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

Pritam P. Karmokar, Quan H. Nguyen, and William J. Beksi

WACV 2025
TUCSON, ARIZONA • FEB 28 - MAR 4

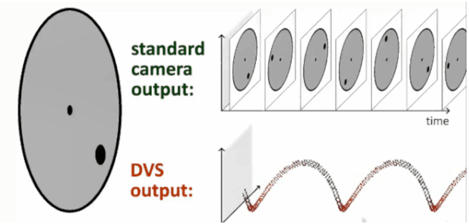
UNIVERSITY OF TEXAS ROBOTIC VISION

Outline

Outline

Event-Based Cameras

- Event-based cameras measure only motion in the scene
- Advantages:
  - High temporal resolution (μs)
  - Low latency (sub-ms)
  - Low bandwidth
  - Low power (mW)
  - High dynamic range (120 dB)



Video source: [1]

[1] P. Lichtsteiner et al., 2008, "A 128× 128 120 dB 15 μs latency asynchronous temporal contrast vision sensor."

Outline

Event Optical Flow – Related Works

- Model-based approaches:
  - Local plane-fitting [2]
  - Block-matching [3]
- Learning-based approaches:

[2] R. Benosman et al., 2012, "Asynchronous frameless event-based optical flow."
[3] M. Liu et al., 2018, "Adaptive time-slice block-matching optical flow algorithm for dynamic vision sensors."

Event Optical Flow – Related Works

- Model-based approaches:
  - Local plane-fitting [2]
  - Block-matching [3]
  - Lucas-Kanade variants [4]

- Learning-based approaches:

[2] R. Benosman et al., 2012, "Asynchronous frameless event-based optical flow."
[3] M. Liu et al., 2018, "Adaptive time-slice block-matching optical flow algorithm for dynamic vision sensors."
[4] D. Gehrig et al., 2020, "EKLT: Asynchronous photometric feature tracking using events and frames."

Event Optical Flow – Related Works

- Model-based approaches:
  - Local plane-fitting [2]
  - Block-matching [3]
  - Lucas-Kanade variants [4]
  - Contrast maximization (CM) [5]
- Learning-based approaches:

[2] R. Benosman et al., 2012, "Asynchronous frameless event-based optical flow."
[3] M. Liu et al., 2018, "Adaptive time-slice block-matching optical flow algorithm for dynamic vision sensors."
[4] D. Gehrig et al., 2020, "EKLT: Asynchronous photometric feature tracking using events and frames."
[5] G. Gallego et al., 2018, "A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation."

Event Optical Flow – Related Works

- Model-based approaches:
  - Local plane-fitting [2]
  - Block-matching [3]
  - Lucas-Kanade variants [4]
  - Contrast maximization (CM) [5]
- Learning-based approaches:
  - Largely adopt frame-based models [6], [7]
  - Rarely exploit working principles of camera modalities

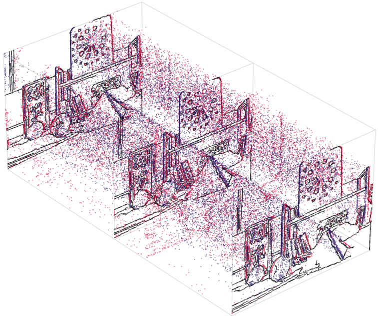
[2] R. Benosman et al., 2012, "Asynchronous frameless event-based optical flow."
[3] M. Liu et al., 2018, "Adaptive time-slice block-matching optical flow algorithm for dynamic vision sensors."
[4] D. Gehrig et al., 2020, "EKLT: Asynchronous photometric feature tracking using events and frames."
[5] G. Gallego et al., 2018, "A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation."
[6] M. Gehrig et al., 2021, "E-raft: Dense optical flow from event cameras."
[7] A. Z. Zhu et al., 2018, "EV-FlowNet: Self-supervised optical flow estimation for event-based cameras."

Outline

Motivation and Contributions

Motivation:

Event-based cameras capture moving edges but ...



Motivation and Contributions

Motivation:

Event-based cameras capture moving edges but ...

- Can appear like smears
- Obscure spatial structure of generating edges

Motivation and Contributions

Motivation:

Event-based cameras capture moving edges but ...

