

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

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





Outline

1 Introduction

- Event-Based Cameras
- Event Optical Flow – Related Work
- Motivation and Contributions

2 Method

- The Problem
- Extracting Edges
- The Warp Model
- Bi-Modal Objective Function

■ Multiple Reference

■ Multiple Scales

3 Challenging Data and Ground Truth

■ Overcoming Imperfections

4 Results & Conclusion

■ Qualitative

■ Quantitative

■ Ablations

■ Output Visuals

■ Conclusion

Outline



1 Introduction

■ Event-Based Cameras

- Event Optical Flow – Related Work
- Motivation and Contributions

2 Method

- The Problem
- Extracting Edges
- The Warp Model
- Bi-Modal Objective Function

■ Multiple Reference

■ Multiple Scales

3 Challenging Data and Ground Truth

- Overcoming Imperfections

4 Results & Conclusion

■ Qualitative

■ Quantitative

■ Ablations

■ Output Visuals

■ Conclusion

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



1 Introduction

- Event-Based Cameras
- Event Optical Flow – Related Work
- Motivation and Contributions

2 Method

- The Problem
- Extracting Edges
- The Warp Model
- Bi-Modal Objective Function

- Multiple Reference

- Multiple Scales

3 Challenging Data and Ground Truth

- Overcoming Imperfections

4 Results & Conclusion

- Qualitative

- Quantitative

- Ablations

- Output Visuals

- Conclusion

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



1 Introduction

- Event-Based Cameras
- Event Optical Flow – Related Work
- Motivation and Contributions

2 Method

- The Problem
- Extracting Edges
- The Warp Model
- Bi-Modal Objective Function

- Multiple Reference
- Multiple Scales

3 Challenging Data and Ground Truth

- Overcoming Imperfections

4 Results & Conclusion

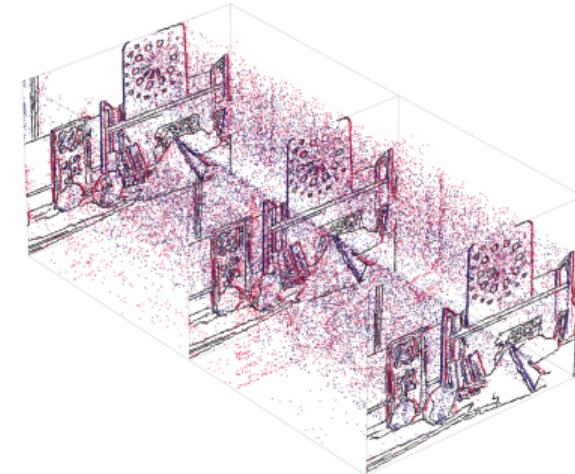
- Qualitative
- Quantitative
- Ablations
- Output Visuals
- Conclusion

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



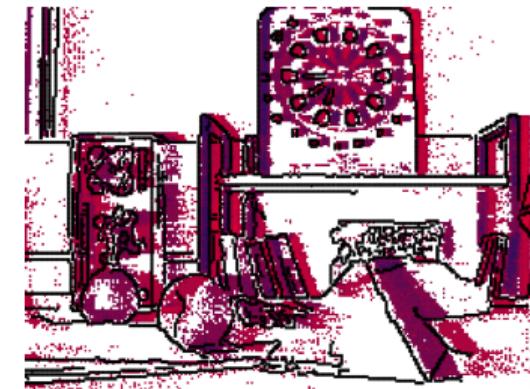
Motivation and Contributions



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."

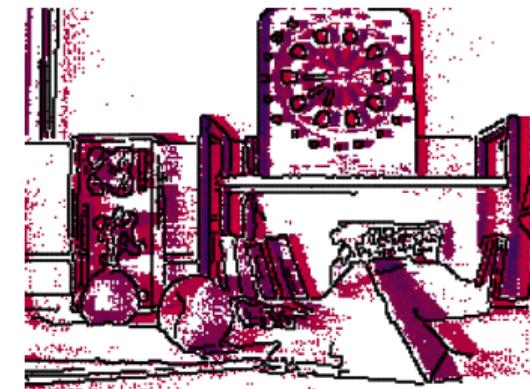
Motivation and Contributions



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



1 Introduction

- Event-Based Cameras
- Event Optical Flow – Related Work
- Motivation and Contributions

