\boldsymbol{p}$. Let the linear regression line satisfies the equation $\sum \sigma_i^{\boldsymbol{p}} = m^{\boldsymbol{p}}(\sum |\sigma_i^{\boldsymbol{p}}|) + d^{\boldsymbol{p}}$. Second, the contrast threshold can be computed by computing the root mean square error $r^{\boldsymbol{p}}$ of $\sum \sigma_i^{\boldsymbol{p}}$ from the linear regressed line. Finally, the bias $b^{\boldsymbol{p}}$ and contrast threshold $c^{\boldsymbol{p}}$ are

$$
c^{\boldsymbol{p}} = \frac{\mathrm{median}(r^{\boldsymbol{p}})}{r^{\boldsymbol{p}}}, \tag{11}
$$
$$
b^{\boldsymbol{p}} = -c^{\boldsymbol{p}} m^{\boldsymbol{p}}. \tag{12}
$$

5   Experiments

Dataset

We evaluated the proposed calibration methods on the publicly available event-camera dataset of `Dynamic Translation` (indoor environment with moving person, increasing speed, low lighting condition), `Office Spiral` (6-DOF, slow motion, low lighting condition) [18]. In addition, we demonstrated the calibration results on our `Bright Grass` dataset (unstructured outdoor environment, bright lighting condition), which was recorded using another DAVIS240C event camera in a bright natural scene with both fast and slow motion. The dataset is available online[1].

Implementation Details

For offline model *OffEI*, we set $d = 40$ frames in (8) and used the whole sequence of each dataset to estimate $b^{\boldsymbol{p}}$ and $c^{\boldsymbol{p}}$. For online model *OnEI*, experimentally, we found out that when the large buffer contains a maximum of 1,700,000 events and each small buffer stores a maximum of 200,000 events for least square estimation in (8), our online method performs well for different datasets. The calibration algorithms were implemented in MATLAB.

Dynamic Translation	RMSE	PSNR	SSIM
DI	34.01	35.13	0.35
OffE	26.11	39.65	0.47
[5]	34.09	35.11	0.43
OffEI	**22.99**	**42.02**	**0.55**
OnEI	**26.23**	**39.68**	**0.54**
Office Spiral	RMSE	PSNR	SSIM
DI	41.45	31.66	0.58
OffE	35.46	34.32	0.71
[5]	32.32	35.96	0.78
OffEI	**27.40**	**38.77**	**0.85**
OnEI	**29.79**	**37.36**	**0.82**
Bright Grass	RMSE	PSNR	SSIM
DI	81.16	20.02	0.36
OffE	64.80	23.89	0.42
[5]	76.43	20.99	0.36
OffEI	**61.65**	**24.73**	**0.50**
OnEI	**61.78**	**24.70**	**0.50**

Table 2: DI is the direct intergration without calibration. *OffE* and *OffEI* are two offline methods we proposed, event-only and event-image calibration. The Root Mean Square Error (RMSE), Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) [29] are shown. The best resluts in each category are marked in bold.

