

Simultaneous Motion And Noise Estimation with Event Cameras

Shintaro Shiba^{1,2,3}, Yoshimitsu Aoki¹ and Guillermo Gallego^{3,4}

¹ Keio University, ² Woven by Toyota, Inc., Japan. ³ Technische Universität Berlin, ⁴ Einstein Center Digital Future, Robotics Institute Germany, and Science of Intelligence Excellence Cluster, Germany.

Abstract

Event cameras are emerging vision sensors whose noise is challenging to characterize. Existing denoising methods for event cameras are often designed in isolation and thus consider other tasks, such as motion estimation, separately (i.e., sequentially after denoising). However, motion is an intrinsic part of event data, since scene edges cannot be sensed without motion. We propose, to the best of our knowledge, the first method that simultaneously estimates motion in its various forms (e.g., ego-motion, optical flow) and noise. The method is flexible, as it allows replacing the one-step motion estimation of the widely-used Contrast Maximization framework with any other motion estimator, such as deep neural networks. The experiments show that the proposed method achieves state-of-the-art results on the E-MLB denoising benchmark and competitive results on the DND21 benchmark, while demonstrating effectiveness across motion estimation and intensity reconstruction tasks. Our approach advances event-data denoising theory and expands practical denoising use-cases via open-source code. Project page: <https://github.com/tub-rip/ESMD>

1. Introduction

Event cameras are emerging vision sensors that overcome shortcomings of conventional cameras (e.g., motion blur, limited dynamic range, data and power efficiency), yet they suffer from a considerable amount of noise due to their novelty and operation in low-power (transistor subthreshold) conditions [19, 36]. Since event cameras are suitable for many computer vision tasks, especially motion-related tasks, it is paramount to classify event data that is related to motion (i.e., *signal*) and that is not (i.e., *noise*). However, denoising event data is a challenging problem because noise properties have yet to be fully characterized [22, 24] and it is infeasible to define ground-truth (GT) noise labels in real-world data recordings. Previous works either leveraged simulation or prepared real event data and aggressively suppressed noise occurrence.

However, such artificial modifications on the real event

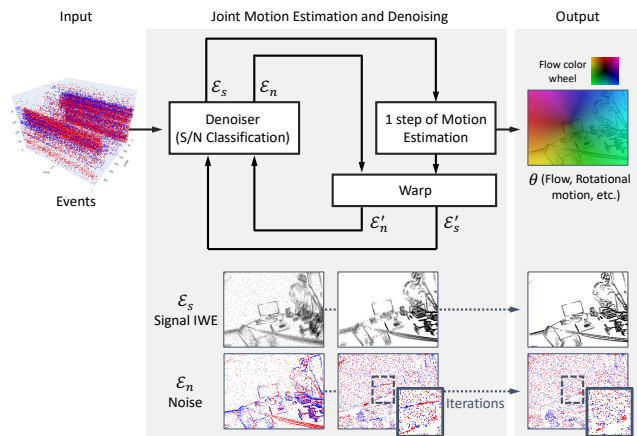


Figure 1. Overview of the proposed method. We use only raw events as input, and estimate signal events, noise events, and motion (e.g., dense optical flow and ego-motion estimation).

data could change event signal/noise characteristics, possibly deteriorating the information that the original event had for downstream tasks (e.g., motion estimation). Our work is inspired by the observation that motion is an intrinsic part of event data since scene edges cannot be sensed without motion. Rather than solving denoising and motion estimation tasks separately, our idea is to leverage motion information to improve denoising (as “vice-versa” is well known). In this work (Fig. 1), we rethink and investigate the connection between denoising and motion estimation, by asking ourselves whether they can be combined in a more integrated way to benefit the results of both tasks simultaneously. We formulate the simultaneous estimation problem by leveraging the idea of motion compensation in the Contrast Maximization (CMax) framework, defining novel metrics about the contribution to the contrast score for each event. In summary, we present several distinctive contributions:

1. We propose a novel method to jointly estimate noise and motion only from event data (Sec. 3).
2. We demonstrate that our joint estimation method (i) achieves state-of-the-art results in existing denoising benchmarks (Sec. 4.2), and (ii) improves the robustness of motion estimation (CMax framework) (Sec. 4.3).

3. We show the flexibility of our method on (i) further applications, such as intensity reconstruction, validating its image quality (Secs. 4.4 and 4.5), and (ii) in combination with learning-based motion estimators (Sec. 4.6).

We hope our work opens a new line of research for event cameras, by considering the problems of motion estimation and denoising together to take advantage of their coaction.

2. Related Work

Denoising is a fundamental problem in event cameras since there are a lot of leakage events [24], and event-processing algorithms may suffer in scenarios with high noise [3]. Removing event noise has a double effect: it reduces the load of subsequent processing stages (as the load is proportional to the number of events) and improves their output (as the input data becomes of higher quality). In this work, we focus on noise due to the leakage inside the camera hardware, such as background activity (BA) events [36]. Events due to flickering or active lights are therefore not our priority.

Classical denoising works focus on the spatio-temporal correlation of events, as BA events have fewer neighboring events in space-time coordinates than signal events. This approach includes widely-used BA Filter (BAF) [9], and other types of spatio-temporal filters [26, 31, 37, 56]. Some works utilize additional sensor inputs (e.g., frames) to estimate noise probabilities [4, 12, 30, 55, 58], since edges are informative of the signal events produced.

Recently, learning-based approaches have gained more attention, such as multilayer perceptrons [24], convolutional neural networks [1, 4, 12], graph neural networks [2], and transformers [29]. They provide the probability of noise (i.e., signal-noise classification). Nonetheless, the challenge of learning-based approaches is the need for ground truth (GT) noise labels for training. GT labels have been obtained through simulation (e.g., [13, 24]) as well as real-world recordings, however, real-world recordings aggressively remove noisy events using existing filters (e.g., [24]) or use unnatural illumination settings (e.g., [29]), by leveraging the differences of noise occurrence due to different scene brightness [24, 35]. While these approaches are effective for benchmarking, the challenge of obtaining GT labels for practical real-world datasets remains unsolved.

Our work builds upon a series of successful methods for motion estimation and their applications. Event-based motion estimation is an important topic since event cameras naturally respond to motion in the scenes [19]. In particular, Contrast Maximization (CMax) [17, 50] is a state-of-the-art framework for motion estimation, which has various applications, such as rotational motion estimation [15, 23, 32, 42, 45], optical flow estimation [27, 28, 44, 48, 64], and motion segmentation [51, 57, 60]. We investigate the connection and potential benefits of applying the CMax framework to denoising during the estimation process (i.e., simul-

taneously), demonstrating motion estimation improvements as well as competitive denoising accuracy.

A prior work, “ProgressiveMotionSeg” [8], has extended the CMax-based segmentation method [51] with extra weights that accounted for how much each event contributed to the estimated motion clusters. However, [8] mainly focuses on the task of motion segmentation using low degrees-of-freedom (DOF) warps (e.g., 2-DOF feature flow) and does not discuss denoising efficacy or evaluate on standard denoising benchmarks. In contrast, our method supports various types of motion estimation (both low-DOF and high-DOF warps, such as optical flow), and considers that noise shall not be used to estimate motion, as from first principles we deem noise and motion to be uncorrelated.

3. Methodology

Event cameras consist of independent pixels that generate asynchronous “events” when the logarithmic brightness at the pixel increases or decreases by a predefined contrast sensitivity. Each event $e_k \doteq (\mathbf{x}_k, t_k, p_k)$ contains the pixel coordinates $\mathbf{x}_k$, the timestamp t_k , and polarity $p_k \in \{+1, -1\}$ of the brightness change. Events occur asynchronously and sparsely on the pixel grid, resulting in a variable data rate based on the scene texture and dynamics.