2 Method

■ The Problem

- Extracting Edges
- The Warp Model
- Bi-Modal Objective Function

- Multiple Reference
- Multiple Scales

3 Challenging Data and Ground Truth

- Overcoming Imperfections

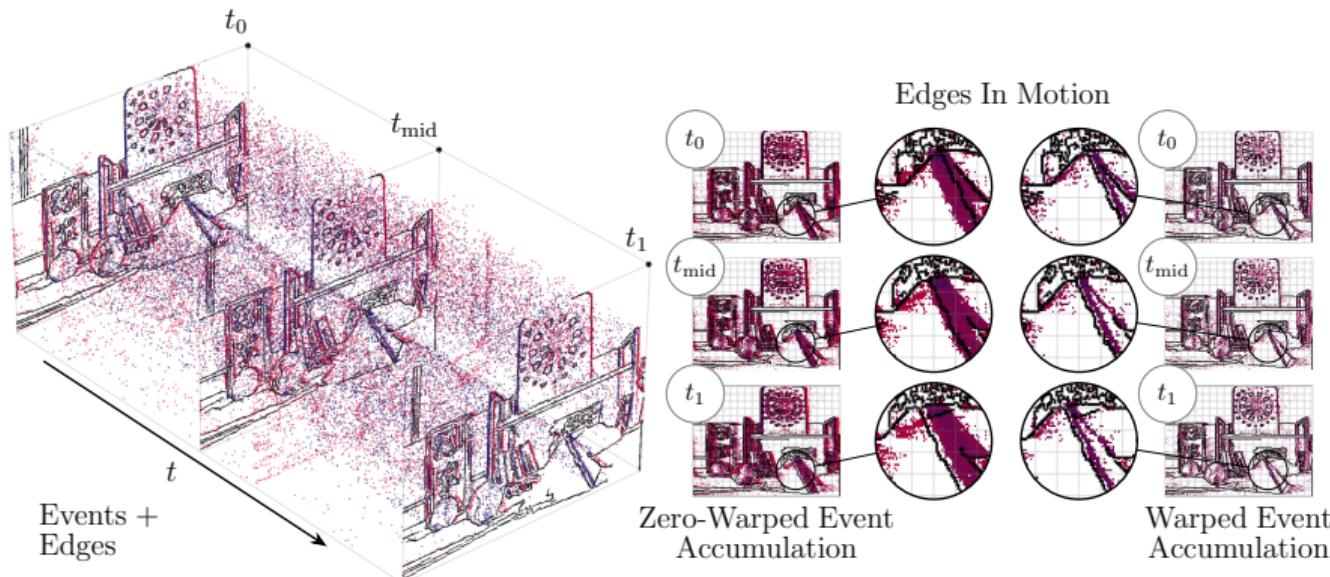
4 Results & Conclusion

- Qualitative
- Quantitative
- Ablations
- Output Visuals
- Conclusion

Problem Setting



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



Outline



1 Introduction

- Event-Based Cameras
- Event Optical Flow – Related Work
- Motivation and Contributions

2 Method

- The Problem
- **Extracting Edges**
- The Warp Model
- Bi-Modal Objective Function

- Multiple Reference
- Multiple Scales

3 Challenging Data and Ground Truth

- Overcoming Imperfections

4 Results & Conclusion

- Qualitative
- Quantitative
- Ablations
- Output Visuals
- Conclusion

Edge Extraction Pipeline



Original



Denoise



CLAHE



Sharpen



Filter



Canny



Blur



Outline



1 Introduction

- Event-Based Cameras
- Event Optical Flow – Related Work
- Motivation and Contributions

2 Method

- The Problem
- Extracting Edges
- **The Warp Model**
- Bi-Modal Objective Function

- Multiple Reference
- Multiple Scales

3 Challenging Data and Ground Truth

- Overcoming Imperfections

4 Results & Conclusion

- Qualitative
- Quantitative
- Ablations
- Output Visuals
- Conclusion



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

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

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

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

Outline



1 Introduction

- Event-Based Cameras
- Event Optical Flow – Related Work
- Motivation and Contributions

2 Method

- The Problem
- Extracting Edges
- The Warp Model
- Bi-Modal Objective Function

- Multiple Reference
- Multiple Scales

3 Challenging Data and Ground Truth

- Overcoming Imperfections

4 Results & Conclusion

- Qualitative
- Quantitative
- Ablations
- Output Visuals
- Conclusion



Contrast and Edge Correlation

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

Contrast

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

Correlation

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

Outline



1 Introduction

- Event-Based Cameras
- Event Optical Flow – Related Work
- Motivation and Contributions

