Secrets of Edge-Informed Contrast Maximization for Event-Based Vision

Pritam P. Karmokar*, Quan H. Nguyen*, and William J. Beksi
Department of Computer Science and Engineering, The University of Texas at Arlington, *Equal contribution

WACV 2025 TUCSON, ARIZONA • FEB 28 - MAR 4

UNIVERSITY OF TEXAS ROBOTIC VISION

The Goal

Extend uni-modal contrast maximization, an optimization framework to estimate motion trajectories of events, to a bi-modal edge-informed contrast maximization (**EINCM**) framework.

Introduction

- Event-based cameras capture the motion of intensity gradients (edges) in the image plane in the form of rapid asynchronous events.
- When accumulated in 2D histograms, these events depict overlays of the edges in motion, consequently obscuring the spatial structure of the generating edges.
- Contrast maximization (CM) is an optimization framework that can reverse this effect by transporting events along their estimated motion trajectories to produce sharp spatial structures that resemble the moving edges.

Figure 1: Events and edge images (top), and corresponding IWEs with zero and optimal warps (bottom).

Our Contributions (EINCM)

- Extended CM – simultaneously maximized image of warped events (IWE) *contrast* and event–edge *correlation*
- Introduced bi-modal multiscale and multireference
- Developed processing strategies – improved convergence and performance
- Scored best sharpness – state-of-the-art event-based optical flow benchmarks on MVSEC, DSEC, and ECD datasets

Background

CM was introduced in [1] as a model-based method to estimate motion parameters such that warped events achieve maximum alignment. However, these parameters may converge to a global optimum that warps events to too few pixels (i.e., event collapse). This issue was mitigated by [2] using several "secrets" for regularization, namely, patch-based processing and multiple warp reference times. On the other hand, the useful relationship between images and events, which can be captured simultaneously by systems such as the DAVIS346, has been underexplored and is the motivation for our work.

Method

A warping function $\mathbf{W}(\mathbf{x}_k, t_k; \boldsymbol{\theta}_k, t_{\text{ref}})$ defines a mapping $\mathcal{E} \mapsto \mathcal{E}'_{t_{\text{ref}}}$ that transports $\mathbf{x}_k \in \mathbf{X}_{\mathcal{E}}$ to $\mathbf{x}'_k \in \mathbf{X}_{\mathcal{E}'_{t_{\text{ref}}}}$ at the reference time t_{ref} using the motion parameter $\boldsymbol{\theta}_k \in \boldsymbol{\Theta}$ according to

$$\mathbf{x}'_k \doteq \mathbf{W}(\mathbf{x}_k, t_k; \boldsymbol{\theta}_k, t_{\text{ref}}) = \mathbf{x}_k + \boldsymbol{\theta}_k(t_{\text{ref}} - t_k), \quad 1 \le k \le N_e. \tag{1}$$

These warped events $\mathcal{E}'_{t_{\text{ref}}}$ are utilized to construct an IWE by organizing them by their coordinates $\mathbf{x}'_k \in \mathbf{X}_{\mathcal{E}'_{t_{\text{ref}}}}$ and aggregating them on the image plane as

$$I_{\text{events}}(\mathbf{x}; \boldsymbol{\Theta}, t_{\text{ref}}) \doteq \sum_{k=1}^{N_e} \delta(\mathbf{x} - \mathbf{x}'_k). \tag{2}$$

The IWE *contrast* and its *correlation* with the edge image E are

Contrast

$$G(\boldsymbol{\Theta}; t_{\text{ref}}) \doteq \frac{1}{|\Omega|} \int_{\Omega} \|\nabla I_{\text{events}}(\mathbf{x}; \boldsymbol{\Theta}, t_{\text{ref}})\|^2 d\mathbf{x}, \tag{3}$$

Correlation

$$\text{MSE}(\boldsymbol{\Theta}; t_{\text{ref}}) = \frac{1}{|\Omega|} \int_{\Omega} \big(I_{\text{events}}(\mathbf{x}; \boldsymbol{\Theta}, t_{\text{ref}}) - E(\mathbf{x}, t_{\text{ref}})\big)^2 d\mathbf{x}. \tag{4}$$