---

[1]github.com/ziweiWWANG/Event-Camera-Calibration

5.1   Experimental Results

To evaluate our proposed calibration algorithms, we directly integrated events between frame I_{k-n} and I_k as the estimated intensity image $\hat{I}_k = \exp\big(\log I_{k-n} + \Sigma_{i=k-n}^{k} c(\boldsymbol{p}_i)\sigma_i + b(\boldsymbol{p}_i)\big)$. Then, we compared the integration results $\hat{I}_k$ with the referenced intensity image I_k, which is produced by a DAVIS240C event camera. Note that in our event-only method *OffE*, intensity images I_{k-n} are not used to estimate bias and contrast threshold. We also compared the *OnEI* method with the [5] calibration method.
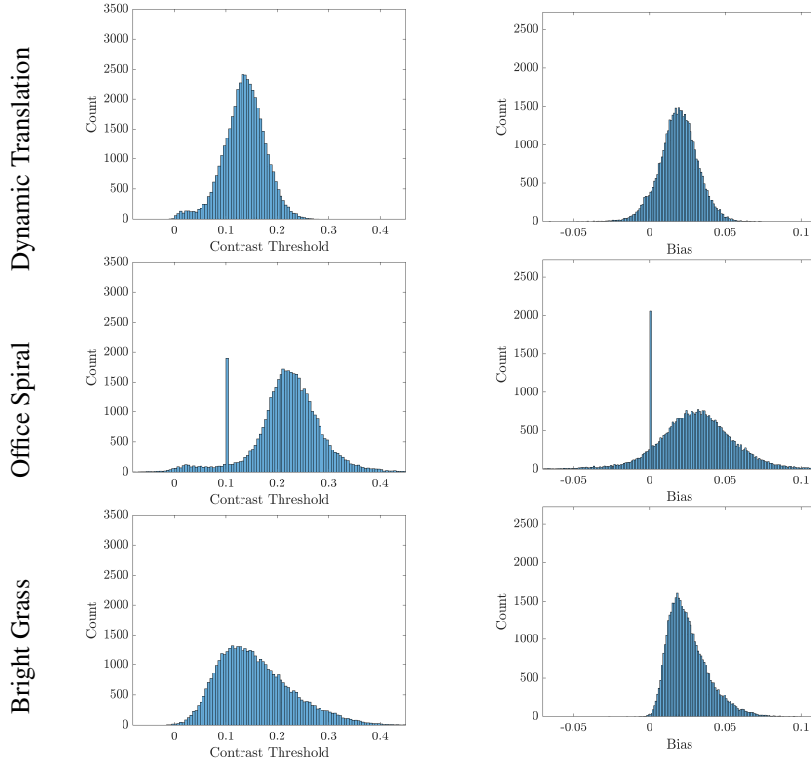

Figure 4: The Histogram of Estimated Bias $b^{\boldsymbol{p}}$ and Contrast Threshold $c^{\boldsymbol{p}}$ Using Different Datasets. First row: `Dynamic Translation`; second row: `Office Spiral`; third row: `Bright Grass`.

Exemplary integration results of different methods on three datasets are shown in Figure 5–7. The three datasets are recorded under different lighting conditions and motions. Because most of the pixels are cool pixel ($b^{\boldsymbol{p}}$ larger than 0), as shown in Figure 4, after direct integration without calibration, the Direct Integrated (DI) images will become darker. The hot pixels (white spots) are also obvious in DI images. Comparing with direct integration using a constant of 0.1 as the contrast threshold, the effect of special pixels is significantly reduced after calibration. For example, in Figure 6, the "shadow" left by the previous location of the keyboard in DI image is mostly removed after calibration. Our intensity image-based method *OffEI* and *OnEI* provide the best calibration results among the four methods presented. For two online methods with intensity images as reference, our online method *OnEI* highly reduces the effect of special pixels and noisy background pixels compared to the method in [5], and has higher similarity with the reference intensity image. A quantitative comparison of four calibration methods is further discussed in Section 5.2, where we demonstrate the superiority of our calibration methods.