3.1. Contrast Maximization

The CMax framework [17] assumes events $\mathcal{E} \doteq \{e_k\}_{k=1}^{N_e}$ are caused by moving edges, and it transforms their coordinates according to a motion model $\mathbf{W}$, producing a set of warped events $\mathcal{E}'_{t_{\text{ref}}} \doteq \{e'_k\}_{k=1}^{N_e}$ at a reference time t_{ref} :

$$e_k \doteq (\mathbf{x}_k, t_k, p_k) \mapsto e'_k \doteq (\mathbf{x}'_k, t_{\text{ref}}, p_k). \quad (1)$$

The warp $\mathbf{x}'_k = \mathbf{W}(\mathbf{x}_k, t_k; \boldsymbol{\theta})$ transports each event from t_k to t_{ref} along the motion curve that passes through it. Then, they are aggregated on an image of warped events (IWE):

$$I(\mathbf{x}; \mathcal{E}'_{t_{\text{ref}}}, \boldsymbol{\theta}) \doteq \sum_{k=1}^{N_e} \delta(\mathbf{x} - \mathbf{x}'_k), \quad (2)$$

where each pixel $\mathbf{x}$ sums the number of warped events $\mathbf{x}'_k$ that fall within it. The Dirac delta is approximated by a Gaussian, $\delta(\mathbf{x} - \boldsymbol{\mu}) \approx \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \epsilon^2 \text{Id})$ with $\epsilon = 1$ pixel.

Next, an objective function, such as the contrast of the IWE (2) (image variance [18] $\text{Var}(I(\mathbf{x}; \boldsymbol{\theta})) \doteq \frac{1}{|\Omega|} \int_{\Omega} (I(\mathbf{x}; \boldsymbol{\theta}) - \mu_I)^2 d\mathbf{x}$), measures the goodness of fit between the events and the candidate motion curves (warp).

Finally, an optimization algorithm iterates the above steps to find the motion parameters $\boldsymbol{\theta}^*$ that maximize the alignment of events caused by the same scene edge. Event alignment is measured by the strength of the edges of the IWE, which is related to image contrast [18].

Motion Models. As exemplary warps $\mathbf{W}$ we focus on two popular problem settings: rotational motion (3-DOF angular velocity estimation [23, 25, 42, 45]) and dense optical flow (pixel-wise velocity estimation [43, 48, 50, 64],

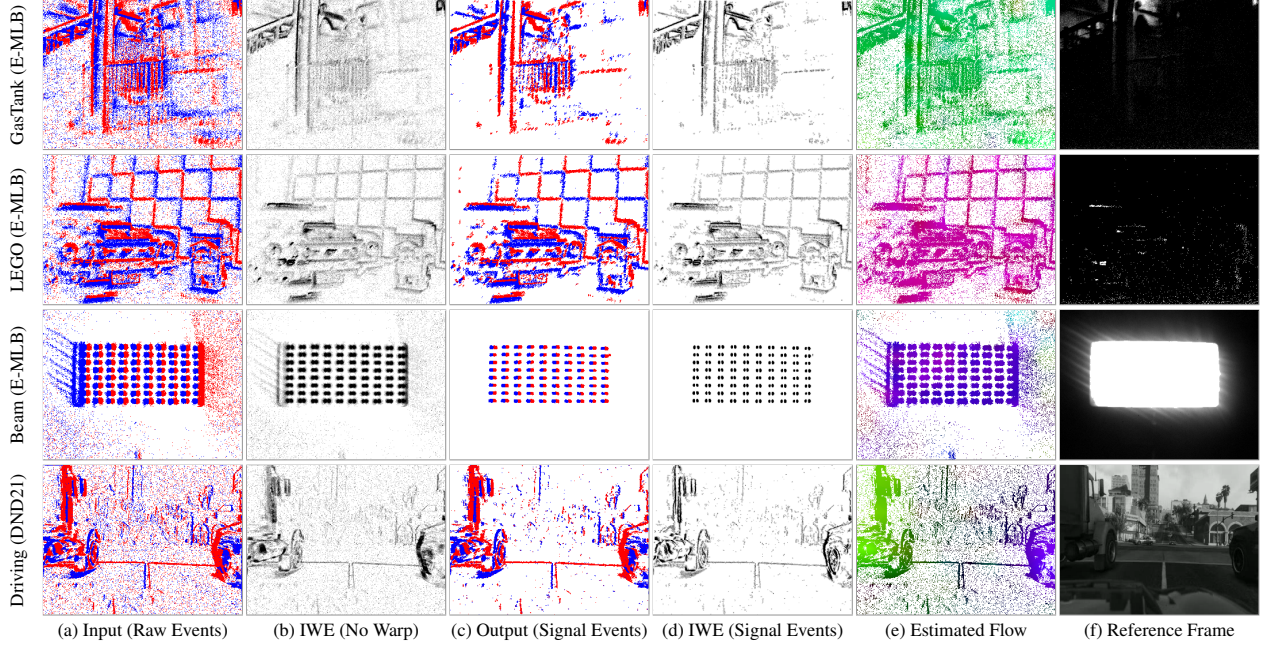


Figure 2. Qualitative result of denoising and flow estimation on E-MLB and DND21. The driving sequence has 5-Hz artificial noise. (b) is an alternative visualization of (a) with the same grayscale coding as (d). Flow is visualized on all events and flow color is given in Fig. 1.

resulting in $2N_p$ -DOFs, where N_p is the number of pixels). For the angular velocity estimation during a small time interval, the warp is parametrized by the angular velocity $\boldsymbol{\theta} \equiv \boldsymbol{\omega} = (\omega_x, \omega_y, \omega_z)^\top$, and $\mathbf{x}^h(t) \sim \mathbf{R}(\boldsymbol{\omega}t) \mathbf{x}^h(0)$, using calibrated homogeneous coordinates $\mathbf{x}^h$ and exponential coordinates $\mathbf{R}(\phi) \doteq \exp(\phi^\wedge)$ [6, 16]. The warp for dense optical flow is $\mathbf{x}'_k = \mathbf{x}_k + (t_k - t_{\text{ref}}) \mathbf{v}(\mathbf{x}_k)$, where $\boldsymbol{\theta} = \{\mathbf{v}(\mathbf{x})\}_{\mathbf{x} \in \Omega}$ is a flow field on the image plane Ω .

3.2. Denoising Based on Local Contrast

As Fig. 1 shows, our method is iterative. In each iteration, motion estimation is refined and events are classified into signal $\mathcal{E}_s$ or noise $\mathcal{E}_n$. Without labels or additional information, events rely on each other to realize this classification. To this end, we compute a score c_k for each event, rank the events according to their scores, and threshold at some value T . The threshold is computed by assuming the amount or ratio of noisy events is known η , i.e., $T \equiv T(\eta)$, which allows us to have high-level control over the denoising process. Hence we classify events into signal (the $\tau \doteq (1 - \eta)$ percentage of events with highest scores) and noise:

$$\begin{aligned} \mathcal{E}_s &\equiv \mathcal{E}_{\text{signal}} \doteq \{e_k \in \mathcal{E} \mid c_k > T(\eta)\}, \\ \mathcal{E}_n &\equiv \mathcal{E}_{\text{noise}} \doteq \mathcal{E} \setminus \mathcal{E}_{\text{signal}} \quad (\text{complement of } \mathcal{E}_{\text{signal}}). \end{aligned} \quad (3)$$

We studied different choices for the per-event score c_k based on the agreement among warped events $\mathcal{E}' = \mathcal{E}'_s \cup \mathcal{E}'_n$, through their IWEs, and concluded on two, as follows. Letting $I_k^s \equiv I_k^s(\mathbf{x}'_k)$, $I_k^n \equiv I_k^n(\mathbf{x}'_k)$ and $I_k \equiv I_k(\mathbf{x}'_k) = I_k^s + I_k^n$

be the IWE values of previously classified signal events, noise events and all events, respectively, the two choices are: (i) ${}^1c_k \doteq I_k$, and (ii) the ratio of signal events ${}^2c_k \doteq I_k^s/I_k$ (probabilistic viewpoint). In fact, both are equivalent: rule ${}^1c_k > T_1$ is the same as rule ${}^2c_k > T_2$ if $T_2 = 1 - I_k^n/T_1$ and the event noise I_k^n is known. We use ${}^1c_k$, the number of warped events (also called local contrast [18, 51]), because it is the simplest to interpret.