2 Method

- The Problem
- Extracting Edges
- The Warp Model
- Bi-Modal Objective Function

■ Multiple Reference

- Multiple Scales

3 Challenging Data and Ground Truth

- Overcoming Imperfections

4 Results & Conclusion

- Qualitative
- Quantitative
- Ablations
- Output Visuals
- Conclusion



Multiple Frames = Multiple Reference Times

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

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

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

The motion parameters are optimized using contrast–correlation maximization

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

Outline



1 Introduction

- Event-Based Cameras
- Event Optical Flow – Related Work
- Motivation and Contributions

2 Method

- The Problem
- Extracting Edges
- The Warp Model
- Bi-Modal Objective Function

■ Multiple Reference

■ Multiple Scales

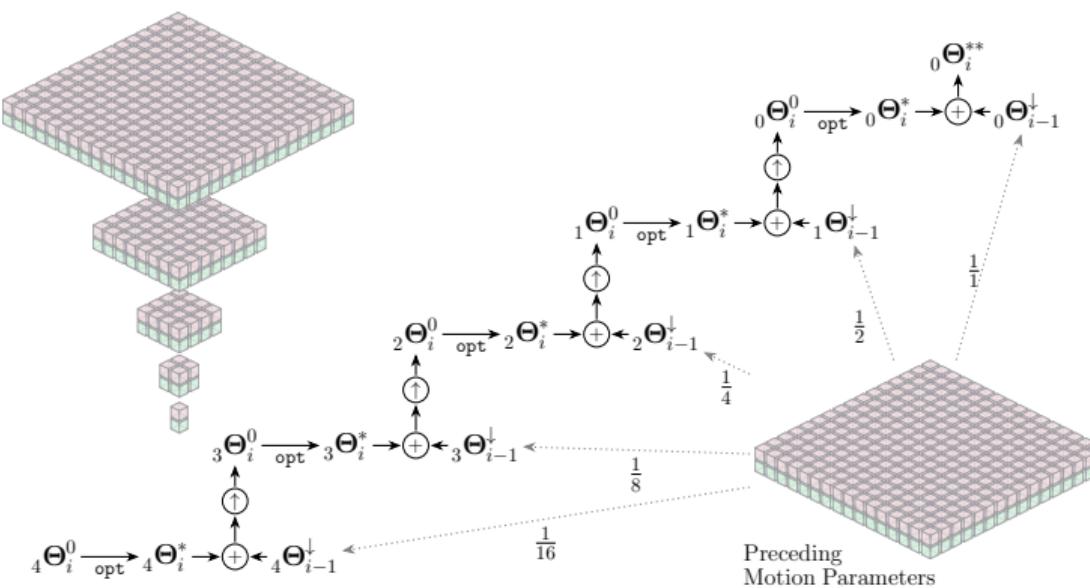
3 Challenging Data and Ground Truth

- Overcoming Imperfections

4 Results & Conclusion

- Qualitative
- Quantitative
- Ablations
- Output Visuals
- Conclusion

Multiscale and Sequential Processing



Outline



1 Introduction

- Event-Based Cameras
- Event Optical Flow – Related Work
- Motivation and Contributions

2 Method

- The Problem
- Extracting Edges
- The Warp Model
- Bi-Modal Objective Function

- Multiple Reference
- Multiple Scales

3 Challenging Data and Ground Truth

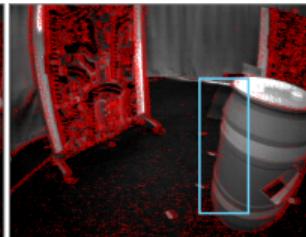
- Overcoming Imperfections

4 Results & Conclusion

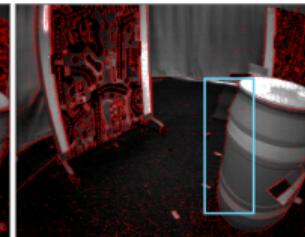
- Qualitative
- Quantitative
- Ablations
- Output Visuals
- Conclusion



Imperfect Ground Truth | Imperfect Data

(a) I_{events} 

(b) IWE (GT)

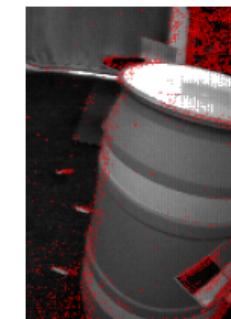


(c) IWE (Ours)