For our offline method *OffEI*, we also showed the estimated histogram of bias and contrast threshold in Figure 4. Three different datasets have similar $b^{\boldsymbol{p}}$ and $c^{\boldsymbol{p}}$ distributions. The values of bias $b^{\boldsymbol{p}}$ are relatively small, and for most pixels, they are larger than 0, which are consistent with the effect of positive leakage voltage in amplifier [30]. The estimated contrast threshold values $c^{\boldsymbol{p}}$ are between 0 to 0.4. For pixels that do not have any triggered event and where the intensity changes between intensity frames are close to zero during the whole dataset, matrix A is not full rank and most of the values in vector z are close to zero in (7). In addition, the value of $b^{\boldsymbol{p}}$ and $c^{\boldsymbol{p}}$ will not affect the result of the direct integration in these pixels, we directly assign $b^{\boldsymbol{p}}$ to zero and $c^{\boldsymbol{p}}$ to 0.1. For example, some pixels on the monitor screen in `Office Spiral` remain at zero intensity value during the whole dataset and generate few events (Figure 6). This causes the presence of the two peaks at $c^{\boldsymbol{p}} = 0.1$ and $b^{\boldsymbol{p}} = 0$ in the second row of Figure 4. We also found that the contrast threshold values are not always close to the commonly used value of 0.1 in Figure 4. This is likely due to the different camera sensitivity settings used in the three datasets. In a darker environment, a lower sensitivity is usually chosen to keep the events generated at a desired rate.
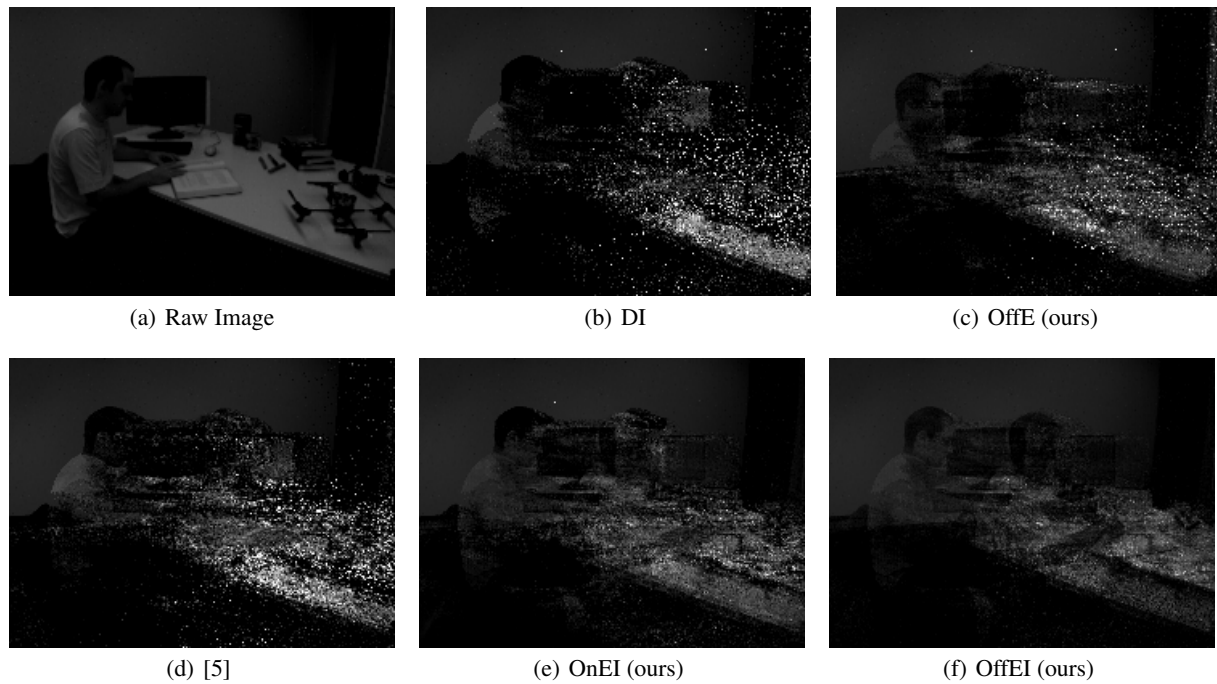
(a) Raw Image

(b) DI

(c) OffE (ours)

(d) [5]

(e) OnEI (ours)

(f) OffEI (ours)