The intuition is that “signal” events (e.g., due to motion) warped according to the true motion lie all at motion-compensated edges, and therefore produce a sharp IWE. Noise events, produced all over the image plane, are not expected to contribute to the IWE’s sharpness. Hence, we may classify events into signal or noise according to how much they contribute to edge strength. The IWE (2) provides one such measure of edge strength: the higher $I(\mathbf{x})$, the more events are warped to that pixel location $\mathbf{x}$ (i.e., more events support the same scene edge), thus producing a sharper IWE. Hence, we define how much an event e_k contributes to edge strength by means of the IWE (2), ${}^1c_k$.

Invariance. Note that (3) is invariant to monotonically increasing functions of c_k , as such functions preserve the order. The same $\mathcal{E}_{\text{signal}}$ is obtained if one ranks and thresholds (at the corresponding level) values $\{c_k\}$, $\{\ln(c_k)\}$, $\{e^{c_k}\}$, etc. They are equivalent under the ranking rule (3).

Robustness to Various Edge Strengths. The Gaussian kernel in the IWE controls the sensitivity to edge strength. Since the proposed method is based on both signal and noise

IWEs, increasing the size of the Gaussian kernel in the IWE can emphasize contour edges over isolated points. This preserves signal events better in regions with low IWE intensities, for example at pixels with large depth values.

3.3. Alternating Optimization

Classifying events into signal and noise requires knowledge of the true motion, and estimation of the true motion (e.g., using CMax) requires knowledge of the signal events because, by definition, noise events carry no information about motion. Hence, it is a circular dependency problem.

We approach its solution iteratively, as shown in Fig. 1, by combining motion estimation (Sec. 3.1) and S/N classification (Sec. 3.2). Assuming some initialization, the current signal events $\mathcal{E}_{\text{signal}}$ are used to estimate the motion via CMax (1 step is enough); then all events are warped to compute the IWE and the per-event scores c_k , from which the sets (3) are recomputed, in preparation for the next iteration. In practice, we initialize with a random split of $\mathcal{E}$ into $\mathcal{E}_{\text{signal}}$ and $\mathcal{E}_{\text{noise}}$, update $\mathcal{E}_{\text{signal}}$ at every warp during the optimization iteration until the process converges (as flagged by the convergence of the motion parameters in CMax).

Computational Complexity. The computational complexity of one iteration of the proposed method is $O(N_p + N_e \log N_e)$, which is slightly larger than the original CMax (i.e., $O(N_p + N_e)$) due to the search for the highest-ranked events during the optimization.

4. Experiments

We evaluate the proposed method on four datasets using various metrics (in Sec. 4.1). In the following sections using common benchmarks, we assess denoising (Sec. 4.2), joint motion estimation (Sec. 4.3), and the efficacy on intensity reconstruction (Sec. 4.4). We also conduct sensitivity analyses (Sec. 4.5) and an ablation (Sec. 4.6).

4.1. Datasets, Metrics, and Hyper-parameters

Datasets. E-MLB [10] is a large-scale, de-facto dataset to benchmark denoising. It consists of 100 sequences with 4 different brightness levels using neutral-density (ND) filters on a DAVIS346 camera (346×260 px) during day and night.

Historically, the two sequences (*hotel* and *driving*) of the DND21 dataset [24] have been widely used for denoising evaluation [10, 24, 29]. Both sequences are recorded with a DAVIS346 camera [53]. They are regarded as “signal” data, as the sequences have been aggressively filtered offline. Combined with other pure BA noise sequences, the dataset has event-wise noise annotations that are useful for evaluation. We use the above sequences with different noise rates, 1, 3, 5, 7, 10 Hz per pixel, following prior work [29].

The ECD dataset [40] is a standard dataset for various tasks including camera ego-motion estimation [15, 41, 47, 49, 61, 62]. Using a DAVIS240C camera (240×180 px [7]),

		E-MLB (Day)				E-MLB (Night)				DND21
		ND1	ND4	ND16	ND64	ND1	ND4	ND16	ND64	
Model-based	Raw	0.821	0.824	0.815	0.786	0.890	0.824	0.786	0.768	0.869
	BAF [9]	0.861	0.869	0.876	0.890	0.946	<u>0.973</u>	0.992	0.942	0.920
	TS [34]	0.877	0.887	0.870	0.837	1.033	0.944	0.886	0.797	0.985
	KNoise [31]	0.846	0.837	0.830	0.807	0.954	0.956	0.871	0.817	0.887
	EvFlow [54]	0.848	0.878	0.868	0.833	0.969	0.983	<u>0.889</u>	0.797	1.006
	IETS [5]	0.772	0.785	0.777	0.753	0.950	0.823	0.804	0.711	0.900
	Ynoise [14]	0.866	0.863	0.857	0.821	1.009	0.943	0.875	0.792	0.966
	GEF [11]	1.051	<u>0.938</u>	<u>0.935</u>	<u>0.927</u>	<u>1.027</u>	0.955	0.946	<u>0.935</u>	0.932
	DWF [24]	0.878	0.876	0.866	0.865	0.923	0.962	0.988	0.932	0.905
	Ours	<u>0.938</u>	0.958	0.986	0.950	1.037	0.961	0.945	0.932	<u>0.992</u>
Learning	EDnCNN [4]	0.887	0.908	0.903	0.912	1.001	1.024	1.079	<u>1.086</u>	0.977
	EventZoom [12]	0.996	0.988	0.996	0.970	1.055	1.007	1.010	0.988	1.059
	MLPF [24]	0.851	0.855	0.846	0.840	0.926	0.928	0.910	0.906	0.944
	EDformer [29]	<u>0.952</u>	<u>0.955</u>	<u>0.956</u>	<u>0.942</u>	<u>1.048</u>	<u>1.019</u>	<u>1.076</u>	1.099	<u>1.041</u>

Table 1. Mean ESR (MESR $\uparrow$) results among denoising methods on the event denoising datasets E-MLB [10] and DND21 [24]. In each category, the best is in bold and the second best is underlined. Methods are chronologically sorted within each category.

		1Hz		5Hz		10Hz	
		<i>hotel</i>	<i>driving</i>	<i>hotel</i>	<i>driving</i>	<i>hotel</i>	<i>driving</i>
Model-based	BAF [9]	0.954	0.848	0.892	0.793	0.837	0.748
	TS [34]	<u>0.972</u>	<u>0.931</u>	<u>0.961</u>	0.927	0.962	0.920
	KNoise [31]	0.677	0.630	0.670	0.624	0.641	0.614
	Ynoise [14]	0.969	0.941	0.923	<u>0.909</u>	0.899	<u>0.880</u>
	DWF [24]	0.927	0.741	0.862	0.690	0.796	0.656
	Ours	1.014	0.882	0.963	0.855	<u>0.961</u>	0.836
Learning	EDnCNN [4]	0.957	0.887	0.937	0.875	0.901	0.874
	MLPF [24]	<u>0.970</u>	0.889	<u>0.970</u>	<u>0.885</u>	<u>0.963</u>	<u>0.876</u>
	EDformer [29]	0.993	0.954	0.985	0.942	0.970	0.926

Table 2. The AUC $\uparrow$ of ROC on the two DND21 sequence at different noise rates. We follow [29] for baseline methods selection.