- Can appear like smears
- Obscure spatial structure of generating edges
- CM [5] can reverse this effect

[5] G. Gallego et al., 2018, "A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation."

Motivation and Contributions

Motivation:

Event-based cameras capture moving edges but ...

- Can appear like smears
- Obscure spatial structure of generating edges
- CM [5] can reverse this effect
- Challenges such as *event collapse*

[5] G. Gallego et al., 2018, "A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation."

Motivation and Contributions

Motivation:

Event-based cameras capture moving edges but ...

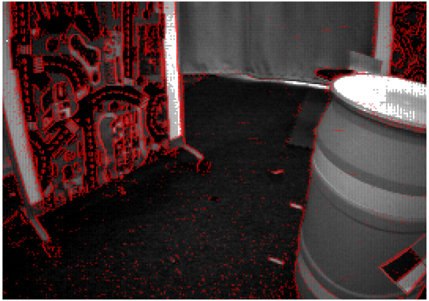
- Can appear like smears
- Obscure spatial structure of generating edges
- CM [5] can reverse this effect
- Challenges such as *event collapse*
- Mitigated by multireference [8]

[5] G. Gallego et al., 2018, "A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation."
[8] S. Shiba et al., 2022, "Secrets of event-based optical flow."

Motivation and Contributions

Hypothesis

Under optimal warp, the image of accumulated events is not only **sharp**, but also has **good spatial correlation** with its corresponding edge image.



Motivation and Contributions

Contributions:

1. Extend CM in a model-based setting by simultaneously maximizing *contrast* and event-edge *correlation*

Motivation and Contributions

Contributions:

1. Extend CM in a model-based setting by simultaneously maximizing *contrast* and event-edge *correlation*
2. Refine existing multiscale and multireference techniques for the bi-modal case

Motivation and Contributions

Contributions:

1. Extend CM in a model-based setting by simultaneously maximizing *contrast* and event-edge *correlation*
2. Refine existing multiscale and multireference techniques for the bi-modal case
3. Develop sophisticated sequential processing strategies to improve convergence and enhance performance

Motivation and Contributions

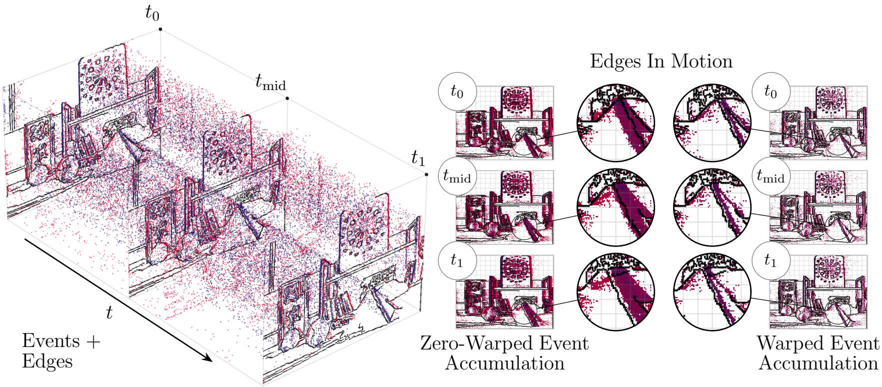
Contributions:

1. Extend CM in a model-based setting by simultaneously maximizing *contrast* and event-edge *correlation*
2. Refine existing multiscale and multireference techniques for the bi-modal case
3. Develop sophisticated sequential processing strategies to improve convergence and enhance performance
4. Score best sharpness and establish new event-based optical flow benchmarks on the MVSEC, DSEC, and ECD datasets

Outline

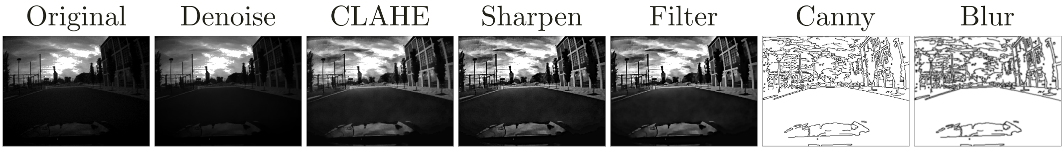
Problem Setting

- Inputs: Events and frames within a small time interval
- Outputs: Dense event optical flow estimation (pixel displacements)