Figure 5: `Dynamic Translation` Dataset. Raw image shows the frame 366 generated by a standard camera sensor. DI is the direct integration without calibration. *OffE* and *OffEI* are two offline methods we proposed, event-only and event-image calibration. We mainly compare our online event-image calibration method OnEI with the [5] method. Four calibration methods integrate events between frame 339-366.
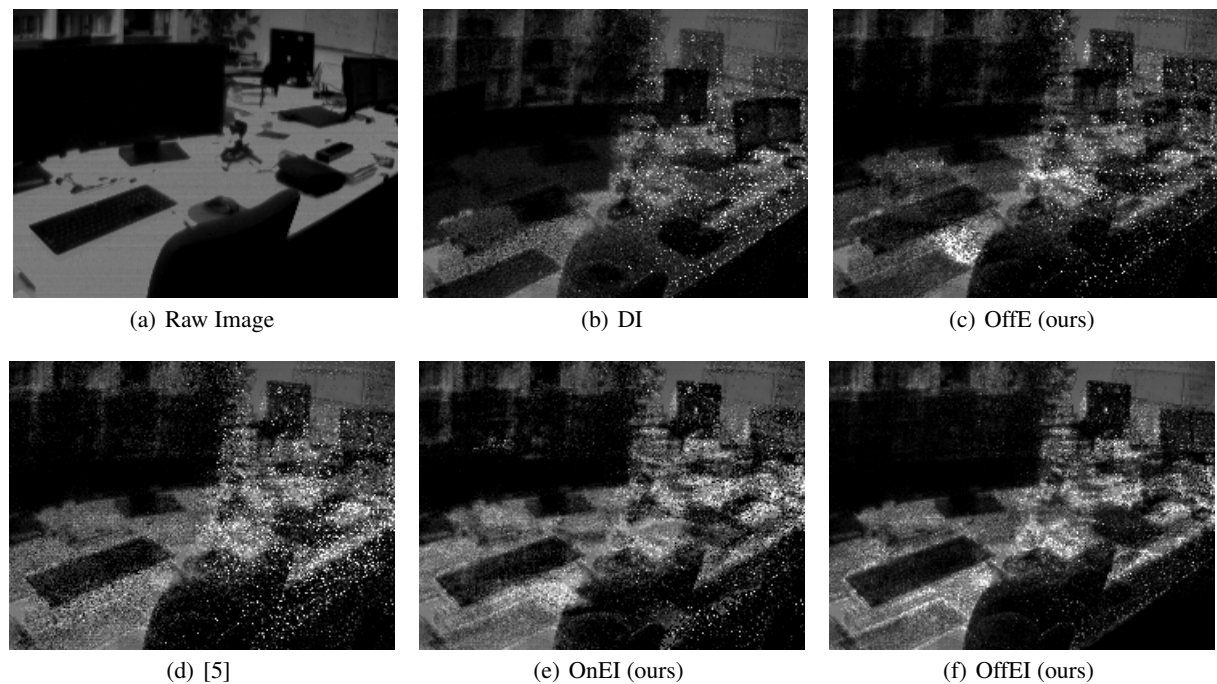

(a) Raw Image

(b) DI

(c) OffE (ours)

(d) [5]

(e) OnEI (ours)

(f) OffEI (ours)