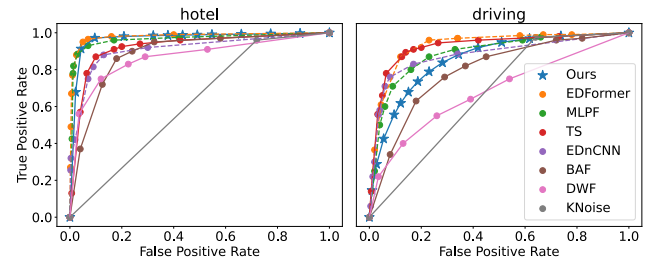


Figure 3. ROC curves on the DND 21 dataset sequences. Dash lines represent learning-based methods (see also Tab. 2).

each sequence provides events, frames, calibration information, IMU data, and ground truth (GT) camera poses (at 200 Hz). We use *boxes_rotation* and *dynamic_rotation* sequences to evaluate rotational motion estimation accuracy.

Finally, we also use sequences from MVSEC [63] and *zurich_city_12a* sequence from the DSEC dataset [20], which are typically used for optical flow estimation. We use MVSEC sequences for quantitative evaluation and the DSEC sequence for qualitative evaluation, since it is captured during the night and hence includes lots of noise.

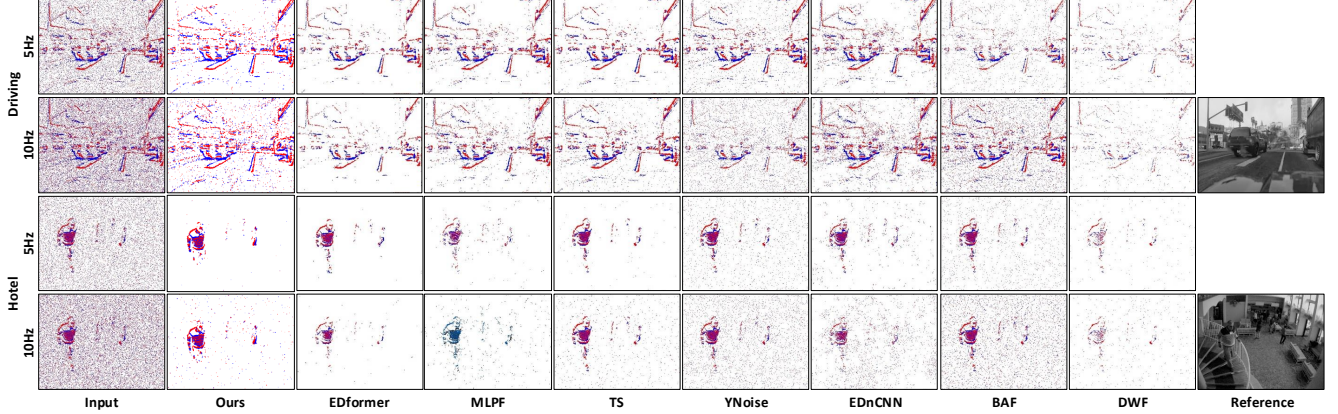


Figure 4. Comparison of denoising methods on the two DND21 sequences at two different noise levels.

Metrics. The metrics used for denoising assessment are: receiver operating characteristic (ROC) curve, area under the curve (AUC) and Mean Event Structural Ratio (MESR). Given signal and noise GT labels, the ROC shows the True-Positive Rate and False-Positive Rate over several S/N ratios. The AUC of the ROC becomes closer to 1, the better the denoiser works. The MESR is a denoising metric that does not require GT labels [10]. We follow previous work to fix the number of signal events to 30000 [10, 29], as well as $M = 20000$ [10, Eq. 10], to discount the difference in the number of signal events from various denoising methods. The higher, the better. Furthermore, we use the RMS error for angular velocity estimation and the EPE for optical flow estimation, following previous works [23, 25, 42, 50], as well as the FWL metric [52], which is the relative variance (sharpness) of the IWE with respect to that of the identity warp. Further details about the experimental settings (e.g., hyper-parameters) are in the supplementary.

4.2. Denoising Results

We evaluate denoising accuracy using AUC on DND21 data, and using MESR on E-MLB and DND21 data, using the optical flow estimation.

Denoising Performance on E-MLB. Table 1 shows the benchmark results on the E-MLB dataset using the MESR metric. We categorize prior work into model-based methods and learning-based methods for convenience. Compared with other model-based techniques, our method ranks first or second, showcasing the efficacy of the proposed method on denoising. Even compared with learning-based ones, our method achieves competitive scores (e.g., second in E-MLB (Day) ND4–ND64 conditions, etc). Notice that, however, learning-based approaches utilize additional information (GT labels) during training, while our method does not. Nonetheless, the results demonstrate the efficacy of our approach, which contributes to the theoretical foundation of event-denoising methodologies. We discuss further details

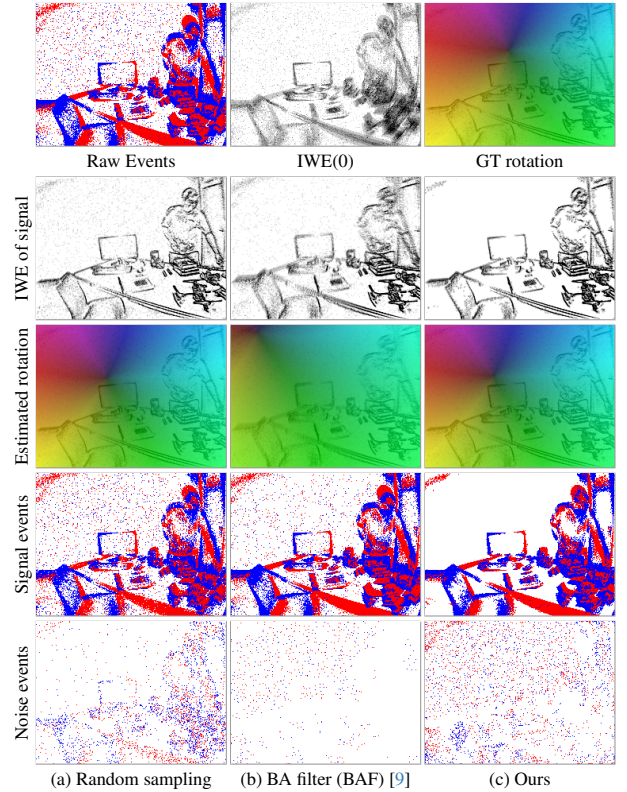


Figure 5. Simultaneous estimation of rotational motion and denoising on the *dynamic_rotation* sequence from ECD dataset [40]. Our method produces sharp IWEs (2nd row) as well as reasonable S/N classification (rows 4 and 5), keeping 90% of data ($\tau = 0.9$).

of the evaluation metrics (ESR) in Sec. 5.

Qualitative results are shown in Fig. 2. The denoised outputs (“Signal”) are reasonable, as dark (underexposed) regions of the scene become cleaner, and the IWEs using the estimated flow provide sharp edges of the original scenes. We compare with state-of-the-art methods in detail next.