Outline

Edge Extraction Pipeline



Outline

Linear Warping Model

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

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

The image of warped events (IWE), I_{events}, is constructed by organizing warped events, $\mathcal{E}'_{t_{\text{ref}}}$, by their coordinates, $\mathbf{x}'_k \in \mathbf{X}_{\mathcal{E}'_{t_{\text{ref}}}}$, and aggregating over the image plane as

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

Outline

Contrast and Edge Correlation

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

Contrast

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

Correlation

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

Outline

Multiple Frames = Multiple Reference Times

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

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

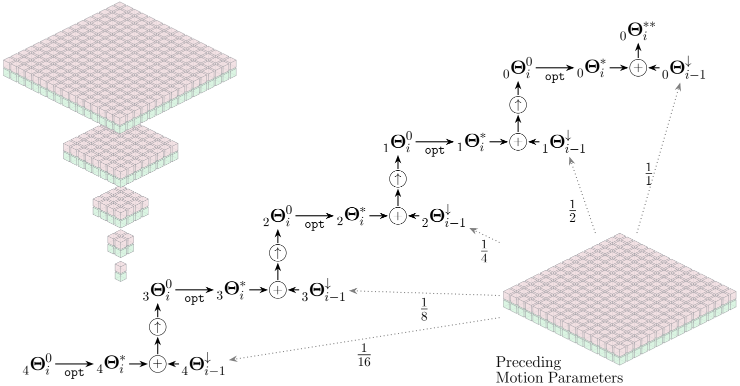
$$g_{\text{rel}}(\boldsymbol{\Theta}) = -\frac{1}{N_{\text{img}}} \sum_{j=1}^{N_{\text{img}}} \frac{\text{MSE}(\boldsymbol{\Theta}; t_{\text{ref}})}{\text{MSE}(\mathbf{0}_{\boldsymbol{\Theta}}; -)} \tag{6}$$

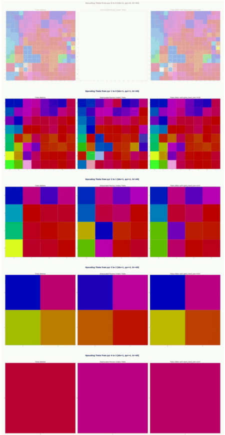
The motion parameters are optimized using contrast–correlation maximization

$$\boldsymbol{\Theta}^* = \underset{\boldsymbol{\Theta}}{\operatorname{argmax}} \Big(\alpha f_{\text{rel}}(\boldsymbol{\Theta}) + \beta g_{\text{rel}}(\boldsymbol{\Theta}) + \gamma \mathcal{R}(\boldsymbol{\Theta})\Big) \tag{7}$$

Outline

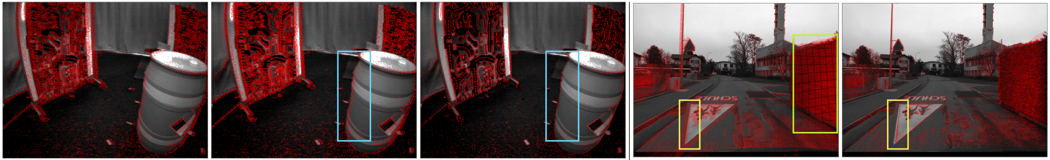
Multiscale and Sequential Processing





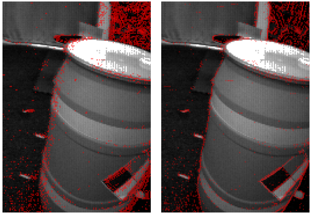
Outline

Imperfect Ground Truth | Imperfect Data

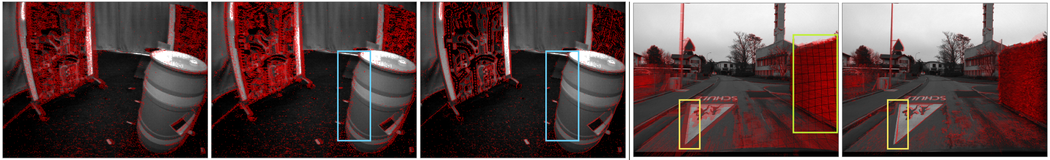


(a) I_{events} (b) IWE (GT) (c) IWE (Ours) (d) I_{events} (e) IWE (Ours)