Figure 6: `Office Spiral` Dataset. Raw image shows the frame 101 generated by a standard camera sensor. DI is the direct intergration without calibration. *OffE* and *OffEI* are two offline methods we proposed, event-only and event-image calibration. We mainly compare our online event-image calibration method *OnEI* with the [5] calibration method. Four calibration methods integrate events between frame 83-101.
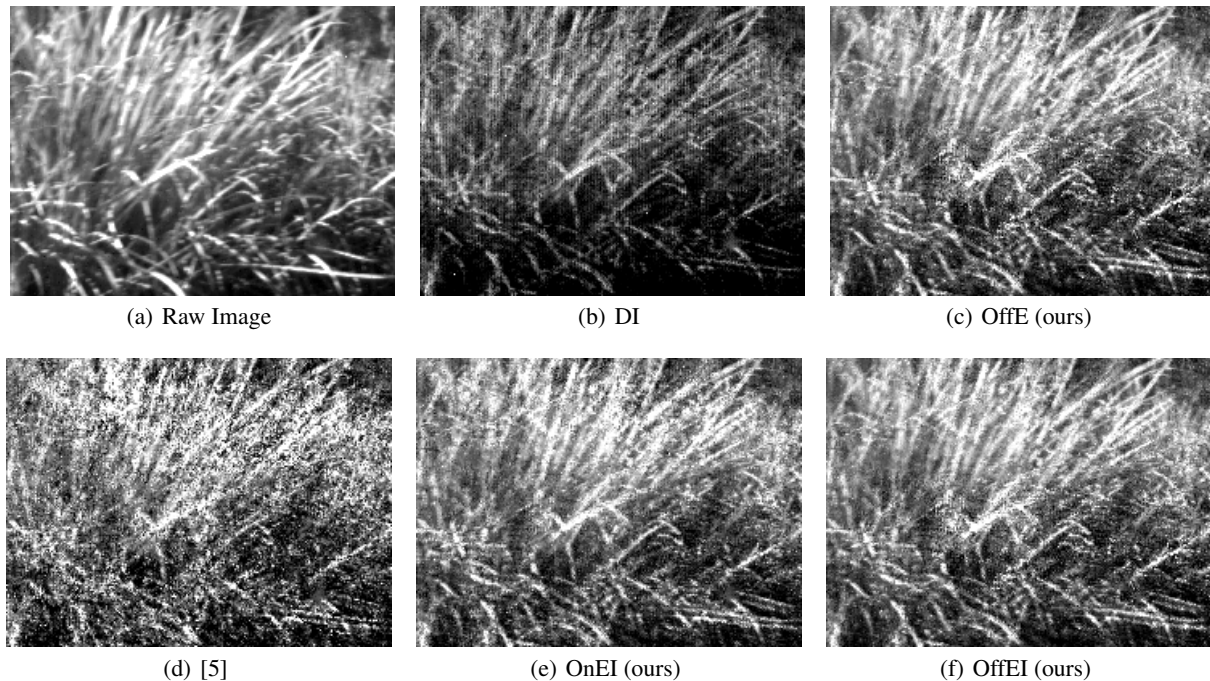
8

(a) Raw Image	(b) DI	(c) OffE (ours)
(d) [5]	(e) OnEI (ours)	(f) OffEI (ours)

Figure 7: `Bright Grass` Dataset. Raw image shows the frame 225 generated by a standard camera sensor. DI is the direct integration without calibration. *OffE* and *OffEI* are two offline methods we proposed, event-only and event-image calibration. We mainly compare our online event-image calibration method *OnEI* with the [5] calibration method. Four calibration methods integrate events between frame 217-225.

5.2 Experimental Metrics

With the corrected parameters b^p and c^p, the direct integrated images should be similar to the corresponding intensity images, even though the intensity images generated by DAVIS cameras could be blurry in fast-motion. The Root Mean Square Error (RMSE), Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) [29] are implemented to evaluate the calibration performance. We compare the referenced intensity images from DAVIS camera with the calibrated direct integrated images. Lower RMSE, higher PSNR and SSIM represent higher similarity to the reference image and a better calibration result.

The average metrics over three different datasets are shown in Table 1. Each output images is direct integrated result for around 500,000 new events from the previous output image. Note that the calibration method in [5] and *OnEI* require time to converge, so the evaluation does not include the output images in the first 5 seconds. The results are consistent with Figure 5–7. Our online method *OnEI* has better calibration performance than the [5] method. In three datasets, the offline model *OffEI* is slightly better than the online model *OnEI*, which is likely due to the bias and contrast threshold being approximately constant for the datasets we used. Another possible reason is that the scenes in the three datasets are relatively unchanged (people and desk, office, grass). Therefore, if the environment and motion speed are known prior to an experiment, it is possible to pre-calibrate the event camera using our offline method. Without intensity images, the performance of our event-only method *OffE* is worse than our image-based methods *OnEI* and *OffEI* (except RMSE in `Dynamic Translation`), which is expected. However, it is still better than the [5] method in `Dynamic Translation` and `Bright Grass`.