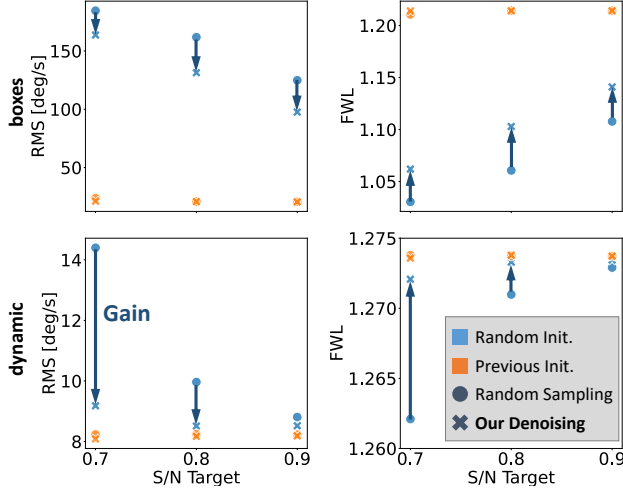


Figure 6. Results on the ego-motion estimation application (top: *boxes*, bottom: *dynamic*) [40]. Our method relaxes the dependency on the initial value for CMax-based rotational motion estimation (indicated as “Gain”).

	<i>dynamic_rot</i>		<i>boxes_rot</i>	
	RMS ↓	FWL ↑	RMS ↓	FWL ↑
CMax [17]	20.001	1.259	124.641	1.129
– w/ Init.	8.275	1.274	20.659	1.214
Downsampling (Best)	8.808	1.273	124.869	1.108
– w/ Init.	<u>8.226</u>	<u>1.274</u>	20.619	<u>1.214</u>
Ours (Best)	8.522	1.273	97.585	1.141
– w/ Init.	8.170	1.274	<u>20.604</u>	1.214
BAF	19.675	1.260	125.028	1.127
– w/ Init.	8.253	1.274	20.550	1.214

Table 3. Angular velocity estimation on ECD dataset [40].

Denoising Accuracy on DND21. Table 2 shows the result of AUC for the ROC result on the DND21 dataset. Our method achieves consistently high AUC compared with existing model-based methods. Note that the DND21 dataset aggressively removes events to generate “pure-signal” sequences [24, Sec.5]. Due to this step, the labeled signal events may be fewer than the actual ones corresponding to the true motion. Hence, the DND21 sequences could result in a slight degradation in the False Positive score for motion-based (unsupervised) denoising methods like ours, which may result in lower AUC values. Nonetheless, our method provides competitive AUC scores among other state-of-the-art. The ROC curves are shown in Fig. 3.

Qualitative comparisons among several denoising methods is given in Fig. 4. From the results, our method achieves a good trade-off between under-denoising (preserving not only scene-edge details but some noise) and over-denoising (removing not only most of the noise but some edge details). This can be confirmed, for example, in the upper-right edges in the driving example, and the person’s arm in the hotel example. Essentially, all denoising methods have

some parameters to control the amount of denoising; our proposed method has the signal ratio, which is intuitive; other optimization methods have some parameters about spatio-temporal kernel size, and learning-based approaches use the probability threshold for the network output. While we fix it for each dataset and condition for fair comparison (we do not tune it for each sequence in the results), this is an essential challenge for denoising algorithms in general, regardless of their underlying character (learning-based or model-based). Hence, we believe it is important for research to discuss denoising effectiveness in combination with downstream tasks, as shown in Secs. 4.3 and 4.4.

4.3. Joint Motion Estimation Results

Ego-Motion on ECD. One of the merits of the proposed joint estimation method is that it can improve motion estimation results. Here, the results of ego-motion (e.g., angular velocity) estimation are displayed in Fig. 5. We used the classical ECD dataset and injected BA noise from the DND21 recordings with different noise rates. As no prior work simultaneously estimates motion and noise, we compare against sequential filters: random sampling (of the input data with a uniform distribution) and BAF [9] that is the most widely-used denoising method to date. As shown in the IWEs, our method produces the sharpest edges (cf. (b) and (c)) and the most reasonable signal-noise classification among others (cf. (a), (b), and (c)). Note that the original ECD recordings have noise events besides the ones that are injected using DND21 recordings, which makes the calculation of ROC (similar to Sec. 3.2) intricate. Nonetheless, we report ROC results in the supplementary.

The quantitative evaluation on angular velocity estimation is reported in Fig. 6, as well as the summary of the supplementary detailed table in Tab. 3. As shown in the first row (“CMax”), the original CMax degrades due to the noise occurrences. Our method achieves the best results in terms of RMS and FWL, compared with the random sampling method as well as the original CMax (no denoising). Also, using better initialization (the result of the previous slice for sequential estimation) always improves the results (“w/ Init.”): the estimation accuracy relies on both the initialization and the denoising strategies. Hence, the result demonstrates that the proposed method relaxes the dependency on the initialization, thus improving the robustness of motion estimation.

Optical Flow on MVSEC. The results on more complex motion estimation (e.g., optical flow) are shown in Tab. 4, using the MVSEC dataset [63], with BA noise from the DND21 dataset. Due to the 5Hz BA noise, the baseline method MultiCM [48] degrades. Our denoising method improves the flow accuracy, showcasing the efficacy of simultaneous noise estimation. A challenge of evaluating denoising in these real-world datasets (e.g., ECD, MVSEC) is the

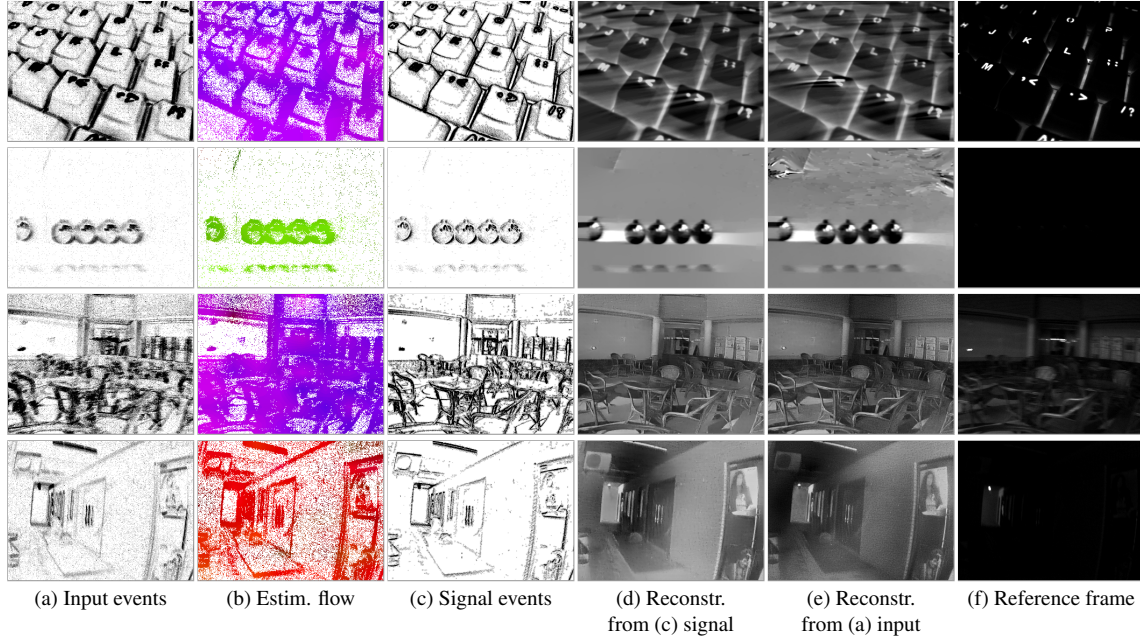


Figure 7. *Image intensity reconstruction* on E-MLB data [10]. The first two rows use the reconstruction method EVILIP [59] and the last two rows use E2VID [46]. Reconstruction on denoised events produces higher quality images (i.e., fewer artifacts) than on raw event data.

