Perturbed State Space Feature Encoders for Optical Flow with Event Cameras

University of Zurich[UZH]

Gokul Raju Govinda Raju*, Nikola Zubić*, Marco Cannici, Davide Scaramuzza

ROBOTICS & PERCEPTION GROUP

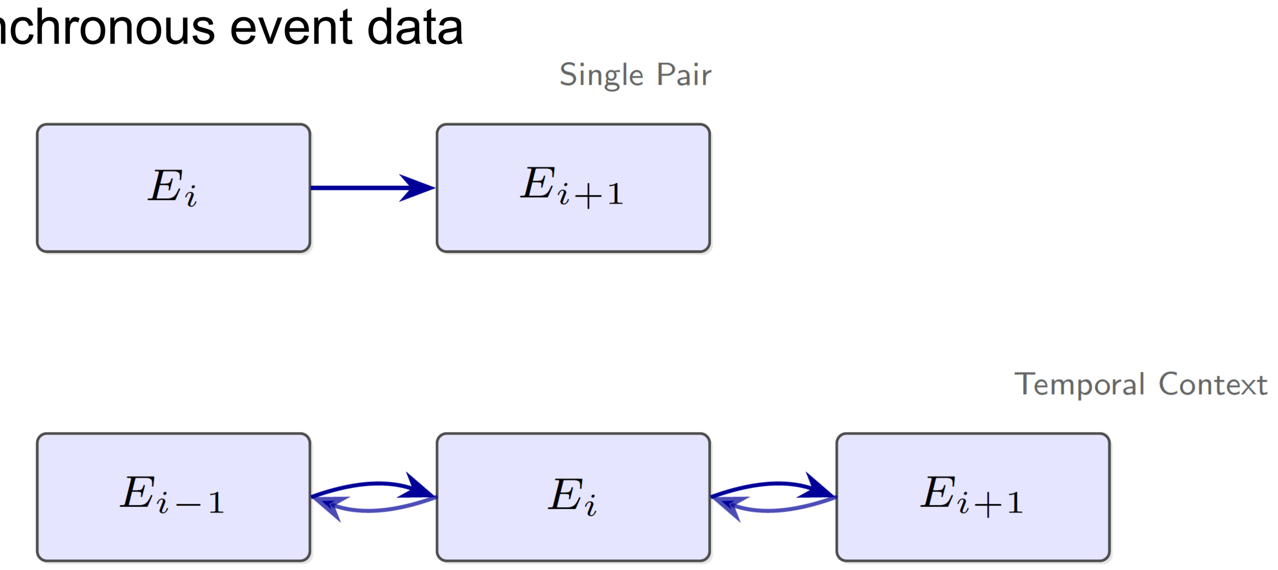
Motivation: **Event cameras offer rich motion data but current optical flow methods have temporal and spatial limits. We introduce Perturbed State Space Feature Encoders (P-SSE) with a novel perturbation technique for improved stability and performance. P-SSE integrates into a multi-frame, bi-directional framework, achieving SOTA on DSEC-Flow and MVSEC datasets for event-based optical flow.**

Goal: **Our aim is Robust & Accurate Multi-Frame Event Optical Flow. To achieve this, we do the following:**

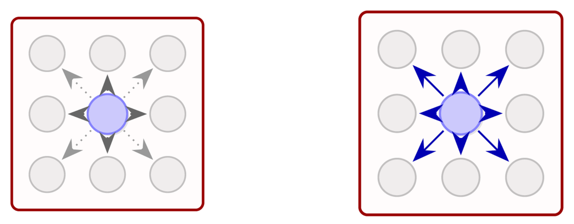
- **Leverage multi-frame bi-directional flow for richer temporal context**
- **Develop a novel encoder (P-SSE) with a global receptive field and linear complexity**
- **Ensure model stability and performance through a unique perturbation technique**

Limitations in Event-Based Optical Flow

- **Temporal Limitations:** Pair-wise designs ignore broader temporal context
- **Spatial Limitations:** Encoders (e.g., in VideoFlow) have restricted receptive fields, constraining global dependency capture
- **SSM Instability:** Standard SSMs can struggle with noisy, asynchronous event data



Pairwise vs. Multi-frame



Receptive Field Comparison

Check out our Paper, Research and more!