6 Conclusion

To the best of our knowledge, an event generation model supported by transistor junction leakage bias and contrast threshold variance has not been previously proposed. We fully take advantage of the low-frequency intensity images provided by DAVIS event cameras, providing both online and offline calibration methods. We also present an offline regression method to calibrate pure event cameras. We demonstrate that our calibration methods greatly eliminate the effects of special pixels and better reflect the actual intensity changes, which will improve the performance of many applications that use event cameras.

9

References

[1] Ignacio Alzugaray and Margarita Chli. Asynchronous corner detection and tracking for event cameras in real time. *IEEE Robotics and Automation Letters*, 3(4):3177–3184, 2018.

[2] Patrick Bardow, Andrew J Davison, and Stefan Leutenegger. Simultaneous optical flow and intensity estimation from an event camera. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 884–892, 2016.

[3] Raphael Berner, Christian Brandli, Minhao Yang, S-C Liu, and Tobi Delbruck. A 240x180 120db 10mw 12us-latency sparse output vision sensor for mobile applications. In *Proceedings of the International Image Sensors Workshop*, number CONF, pages 41–44, 2013.

[4] Raphael Berner and Tobi Delbruck. Event-based pixel sensitive to changes of color and brightness. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 58(7):1581–1590, 2011.

[5] Christian Brandli, Lorenz Muller, and Tobi Delbruck. Real-time, high-speed video decompression using a frame-and event-based davis sensor. In *2014 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 686–689. IEEE, 2014.

[6] Gottfried Graber Christian Reinbacher and Thomas Pock. Real-time intensity-image reconstruction for event cameras using manifold regularisation. In Edwin R. Hancock Richard C. Wilson and William A. P. Smith, editors, *Proceedings of the British Machine Vision Conference (BMVC)*, pages 9.1–9.12. BMVA Press, September 2016.

[7] Matthew Cook, Luca Gugelmann, Florian Jug, Christoph Krautz, and Angelika Steger. Interacting maps for fast visual interpretation. In *The 2011 International Joint Conference on Neural Networks*, pages 770–776. IEEE, 2011.

[8] Tobi Delbruck. Frame-free dynamic digital vision. In *Proceedings of Intl. Symp. on Secure-Life Electronics, Advanced Electronics for Quality Life and Society*, pages 21–26, 2008.

[9] Daniel Gehrig, Henri Rebecq, Guillermo Gallego, and Davide Scaramuzza. Asynchronous, photometric feature tracking using events and frames. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 750–765, 2018.

[10] Daniel Gehrig, Henri Rebecq, Guillermo Gallego, and Davide Scaramuzza. Eklt: Asynchronous photometric feature tracking using events and frames. *International Journal of Computer Vision*, pages 1–18, 2019.

[11] Hanme Kim, Ankur Handa, Ryad Benosman, Sio-Hoi Ieng, and Andrew J Davison. Simultaneous mosaicing and tracking with an event camera. *J. Solid State Circ*, 43:566–576, 2008.

[12] Hanme Kim, Stefan Leutenegger, and Andrew J Davison. Real-time 3d reconstruction and 6-dof tracking with an event camera. In *European Conference on Computer Vision*, pages 349–364. Springer, 2016.

[13] Beat Kueng, Elias Mueggler, Guillermo Gallego, and Davide Scaramuzza. Low-latency visual odometry using event-based feature tracks. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 16–23. IEEE, 2016.

[14] Juan Antonio Leñero-Bardallo, Teresa Serrano-Gotarredona, and Bernabé Linares-Barranco. A 3.6μ s latency asynchronous frame-free event-driven dynamic-vision-sensor. *IEEE Journal of Solid-State Circuits*, 46(6):1443–1455, 2011.