	<i>indoor1</i>		<i>indoor2</i>	
	EPE ↓	%Out ↓	EPE ↓	%Out ↓
MultiCM [48]	2.676	32.098	2.746	33.474
Ours	2.517	29.475	2.695	33.047

	<i>indoor3</i>		<i>outdoor1</i>	
	EPE ↓	%Out ↓	EPE ↓	%Out ↓
MultiCM [48]	2.732	33.219	1.960	18.160
Ours	2.654	32.241	1.865	17.554

Table 4. *Optical flow estimation results* on MVSEC [63] ($dt = 4$ frames, i.e., 89 ms) with BA noise at 5 Hz.

existing noise in the original sequences. Hence, we only evaluate motion estimation metrics, not ROC or AUC.

4.4. Application to Intensity Reconstruction

In real-world use cases, denoising works as preprocessing steps for other tasks (e.g., deblur, 3D reconstruction, motion estimation, etc.), hence it is important for denoising methods to demonstrate applicability. Although the proposed approach intrinsically improves motion estimation (Sec. 4.3), we further evaluate the denoising outcome via intensity reconstruction, a well-known task in event-based vision, using the E-MLB dataset [10]. Figure 7 collects the results. We use two state-of-the-art approaches for reconstruction: E2VID [46], which does not rely on motion, and EVILIP [59], which utilizes optical flow as prior.

Our method provides flow (Fig. 7, column (b)) and signal events (sharp IWEs in column (c)). Comparing the reconstruction with our event denoising method (column (d)) and

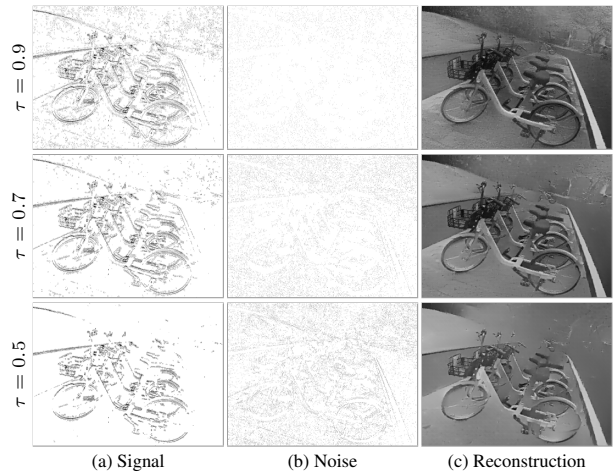


Figure 8. Sensitivity analysis: effect of different target ratio parameters τ on image reconstruction using EVILIP [59].

the reconstruction from the raw event data (column (e)), the images after event denoising have fewer artifacts than those without denoising, for both reconstruction methods tested. We observe that E2VID is more robust against noise than EVILIP, possibly due to noise events in its training (simulation) data. Due to the ND filters, the quality of the reference frames (last column) is limited, and hence, we discuss the results qualitatively. Nevertheless comparing the reconstructed images to the reference frames, evidences the HDR advantages of event cameras over frame-based ones.

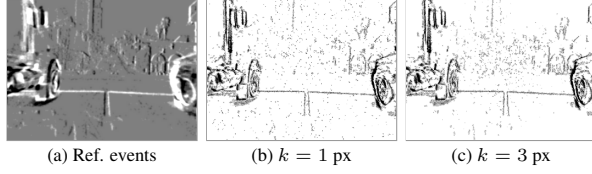


Figure 9. Results of different Gaussian kernel sizes k for sensitivity on different depth scenes.

4.5. Sensitivity

Effect of Noise Ratio. As Sec. 4.2 already showed the sensitivity for denoising accuracy, let us further discuss the sensitivity using image reconstruction. Figure 8 presents the result of the *Bicycle* sequence with ND64 filter [10]. While the GT image is highly underexposed and not informative (as those on the last column of Fig. 7), the reconstructed images clearly reflect the choice of the signal target ratio τ . As more events are removed, (i) the reconstructed image becomes more homogeneous but preserves strong edges, and (ii) the removed events (2nd col.) contain more scene edges (i.e., signal). We quantitatively analyze the former in the supplementary using non-reference image quality metrics.

Effect of the Kernel Size. Figure 9 shows the sensitivity analysis for different kernel sizes. By increasing the size of the Gaussian kernel in the IWE, our method can emphasize contour edges over isolated points, thus better preserving signal events in regions with low IWE intensities (center of the image, corresponding to far away scene points).

4.6. Ablation of Motion Estimator

The proposed approach can be combined with other estimation methods, such as deep-neural-network (DNN) flow estimators. To this end, we show an example of E-RAFT [21]. The model-based method alone [48] is known to output a degraded flow in night sequences with lots of noise, such as *zurich_city_12a* [20]. As shown in Fig. 10, our method can identify and discard noisy events. E-RAFT provides sharper and more natural edges in IWEs after denoising compared to MultiCM [48] (thanks to its more accurate flow). ESR scores are similar: 2.39 for E-RAFT and 2.45 for MultiCM. While we focus on the inference of DNN in the proposed pipeline, future research could look into accommodating training within the proposed joint estimation framework.

4.7. Runtime

Using 30k events from the ECD dataset (240×180 px) on a Mac M1 CPU (2020), the denoising operations (scoring and ranking) take less than 0.1 s extra computation in total, out of 2 s of the entire processing time (i.e., 5% computation increase with respect to the base method). Nonetheless, accommodating learning-based approaches in our framework could be key to the trade-off between accuracy and speed.

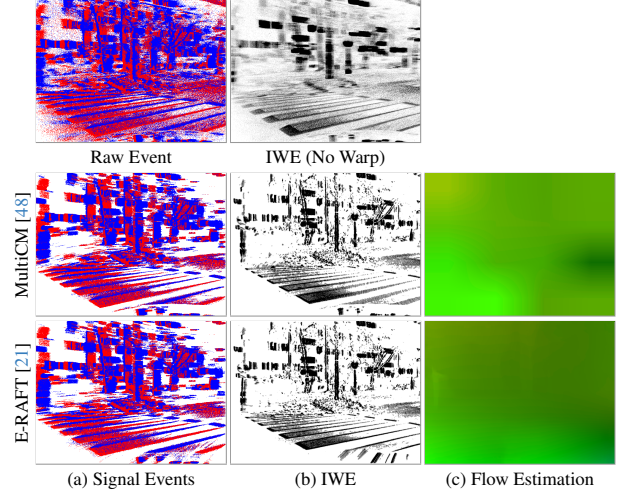


Figure 10. Effect of adopting two different motion estimators (E-RAFT or MultiCM) on the DSEC night scene (*zurich_12a*) [20]. Despite small IWE differences (e.g., in the car’s hood), the proposed framework supports several motion estimators.

5. Limitations

Our method expects as parameter a target noise level, but the intrinsic (“true”) noise level depends on the camera, scene, etc. and is therefore unknown. More research in this direction is needed for automatic parameter tuning. Also, notice that there might be a potential limitation of the denoising metric ESR, since it depends on the motion parameter (see Suppl. Mat. of [29]). Although our denoising efficacy is confirmed from multiple perspectives, we believe that advancing denoising metrics for real-world data (where it gets interesting due to the absence of GT labels) is a central direction for future work in event-based vision.

6. Conclusion

In summary, this work proposes that: (i) Joint estimation of noise and motion is a novel problem setting compared to conventional (isolated) denoising, and is sensitive given the event data’s inherent motion dependency. (ii) The proposed joint estimation framework can accommodate various motion estimators and has shown its efficacy in improving both denoising and motion estimation accuracies. (iii) It is important to evaluate denoising results on both, denoising benchmarks and downstream tasks (e.g., motion estimation, intensity reconstruction), since current denoising benchmarks have shortcomings in terms of metrics, data, and baselines. In particular, our method achieves notable denoising results (state-of-the-art on the E-MLB benchmark and competitive on DND21) while demonstrating effectiveness across motion estimation and intensity reconstruction tasks. Our approach advances both event-denoising theory and practical denoising use-cases.