These contrast–correlation energies are accumulated over multiple frames N_{img}

$$f_{\text{rel}}(\boldsymbol{\Theta}) = \frac{1}{N_{\text{img}} \cdot G(\mathbf{0}_{\boldsymbol{\Theta}}; -)} \sum_{j=1}^{N_{\text{img}}} G(\boldsymbol{\Theta}; t_j), \tag{5}$$

$$g_{\text{rel}}(\boldsymbol{\Theta}) = -\frac{1}{N_{\text{img}}} \sum_{j=1}^{N_{\text{img}}} \frac{\text{MSE}(\boldsymbol{\Theta}; t_{\text{ref}})}{\text{MSE}(\mathbf{0}_{\boldsymbol{\Theta}}; -)}. \tag{6}$$

The motion parameters are optimized using contrast–correlation maximization

C²M

$$\boldsymbol{\Theta}^* = \arg\max_{\boldsymbol{\Theta}} \big(\alpha f_{\text{rel}}(\boldsymbol{\Theta}) + \beta g_{\text{rel}}(\boldsymbol{\Theta}) + \gamma \mathcal{R}(\boldsymbol{\Theta})\big). \tag{7}$$

Figure 2: The *pre-handover* multiscale strategy with an optimize-handover-upsample pipeline.

Figure 3: Edge extraction pipeline.

Results

	$dt = 1$	indoor_flying1 AEE ↓	indoor_flying1 %Out ↓	indoor_flying2 AEE ↓	indoor_flying2 %Out ↓	indoor_flying3 AEE ↓	indoor_flying3 %Out ↓	outdoor_day1 AEE ↓	outdoor_day1 %Out ↓
SL	EV-FlowNet-EST	0.97	0.91	1.38	8.20	1.43	6.47	–	–
SL	EV-FlowNet+	0.56	1.00	0.66	1.00	0.59	1.00	0.68	0.99
SL	E-RAFT	–	–	–	–	–	–	**0.24**	1.70
SSL	EV-FlowNet (original)	1.03	2.20	1.72	15.1	1.53	11.90	0.49	0.20
SSL	Spike-FlowNet	0.84	–	1.28	–	1.11	–	0.49	–
SSL	STE-FlowNet	0.57	0.10	0.79	1.60	0.72	1.30	0.42	**0.00**
USL	EV-FlowNet	0.58	**0.00**	1.02	4.00	0.87	3.00	0.32	**0.00**
USL	FireFlowNet	0.97	2.60	1.67	15.30	1.43	11.00	1.06	6.60
USL	ConvGRU-EV-FlowNet	0.60	0.51	1.17	8.06	0.93	5.64	0.47	0.25
MB	RTEF	0.52	0.10	0.98	5.50	0.71	2.10	0.53	0.20
MB	MultiCM	0.42	0.09	0.60	0.59	0.50	0.29	0.72_{6}	5.49_{4}
MB	Ours (CM events only)	0.37_{8}	0.05_{5}	0.50_{9}	0.17_{7}	0.44_{8}	0.05_{7}	0.64_{9}	4.22_{6}
MB	Ours (EINCM)	$\mathbf{0.37_{2}}$	0.04_{6}	$\mathbf{0.50_{5}}$	$\mathbf{0.15_{5}}$	$\mathbf{0.43_{7}}$	$\mathbf{0.02_{9}}$	0.61_{8}	3.69_{5}
	$dt = 4$								
SSL	EV-FlowNet (original)	2.25	24.70	4.05	45.30	3.45	39.70	**1.23**	**7.30**
SSL	Spike-FlowNet	2.24	–	3.83	–	3.18	–	1.09	–
SSL	STE-FlowNet	1.77	14.70	2.52	26.10	2.23	22.10	0.99	3.90
USL	EV-FlowNet	2.18	24.20	3.85	46.80	3.18	47.80	1.30	9.70
USL	ConvGRU-EV-FlowNet	2.16	21.50	3.90	40.72	3.00	29.60	1.69	12.50
MB	MultiCM	1.68	12.79	2.49	26.31	2.06	18.93	2.07	19.99
MB	Ours (CM events only)	1.53_{8}	9.48_{7}	2.38_{7}	21.84_{6}	1.92_{6}	15.37_{6}	2.83_{8}	25.01_{8}
MB	Ours (EINCM)	$\mathbf{1.43_{9}}$	$\mathbf{7.78_{9}}$	$\mathbf{1.97_{3}}$	$\mathbf{17.18_{3}}$	$\mathbf{1.70_{6}}$	$\mathbf{12.33_{9}}$	1.70_{5}	16.01_{3}