[15] Patrick Lichtsteiner, Christoph Posch, and Tobi Delbruck. A 128×128 120 db 15μ s latency asynchronous temporal contrast vision sensor. *IEEE journal of solid-state circuits*, 43(2):566–576, 2008.

[16] Jacques Manderscheid, Amos Sironi, Nicolas Bourdis, Davide Migliore, and Vincent Lepetit. Speed invariant time surface for learning to detect corner points with event-based cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 10245–10254, 2019.

[17] E. Mueggler, B. Huber, and D. Scaramuzza. Event-based, 6-dof pose tracking for high-speed maneuvers. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2761–2768, Sep. 2014.

[18] Elias Mueggler, Henri Rebecq, Guillermo Gallego, Tobi Delbruck, and Davide Scaramuzza. The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and slam. *The International Journal of Robotics Research*, 36(2):142–149, 2017.

[19] Garrick Orchard, Daniel Matolin, Xavier Lagorce, Ryad Benosman, and Christoph Posch. Accelerated frame-free time-encoded multi-step imaging. In *2014 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 2644–2647. IEEE, 2014.

[20] Liyuan Pan, Cedric Scheerlinck, Xin Yu, Richard Hartley, Miaomiao Liu, and Yuchao Dai. Bringing a blurry frame alive at high frame-rate with an event camera. In *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6820–6829, 2019.

[21] Christoph Posch, Daniel Matolin, and Rainer 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*, 46(1):259–275, 2010.

[22] Christian Reinbacher, Gottfried Graber, and Thomas Pock. Real-time intensity-image reconstruction for event cameras using manifold regularisation. *arXiv preprint arXiv:1607.06283*, 2016.

[23] Cedric Scheerlinck, Nick Barnes, and Robert Mahony. Continuous-time intensity estimation using event cameras. In *Asian Conference on Computer Vision*, pages 308–324. Springer, 2018.

[24] Cedric Scheerlinck, Nick Barnes, and Robert Mahony. Asynchronous spatial image convolutions for event cameras. *IEEE Robotics and Automation Letters*, 4(2):816–822, 2019.

[25] 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. *IEEE Journal of Solid-State Circuits*, 48(3):827–838, 2013.

[26] Bongki Son, Yunjae Suh, Sungho Kim, Heejae Jung, Jun-Seok Kim, Changwoo Shin, Keunju Park, Kyoobin Lee, Jinman Park, Jooyeon Woo, et al. 4.1 a 640× 480 dynamic vision sensor with a 9μm pixel and 300meps address-event representation. In *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, pages 66–67. IEEE, 2017.

[27] David Tedaldi, Guillermo Gallego, Elias Mueggler, and Davide Scaramuzza. Feature detection and tracking with the dynamic and active-pixel vision sensor (davis). In *2016 Second International Conference on Event-based Control, Communication, and Signal Processing (EBCCSP)*, pages 1–7. IEEE, 2016.

[28] A. R. Vidal, H. Rebecq, T. Horstschaefer, and D. Scaramuzza. Ultimate slam? combining events, images, and imu for robust visual slam in hdr and high-speed scenarios. *IEEE Robotics and Automation Letters*, 3(2):994–1001, April 2018.

[29] Zhou Wang, Alan C Bovik, Hamid R Sheikh, Eero P Simoncelli, et al. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[30] Minhao Yang, Shih-Chii Liu, and Tobi Delbruck. A dynamic vision sensor with 1% temporal contrast sensitivity and in-pixel asynchronous delta modulator for event encoding. *IEEE Journal of Solid-State Circuits*, 50(9):2149–2160, 2015.

[31] Yi Zhou, Guillermo Gallego, Henri Rebecq, Laurent Kneip, Hongdong Li, and Davide Scaramuzza. Semi-dense 3d reconstruction with a stereo event camera. In *European Conference on Computer Vision*, pages 242–258. Springer, 2018.