Acknowledgments

We would like to thank Dr. B. Jiang for useful discussions. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2002/1 “Science of Intelligence” – project number 390523135.

References

- [1] Saeed Afshar, Andrew Peter Nicholson, Andre van Schaik, and Gregory Cohen. Event-based object detection and tracking for space situational awareness. *IEEE Sensors Journal*, 20(24):15117–15132, 2020.
- [2] Yusra Alkendi, Rana Azzam, Abdulla Ayyad, Sajid Javed, Lakmal Seneviratne, and Yahya Zweiri. Neuromorphic camera denoising using graph neural network-driven transformers. *IEEE Trans. Neural Netw. Learn. Syst.*, 35(3):4110–4124, 2022.
- [3] Sami Arja, Alexandre Marcireau, Richard L. Balthazor, Matthew G. McHarg, Saeed Afshar, and Gregory Cohen. Density invariant contrast maximization for neuromorphic earth observations. In *IEEE Conf. Comput. Vis. Pattern Recog. Workshops (CVPRW)*, pages 3984–3994, 2023.
- [4] R Baldwin, Mohammed Almatrafi, Vijayan Asari, and Keigo Hirakawa. Event probability mask (epm) and event denoising convolutional neural network (edncnn) for neuromorphic cameras. In *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, pages 1701–1710, 2020.
- [5] R Wes Baldwin, Mohammed Almatrafi, Jason R Kaufman, Vijayan Asari, and Keigo Hirakawa. Inceptive event time-surfaces for object classification using neuromorphic cameras. In *Image Analysis and Recognition: 16th Int. Conf., ICIAR 2019, Waterloo, ON, Canada, August 27–29, 2019, Proceedings, Part II 16*, pages 395–403. Springer, 2019.
- [6] T. D. Barfoot. *State Estimation for Robotics - A Matrix Lie Group Approach*. Cambridge University Press, 2015.
- [7] Christian Brandli, Raphael Berner, Minhao Yang, Shih-Chii Liu, and Tobi Delbruck. A 240x180 130dB 3 μ s latency global shutter spatiotemporal vision sensor. *IEEE J. Solid-State Circuits*, 49(10):2333–2341, 2014.
- [8] Jinze Chen, Yang Wang, Yang Cao, Feng Wu, and Zheng-Jun Zha. Progressivemotionseg: Mutually reinforced framework for event-based motion segmentation. In *AAAI Conf. Artificial Intell.*, pages 303–311, 2022.
- [9] Tobi Delbruck. Frame-free dynamic digital vision. In *Proc. Int. Symp. Secure-Life Electron.*, pages 21–26, 2008.
- [10] Saizhe Ding, Jinze Chen, Yang Wang, Yu Kang, Weiguo Song, Jie Cheng, and Yang Cao. E-MLB: Multilevel benchmark for event-based camera denoising. *IEEE Trans. Multimedia*, 26:65–76, 2023.
- [11] Peiqi Duan, Zihao Wang, Boxin Shi, Oliver Cossairt, Tiejun Huang, and Aggelos Katsaggelos. Guided event filtering: Synergy between intensity images and neuromorphic events for high performance imaging. *IEEE Trans. Pattern Anal. Mach. Intell.*, pages 1–1, 2021.
- [12] Peiqi Duan, Zihao W. Wang, Xinyu Zhou, Yi Ma, and Boxin Shi. EventZoom: Learning to denoise and super resolve neuromorphic events. In *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, pages 12824–12833, 2021.
- [13] Huachen Fang, Jinjian Wu, Leida Li, Junhui Hou, Weisheng Dong, and Guangming Shi. Aednet: Asynchronous event denoising with spatial-temporal correlation among irregular data. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 1427–1435, 2022.
- [14] Yang Feng, Hengyi Lv, Hailong Liu, Yisa Zhang, Yuyao Xiao, and Chengshan Han. Event density based denoising method for dynamic vision sensor. *Applied Sciences*, 10(6): 2024, 2020.
- [15] Guillermo Gallego and Davide Scaramuzza. Accurate angular velocity estimation with an event camera. *IEEE Robot. Autom. Lett.*, 2(2):632–639, 2017.
- [16] Guillermo Gallego and Anthony Yezzi. A compact formula for the derivative of a 3-D rotation in exponential coordinates. *J. Math. Imaging Vis.*, 51(3):378–384, 2014.
- [17] Guillermo Gallego, Henri Rebecq, and Davide Scaramuzza. A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation. In *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, pages 3867–3876, 2018.
- [18] Guillermo Gallego, Mathias Gehrig, and Davide Scaramuzza. Focus is all you need: Loss functions for event-based vision. In *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, pages 12272–12281, 2019.
- [19] Guillermo Gallego, Tobi Delbruck, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew Davison, Jörg Conradt, Kostas Daniilidis, and Davide Scaramuzza. Event-based vision: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(1):154–180, 2022.
- [20] Mathias Gehrig, Willem Aarents, Daniel Gehrig, and Davide Scaramuzza. DSEC: A stereo event camera dataset for driving scenarios. *IEEE Robot. Autom. Lett.*, 6(3):4947–4954, 2021.
- [21] Mathias Gehrig, Mario Millhäusler, Daniel Gehrig, and Davide Scaramuzza. E-RAFT: Dense optical flow from event cameras. In *Int. Conf. 3D Vision (3DV)*, pages 197–206, 2021.
- [22] Rui Graca and Tobi Delbruck. Towards a physically realistic computationally efficient dvs pixel model. In *Int. Image Sensor Workshop (IISW)*, 2025.
- [23] Cheng Gu, Erik Learned-Miller, Daniel Sheldon, Guillermo Gallego, and Pia Bideau. The spatio-temporal Poisson point process: A simple model for the alignment of event camera data. In *Int. Conf. Comput. Vis. (ICCV)*, pages 13495–13504, 2021.
- [24] Shasha Guo and Tobi Delbruck. Low cost and latency event camera background activity denoising. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(1):785–795, 2023.
- [25] Shuang Guo and Guillermo Gallego. CMax-SLAM: Event-based rotational-motion bundle adjustment and SLAM system using contrast maximization. *IEEE Trans. Robot.*, 40: 2442–2461, 2024.
- [26] Shasha Guo, Ziyang Kang, Lei Wang, Shiming Li, and Weixia Xu. Hashheat: An O(C) complexity hashing-based