- Our method yields sharper warped events (□)

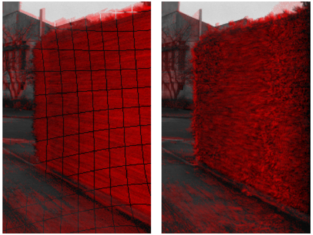


Imperfect Ground Truth | Imperfect Data

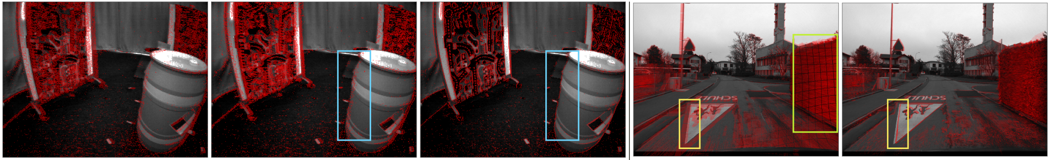


(a) I_{events} (b) IWE (GT) (c) IWE (Ours) (d) I_{events} (e) IWE (Ours)

- Our method yields sharper warped events (□)
- Grid-like rectification artifacts (□) pose additional challenges



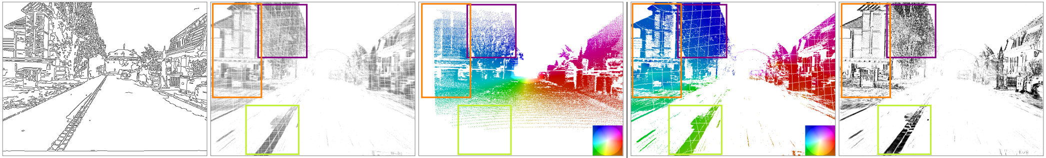
Imperfect Ground Truth | Imperfect Data



(a) I_{events} (b) IWE (GT) (c) IWE (Ours) (d) I_{events} (e) IWE (Ours)

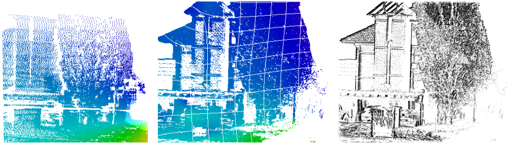
- Our method yields sharper warped events (□)
- Grid-like rectification artifacts (□) pose additional challenges
- Misalignment artifacts (*e.g.*, road markings □) at points near the camera

Absence of Ground Truth at Dynamic Points in Scene



(a) E (b) I_{events} (c) GT (d) Pred. Flow (Ours) (e) IWE (Ours)

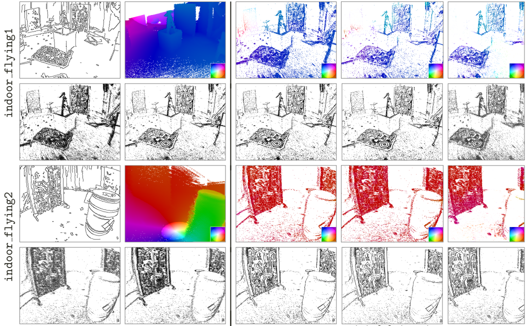
- Regions in the image, such as the top-left (□, □) and the bottom-center (□) are discounted from contributing to the accuracy due to unavailable GT



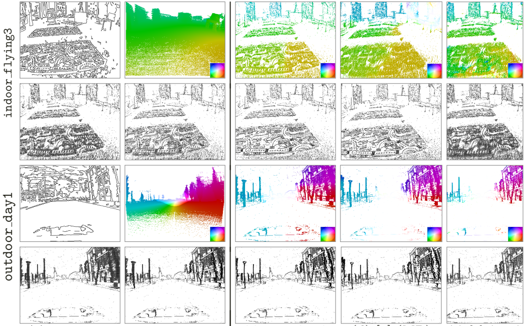
Note that although the IWE looks sharp (e), the predicted flow (d) is only evaluated by the benchmark where valid ground truth exists.