(d) I_{events} 

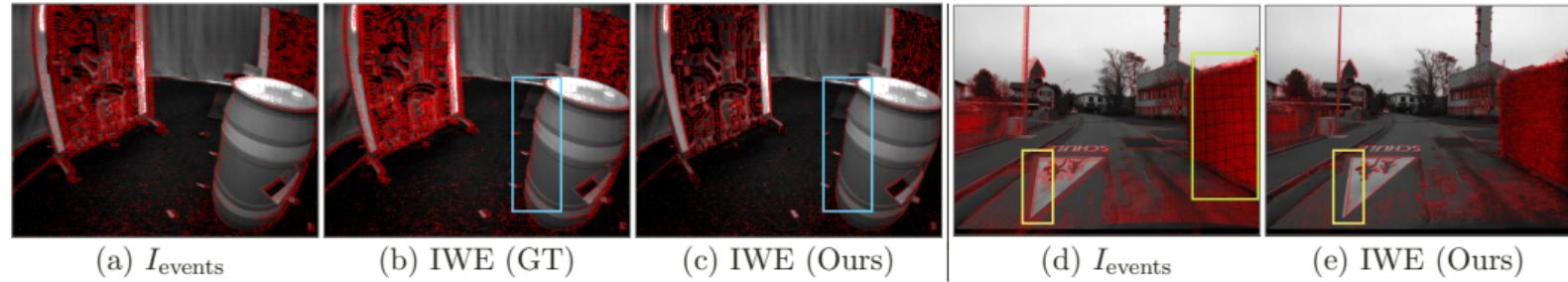
(e) IWE (Ours)

- Our method yields sharper warped events (□)

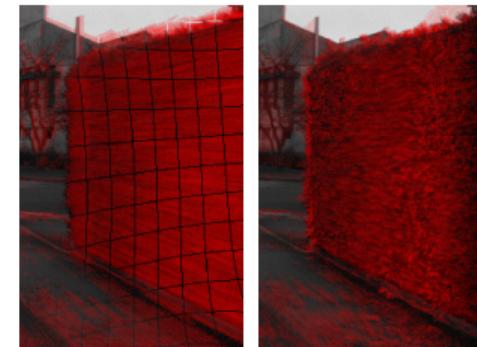




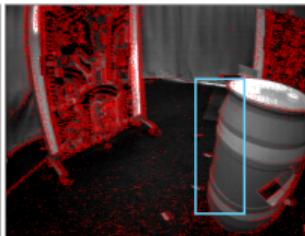
Imperfect Ground Truth | Imperfect Data



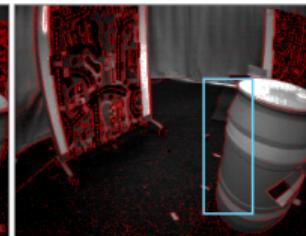
- Our method yields sharper warped events (□)
- Grid-like rectification artifacts (□) pose additional challenges



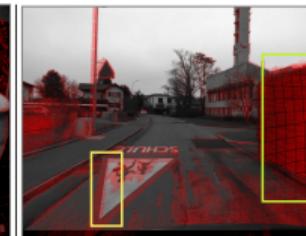
Imperfect Ground Truth | Imperfect Data

(a) I_{events} 

(b) IWE (GT)



(c) IWE (Ours)

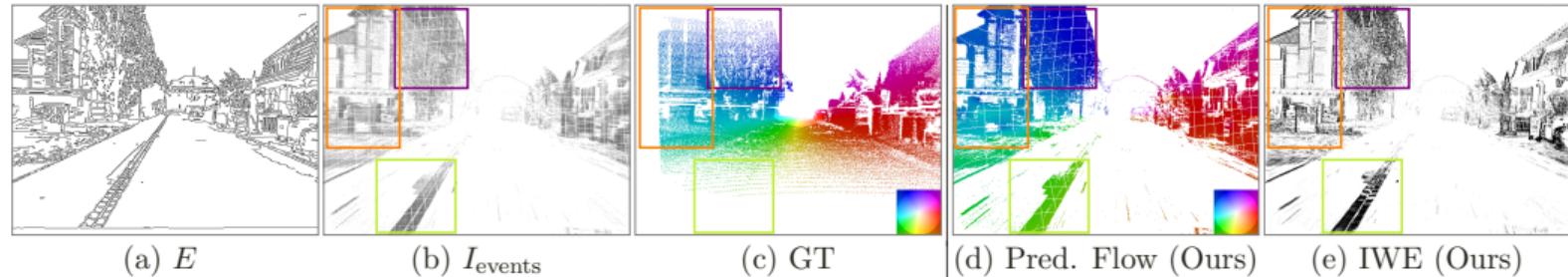
(d) I_{events} 

(e) IWE (Ours)