Table 1: Quantitative results on MVSEC (**best**, second best). %Out indicates a 3-pixel error percentage.

Figure 4: Qualitative comparisons ($dt = 4$) of our approach against two prominent methods [2, 3] on MVSEC. Column (a) shows our preprocessed edge images and the images of (original) events. Column (b) displays the ground-truth (GT) flows and the corresponding IWEs. Columns (c-e) display the predicted flows masked by the original events and the constructed IWEs for each method.

Results (cont.)

	MVSEC ($dt = 4$)				ECD	DSEC (train)	
	ind_fly1	ind_fly2	ind_fly3	out_day1	slider_depth	thu_00_a	zur_07_a
Ground truth	1.09	1.20	1.12	1.07	–	1.01	1.04
MultiCM	1.17	1.30	1.23	1.11	1.93	1.42	1.63
Ours (EINCM)	$\mathbf{1.33_{8}}$	$\mathbf{1.45_{5}}$	$\mathbf{1.39_{2}}$	$\mathbf{1.23_{9}}$	**1.97**	$\mathbf{1.53_{2}}$	$\mathbf{1.63_{6}}$

Table 2: Flow warp loss (FWL) on MVSEC, ECD, and DSEC (train). *Bold* indicates **best**.

Figure 5: Qualitative results for the sequences `slider_depth` (ECD), `thun_01_a` (DSEC), and `thun_01_b` (DSEC).

	int_00_b FWL ↑	int_01_a FWL ↑	thu_01_a FWL ↑	thu_01_b FWL ↑	zur_12_a FWL ↑	zur_14_c FWL ↑	zur_15_a FWL ↑
Ours (EINCM-SHO)	1.51_{3}	1.70_{2}	1.30_{6}	1.32_{6}	0.67_{4}	1.52_{6}	1.47_{5}
Ours (EINCM-FHO)	1.65_{9}	1.74_{6}	1.32_{6}	1.36_{7}	1.15_{9}	1.38_{6}	1.53_{7}
Ours (EINCM-FSHO)	$\mathbf{1.75_{8}}$	$\mathbf{1.75_{5}}$	$\mathbf{1.45_{8}}$	$\mathbf{1.40_{5}}$	$\mathbf{1.34_{3}}$	$\mathbf{1.54_{2}}$	$\mathbf{1.61_{3}}$

Table 3: *Handover* sub-strategy ablations evaluated on the DSEC test set. We report FWL scores using three *handover* sub-strategies: solved handover (SHO), fixed handover (FHO), and fixed+solved handover (FSHO). For SHO, w_{ho}^* is solved by optimizing for the CM objective at all pyramid levels. For FHO, $w_{\text{ho}} = 0.5$ for all pyramid levels. For FSHO, $w_{\text{ho}} = 0.5$ for levels 4, 3, 2, and solved w_{ho}^* at levels 1 and 0. *Bold* and *underline* typefaces indicate the **best** and the second best, respectively.

References

[1] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3867–3876, 2018.

[2] Shintaro Shiba, Yoshimitsu Aoki, and Guillermo Gallego. Secrets of event-based optical flow. In *Proceedings of the European Conference on Computer Vision*, pages 628–645. Springer, 2022.

[3] Alex Zihao Zhu, Liangzhe Yuan, Kenneth Chaney, and Kostas Daniilidis. Ev-flownet: Self-supervised optical flow estimation for event-based cameras. In *Proceedings of Robotics: Science and Systems*, 2018.

Acknowledgments

This material is based upon work supported by the Air Force Research Laboratory under award number FA8571-23-C-0041.