Outline

MVSEC Qualitative Comparisons

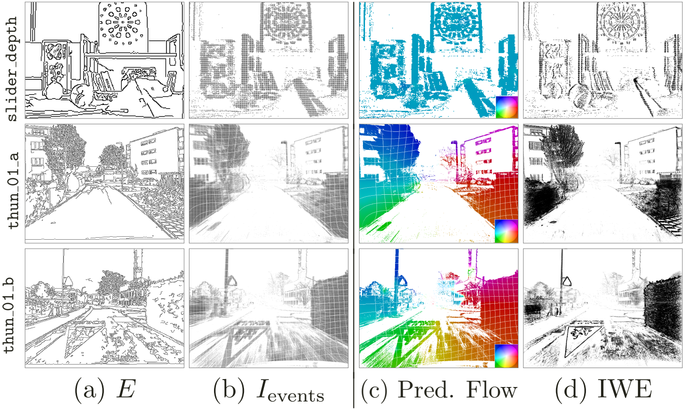


(a) E, I_{events} (b) GT (c) Ours (MB) (d) [8] (MB) (e) [7] (SSL)



(a) E, I_{events} (b) GT (c) Ours (MB) (d) [8] (MB) (e) [7] (SSL)

ECD and DSEC Qualitative Results



Outline

Quantitative Comparisons

		indoor_flying1		indoor_flying2		indoor_flying3		outdoor_day1	
	dt = 1	AEE ↓	%Out ↓	AEE ↓	%Out ↓	AEE ↓	%Out ↓	AEE ↓	%Out ↓
SL	EV-FlowNet-EST [9]	0.97	0.91	1.38	8.20	1.43	6.47	–	–
SL	EV-FlowNet+ [10]	0.56	1.00	0.66	1.00	0.59	1.00	0.68	0.99
SL	E-RAFT [6]	–	–	–	–	–	–	**0.24**	1.70
SSL	EV-FlowNet (original) [7]	1.03	2.20	1.72	15.1	1.53	11.90	0.49	0.20
SSL	Spike-FlowNet [11]	0.84	–	1.28	–	1.11	–	0.49	–
SSL	STE-FlowNet [12]	0.57	0.10	0.79	1.60	0.72	1.30	0.42	**0.00**
USL	EV-FlowNet [13]	0.58	**0.00**	1.02	4.00	0.87	3.00	$\underline{0.32}$	**0.00**
USL	FireFlowNet [14]	0.97	2.60	1.67	15.30	1.43	11.00	1.06	6.60
USL	ConvGRU-EV-FlowNet [15]	0.60	0.51	1.17	8.06	0.93	5.64	0.47	0.25
MB	Nagata [16]	0.62	–	0.93	–	0.84	–	0.77	–
MB	Akolkar [17]	1.52	–	1.59	–	1.89	–	2.75	–
MB	Brebion [18]	0.52	0.10	0.98	5.50	0.71	2.10	0.53	$\underline{0.20}$
MB	Shiba [8]	0.42	0.09	0.60	0.59	0.50	0.29	0.72_5	5.49_4
MB	Ours (CM events only)	$\underline{0.37_8}$	0.05_5	$\underline{0.50_9}$	$\underline{0.17_2}$	$\underline{0.44_3}$	$\underline{0.05_7}$	0.64_9	4.22_6
MB	Ours (EINCM)	$\mathbf{0.37_2}$	$\underline{0.04_6}$	$\mathbf{0.50_5}$	$\mathbf{0.15_5}$	$\mathbf{0.43_7}$	$\mathbf{0.02_9}$	0.61_8	3.69_3
	dt = 4								
SSL	EV-FlowNet (original) [7]	2.25	24.70	4.05	45.30	3.45	39.70	**1.23**	**7.30**
SSL	Spike-FlowNet [11]	2.24	–	3.83	–	3.18	–	1.09	–
SSL	STE-FlowNet [12]	1.77	14.70	2.52	26.10	2.23	22.10	0.99	3.90
USL	EV-FlowNet [13]	2.18	24.20	3.85	46.80	3.18	47.80	$\underline{1.30}$	$\underline{9.70}$
USL	ConvGRU-EV-FlowNet [15]	2.16	21.50	3.90	40.72	3.00	29.60	1.69	12.50
MB	Shiba [8]	1.68	12.79	2.49	26.31	2.06	18.93	2.07	19.99
MB	Ours (CM events only)	$\underline{1.53_9}$	$\underline{9.48_2}$	$\underline{2.38_2}$	$\underline{21.84_9}$	$\underline{1.92_3}$	$\underline{15.37_8}$	2.83_1	25.01_1
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_4	16.01_3