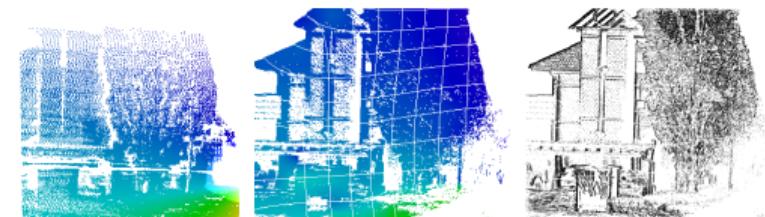
- Our method yields sharper warped events (□)
- Grid-like rectification artifacts (□) pose additional challenges
- Misalignment artifacts (e.g., road markings □) at points near the camera



Absence of Ground Truth at Dynamic Points in Scene



- Regions in the image, such as the top-left (\square , \square) and the bottom-center (\square) 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



1 Introduction

- Event-Based Cameras
- Event Optical Flow – Related Work
- Motivation and Contributions

2 Method

- The Problem
- Extracting Edges
- The Warp Model
- Bi-Modal Objective Function

- Multiple Reference
- Multiple Scales

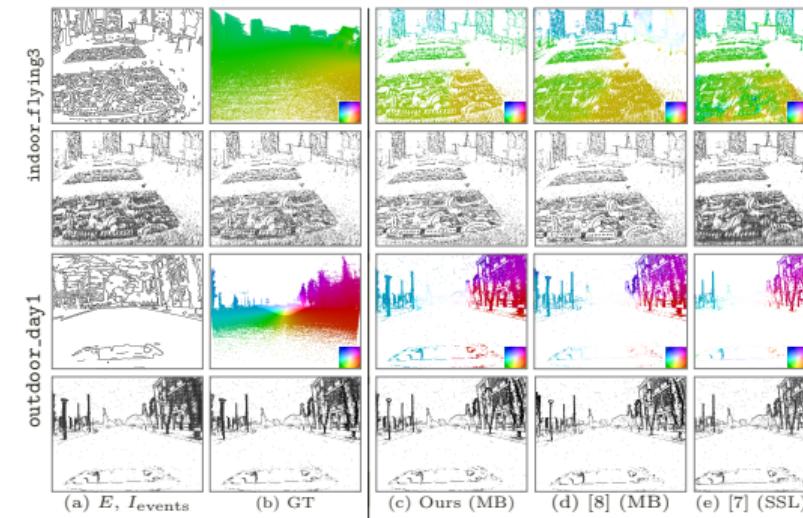
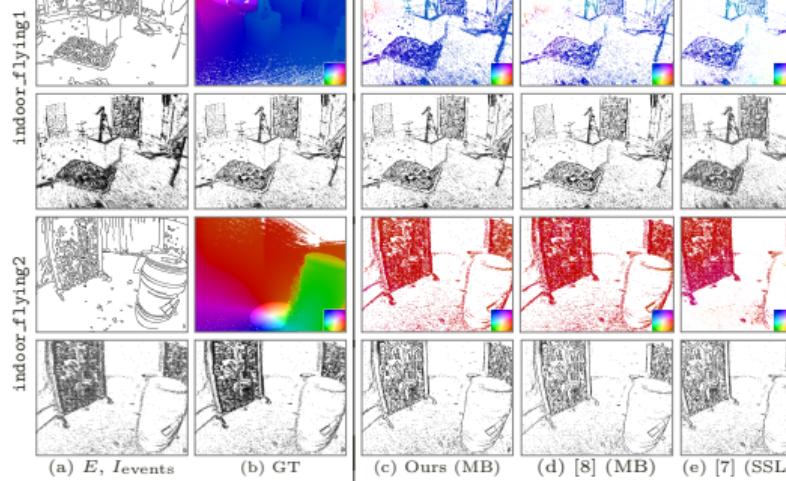
3 Challenging Data and Ground Truth