- filter for dynamic vision sensor. In *25th Asia and South Pacific Design Automation Conference (ASP-DAC)*, pages 452–457, 2020.
- [27] Shuang Guo, Friedhelm Hamann, and Guillermo Gallego. Unsupervised joint learning of optical flow and intensity with event cameras. In *Int. Conf. Comput. Vis. (ICCV)*, 2025.
- [28] Friedhelm Hamann, Ziyun Wang, Ioannis Asmanis, Kenneth Chaney, Guillermo Gallego, and Kostas Daniilidis. Motion-prior contrast maximization for dense continuous-time motion estimation. In *Eur. Conf. Comput. Vis. (ECCV)*, pages 18–37, 2024.
- [29] Bin Jiang, bo Xiong, Bohan Qu, M. Slaman Asif, you Zhou, and Zhan Ma. EDformer: Transformer-based event denoising across varied noise levels. In *Eur. Conf. Comput. Vis. (ECCV)*, pages 1–12, 2024.
- [30] Pritam P. Karmokar, Quan H. Nguyen, and William J. Beks. Secrets of edge-informed contrast maximization for event-based vision. In *IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, 2025.
- [31] Alireza Khodamoradi and Ryan Kastner. O(N)-space spatiotemporal filter for reducing noise in neuromorphic vision sensors. *IEEE Trans. Emerg. Topics Comput.*, PP(99):1–1, 2018.
- [32] Haram Kim and H. Jin Kim. Real-time rotational motion estimation with contrast maximization over globally aligned events. *IEEE Robot. Autom. Lett.*, 6(3):6016–6023, 2021.
- [33] Simon Klenk, Jason Chui, Nikolaus Demmel, and Daniel Cremers. TUM-VIE: The TUM stereo visual-inertial event dataset. In *IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*, pages 8601–8608, 2021.
- [34] Xavier Lagorce, Garrick Orchard, Francesco Gallupi, Bertram E. Shi, and Ryad Benosman. HOTS: A hierarchy of event-based time-surfaces for pattern recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39(7):1346–1359, 2017.
- [35] Patrick Lichtsteiner, Christoph Posch, and Tobi Delbruck. A 128x128 120dB 30mW asynchronous vision sensor that responds to relative intensity change. In *IEEE Int. Solid-State Circuits Conf. (ISSCC)*, pages 2060–2069, 2006.
- [36] Patrick Lichtsteiner, Christoph Posch, and Tobi Delbruck. A 128x128 120 dB 15 μ s latency asynchronous temporal contrast vision sensor. *IEEE J. Solid-State Circuits*, 43(2):566–576, 2008.
- [37] Hongjie Liu, Christian Brandli, Chenghan Li, Shih-Chii Liu, and Tobi Delbruck. Design of a spatiotemporal correlation filter for event-based sensors. In *IEEE Int. Symp. Circuits Syst. (ISCAS)*, pages 722–725, 2015.
- [38] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Trans. Image Process.*, 21(12):4695–4708, 2012.
- [39] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a “completely blind” image quality analyzer. *IEEE Signal Process. Lett.*, 20(3):209–212, 2012.
- [40] 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. *Int. J. Robot. Research*, 36(2):142–149, 2017.
- [41] Elias Mueggler, Guillermo Gallego, Henri Rebecq, and Davide Scaramuzza. Continuous-time visual-inertial odometry for event cameras. *IEEE Trans. Robot.*, 34(6):1425–1440, 2018.
- [42] Urbano Miguel Nunes and Yiannis Demiris. Robust event-based vision model estimation by dispersion minimisation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(12):9561–9573, 2022.
- [43] Federico Paredes-Vallés and Guido C. H. E. de Croon. Back to event basics: Self-supervised learning of image reconstruction for event cameras via photometric constancy. In *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, pages 3445–3454, 2021.
- [44] Federico Paredes-Vallés, Kirk YW Scheper, Christophe De Wagter, and Guido CHE de Croon. Taming contrast maximization for learning sequential, low-latency, event-based optical flow. In *Int. Conf. Comput. Vis. (ICCV)*, pages 9661–9671, 2023.
- [45] Xin Peng, Ling Gao, Yifu Wang, and Laurent Kneip. Globally-optimal contrast maximisation for event cameras. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(7):3479–3495, 2022.
- [46] Henri Rebecq, René Ranftl, Vladlen Koltun, and Davide Scaramuzza. High speed and high dynamic range video with an event camera. *IEEE Trans. Pattern Anal. Mach. Intell.*, 43(6):1964–1980, 2021.
- [47] Antoni Rosinol Vidal, Henri Rebecq, Timo Horstschaefer, and Davide Scaramuzza. Ultimate SLAM? combining events, images, and IMU for robust visual SLAM in HDR and high speed scenarios. *IEEE Robot. Autom. Lett.*, 3(2):994–1001, 2018.
- [48] Shintaro Shiba, Yoshimitsu Aoki, and Guillermo Gallego. Secrets of event-based optical flow. In *Eur. Conf. Comput. Vis. (ECCV)*, pages 628–645, 2022.
- [49] Shintaro Shiba, Yoshimitsu Aoki, and Guillermo Gallego. Event collapse in contrast maximization frameworks. *Sensors*, 22(14):1–20, 2022.
- [50] Shintaro Shiba, Yannick Klose, Yoshimitsu Aoki, and Guillermo Gallego. Secrets of event-based optical flow, depth, and ego-motion by contrast maximization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 46(12):7742–7759, 2024.
- [51] Timo Stoffregen, Guillermo Gallego, Tom Drummond, Lindsay Kleeman, and Davide Scaramuzza. Event-based motion segmentation by motion compensation. In *Int. Conf. Comput. Vis. (ICCV)*, pages 7243–7252, 2019.
- [52] Timo Stoffregen, Cedric Scheerlinck, Davide Scaramuzza, Tom Drummond, Nick Barnes, Lindsay Kleeman, and Robert Mahony. Reducing the sim-to-real gap for event cameras. In *Eur. Conf. Comput. Vis. (ECCV)*, pages 534–549, 2020.
- [53] Gemma Taverni, Diederik Paul Moeys, Chenghan Li, Celso Cavaco, Vasyl Motsnyi, David San Segundo Bello, and Tobi Delbruck. Front and back illuminated Dynamic and Active Pixel Vision Sensors comparison. *IEEE Trans. Circuits Syst. II (TCSII)*, 65(5):677–681, 2018.
- [54] Yanxiang Wang, Bowen Du, Yiran Shen, Kai Wu, Guangrong Zhao, Jianguo Sun, and Hongkai Wen. Ev-gait: Event-based robust gait recognition using dynamic vision sensors.

- In *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, pages 6358–6367, 2019.
- [55] Zihao W. Wang, Peiqi Duan, Oliver Cossairt, Aggelos Katsaggelos, Tiejun Huang, and Boxin Shi. Joint filtering of intensity images and neuromorphic events for high-resolution noise-robust imaging. In *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, pages 1606–1616, 2020.
 - [56] Jinjian Wu, Chuanwei Ma, Leida Li, Weisheng Dong, and Guangming Shi. Probabilistic undirected graph based denoising method for dynamic vision sensor. *IEEE Trans. Multimedia*, 23:1148–1159, 2020.
 - [57] Ryo Yamaki, Shintaro Shiba, Gallego Guillermo, and Yoshimitsu Aoki. Iterative event-based motion segmentation by variational contrast maximization. In *IEEE Conf. Comput. Vis. Pattern Recog. Workshops (CVPRW)*, pages 4918–4927, 2025.
 - [58] Pei Zhang, Haosen Liu, Zhou Ge, Chutian Wang, and Edmund Y Lam. Neuromorphic imaging with joint image deblurring and event denoising. *IEEE Trans. Image Process.*, 2024.
 - [59] Zelin Zhang, Anthony Yezzi, and Guillermo Gallego. Formulating event-based image reconstruction as a linear inverse problem with deep regularization using optical flow. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(7):8372–8389, 2023.
 - [60] Yi Zhou, Guillermo Gallego, Xiuyuan Lu, Siqi Liu, and Shaojie Shen. Event-based motion segmentation with spatio-temporal graph cuts. *IEEE Trans. Neural Netw. Learn. Syst.*, pages 1–13, 2021.
 - [61] Yi Zhou, Guillermo Gallego, and Shaojie Shen. Event-based stereo visual odometry. *IEEE Trans. Robot.*, 37(5):1433–1450, 2021.
 - [62] Alex Zihao Zhu, Nikolay Atanasov, and Kostas Daniilidis. Event-based visual inertial odometry. In *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, pages 5816–5824, 2017.
 - [63] Alex Zihao Zhu, Dinesh Thakur, Tolga Ozaslan, Bernd Pfrommer, Vijay Kumar, and Kostas Daniilidis. The multivehicle stereo event camera dataset: An event camera dataset for 3D perception. *IEEE Robot. Autom. Lett.*, 3(3):2032–2039, 2018.
 - [64] Alex Zihao Zhu, Liangzhe Yuan, Kenneth Chaney, and Kostas Daniilidis. Unsupervised event-based learning of optical flow, depth, and egomotion. In *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, pages 989–997, 2019.