	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
Shiba [8]	1.17	1.30	1.23	1.11	1.93	1.42	1.63
Ours (EINCM)	$\mathbf{1.33_3}$	$\mathbf{1.45_5}$	$\mathbf{1.39_2}$	$\mathbf{1.23_0}$	**1.97**	$\mathbf{1.53_2}$	$\mathbf{1.63_6}$

Outline

Multiscale Handover Ablations

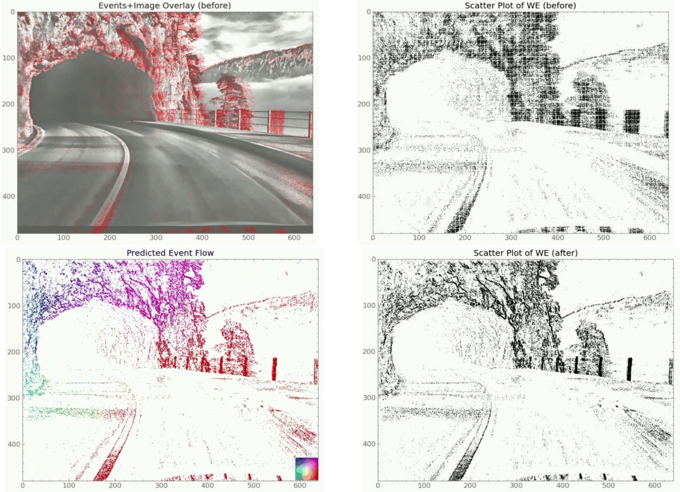
Handover sub-strategy ablations:

- FWL scores using three handover sub-strategies:
  1. Solved handover (SHO): w^*_{ho} is solved for all pyramid levels
  2. Fixed handover (FHO): $w_{\text{ho}} = 0.5$ for all pyramid levels
  3. Fixed+Solved handover (FSHO): $w_{\text{ho}} = 0.5$ for levels 4, 3, 2, and solved w^*_{ho} at levels 1 and 0

	int_00_b	int_01_a	thu_01_a	thu_01_b	zur_12_a	zur_14_c	zur_15_a
	FWL ↑	FWL ↑	FWL ↑	FWL ↑	FWL ↑	FWL ↑	FWL ↑
Ours (EINCM-SHO)	1.51_3	1.70_2	1.30_3	1.32_9	0.67_4	$\underline{1.52_6}$	1.47_3
Ours (EINCM-FHO)	$\underline{1.65_9}$	$\underline{1.74_6}$	$\underline{1.32_6}$	$\underline{1.36_2}$	$\underline{1.15_9}$	1.38_6	$\underline{1.53_7}$
Ours (EINCM-FSHO)	$\mathbf{1.75_8}$	$\mathbf{1.75_5}$	$\mathbf{1.45_8}$	$\mathbf{1.40_5}$	$\mathbf{1.34_3}$	$\mathbf{1.54_2}$	$\mathbf{1.61_3}$

Outline

Example Outcome for DSEC sequence `interlaken_00_b`



Outline

Conclusion

- We introduced a **bi-modal extension for CM** that estimates event optical flow using both events and frames

Conclusion

- We introduced a **bi-modal extension for CM** that estimates event optical flow using both events and frames
- We developed sophisticated **multiscale handover strategies**

Conclusion

- We introduced a **bi-modal extension for CM** that estimates event optical flow using both events and frames
- We developed sophisticated **multiscale handover strategies**
- Our method achieves **superior sharpness and competitive accuracy** on multiple real-world public datasets and benchmarks

Visit Our Poster