- Overcoming Imperfections

4 Results & Conclusion

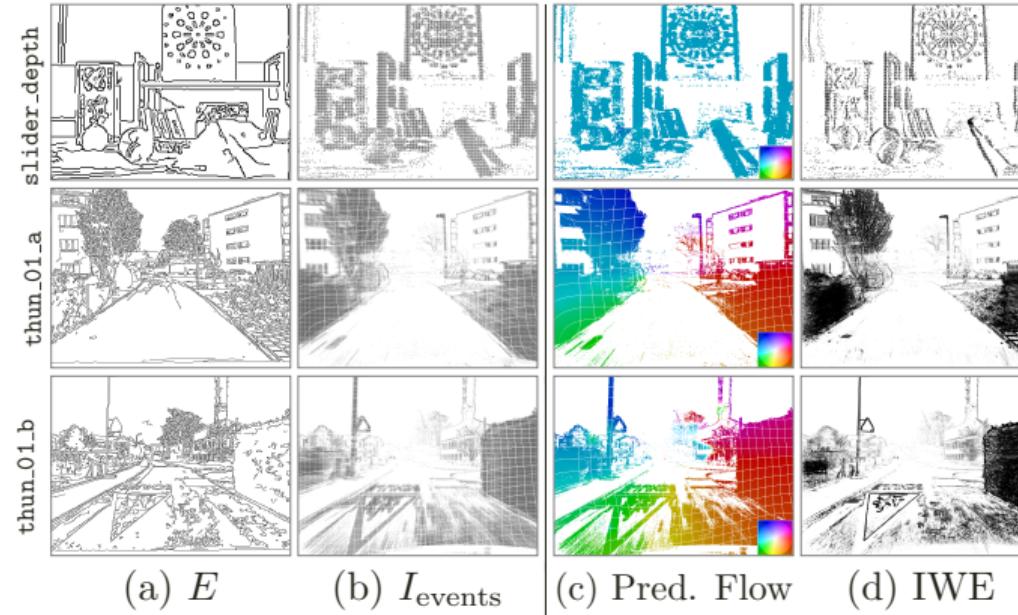
- Qualitative
- Quantitative
- Ablations
- Output Visuals
- Conclusion

MVSEC Qualitative Comparisons





ECD and DSEC Qualitative Results



Outline



1 Introduction

- Event-Based Cameras
- Event Optical Flow – Related Work
- Motivation and Contributions

2 Method

- The Problem
- Extracting Edges
- The Warp Model
- Bi-Modal Objective Function

- Multiple Reference
- Multiple Scales

3 Challenging Data and Ground Truth

- Overcoming Imperfections

4 Results & Conclusion

- Qualitative
- **Quantitative**
- Ablations
- Output Visuals
- Conclusion

Outline



1 Introduction

- Event-Based Cameras
- Event Optical Flow – Related Work
- Motivation and Contributions

2 Method

- The Problem
- Extracting Edges
- The Warp Model
- Bi-Modal Objective Function

- Multiple Reference
- Multiple Scales

3 Challenging Data and Ground Truth

- Overcoming Imperfections

4 Results & Conclusion

- Qualitative
- Quantitative
- Ablations
- Output Visuals
- Conclusion



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): $w_{ho} = 0.5$ for all pyramid levels
- 3 Fixed+Solved handover (FSHO): $w_{ho} = 0.5$ 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 ↑						
Ours (EINCM-SHO)	1.51 ₃	1.70 ₂	1.30 ₃	1.32 ₉	0.67 ₄	<u>1.52</u> ₆	1.47 ₃
Ours (EINCM-FHO)	<u>1.65</u> ₉	<u>1.74</u> ₆	<u>1.32</u> ₆	<u>1.36</u> ₂	<u>1.15</u> ₉	1.38 ₆	<u>1.53</u> ₇
Ours (EINCM-FSHO)	1.75 ₈	1.75 ₅	1.45 ₈	1.40 ₅	1.34 ₃	1.54 ₂	1.61 ₃

Outline



1 Introduction

- Event-Based Cameras
- Event Optical Flow – Related Work
- Motivation and Contributions

2 Method

- The Problem
- Extracting Edges
- The Warp Model
- Bi-Modal Objective Function

- Multiple Reference
- Multiple Scales

3 Challenging Data and Ground Truth

- Overcoming Imperfections

4 Results & Conclusion

- Qualitative
- Quantitative
- Ablations
- **Output Visuals**
- Conclusion

Example Outcome for DSEC sequence interlaken_00_b



Outline



1 Introduction

- Event-Based Cameras
- Event Optical Flow – Related Work
- Motivation and Contributions

2 Method

- The Problem
- Extracting Edges
- The Warp Model
- Bi-Modal Objective Function

- Multiple Reference

- Multiple Scales

3 Challenging Data and Ground Truth

- Overcoming Imperfections

4 Results & Conclusion

- Qualitative

- Quantitative

- Ablations

- Output Visuals

- Conclusion

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