Paper link

Vision research at RPG

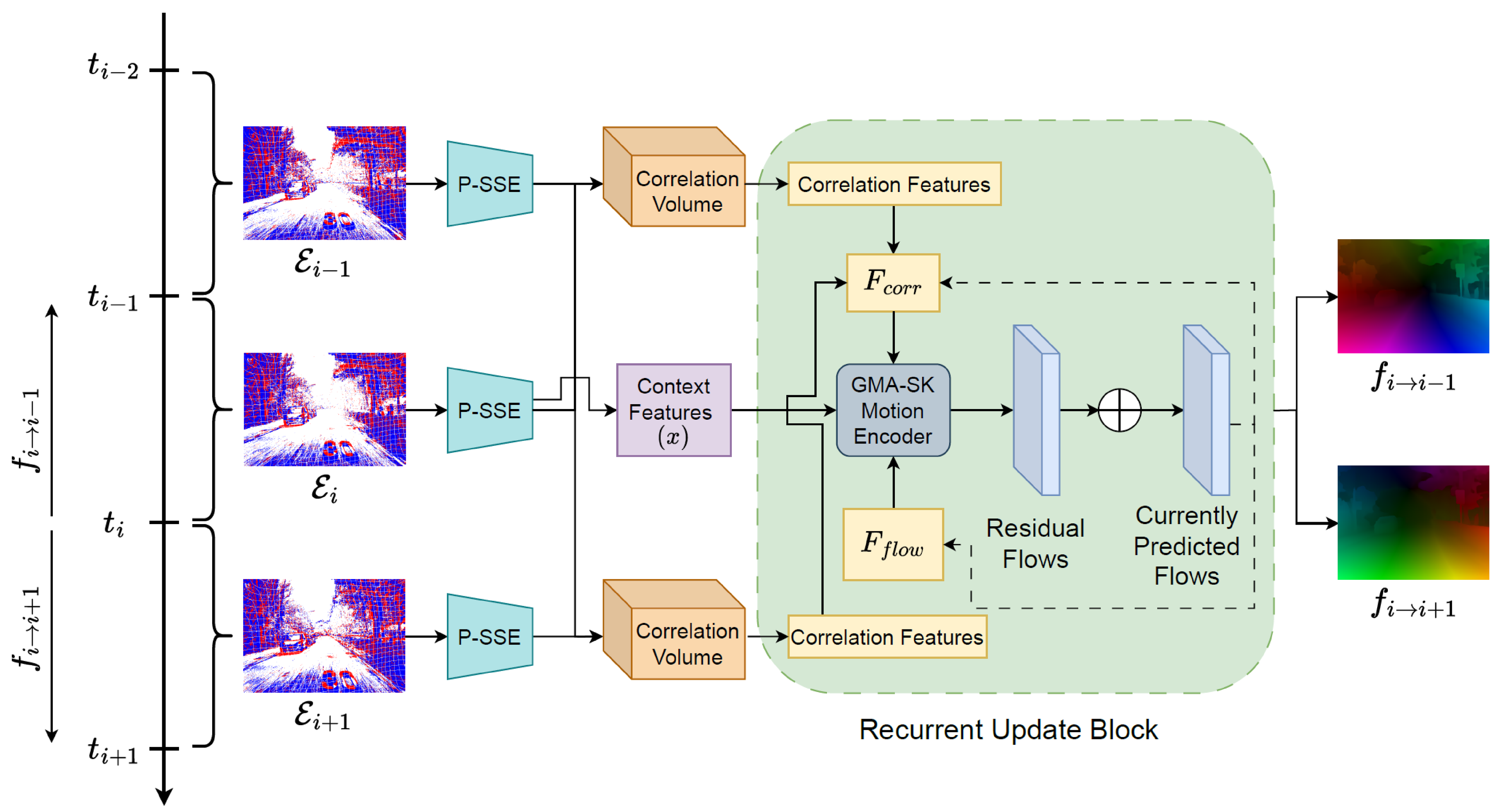
ML research at RPG

Nikola Zubić's homepage

Sponsors: Horizon2020 European Union Funding for Research & Innovation; erc

P-SSE Framework:

- **Foundation**: Builds on State Space Models (SSMs) for linear complexity and Transformers for global receptive fields
- **Key Idea:** Perturbation Technique:
  - Addresses SSM instability with complex event data
  - Applies a carefully designed perturbation (PTD: Perturb-then-Diagonalize) to the state dynamics matrix (A)
  - A* = A + E (small perturbation E, ~10% of A's magnitude)
  - Improves stability, robustness to noise, and overall performance
- **Scanning:** Utilizes VMamba-like 2D scanning to process feature maps



A simplified block diagram of the P-SSE encoder for tri-frame (E-TROF) case. Perturbed "A" matrix is used and the P-SSE is pretrained on ImageNet, then used in the following architecture

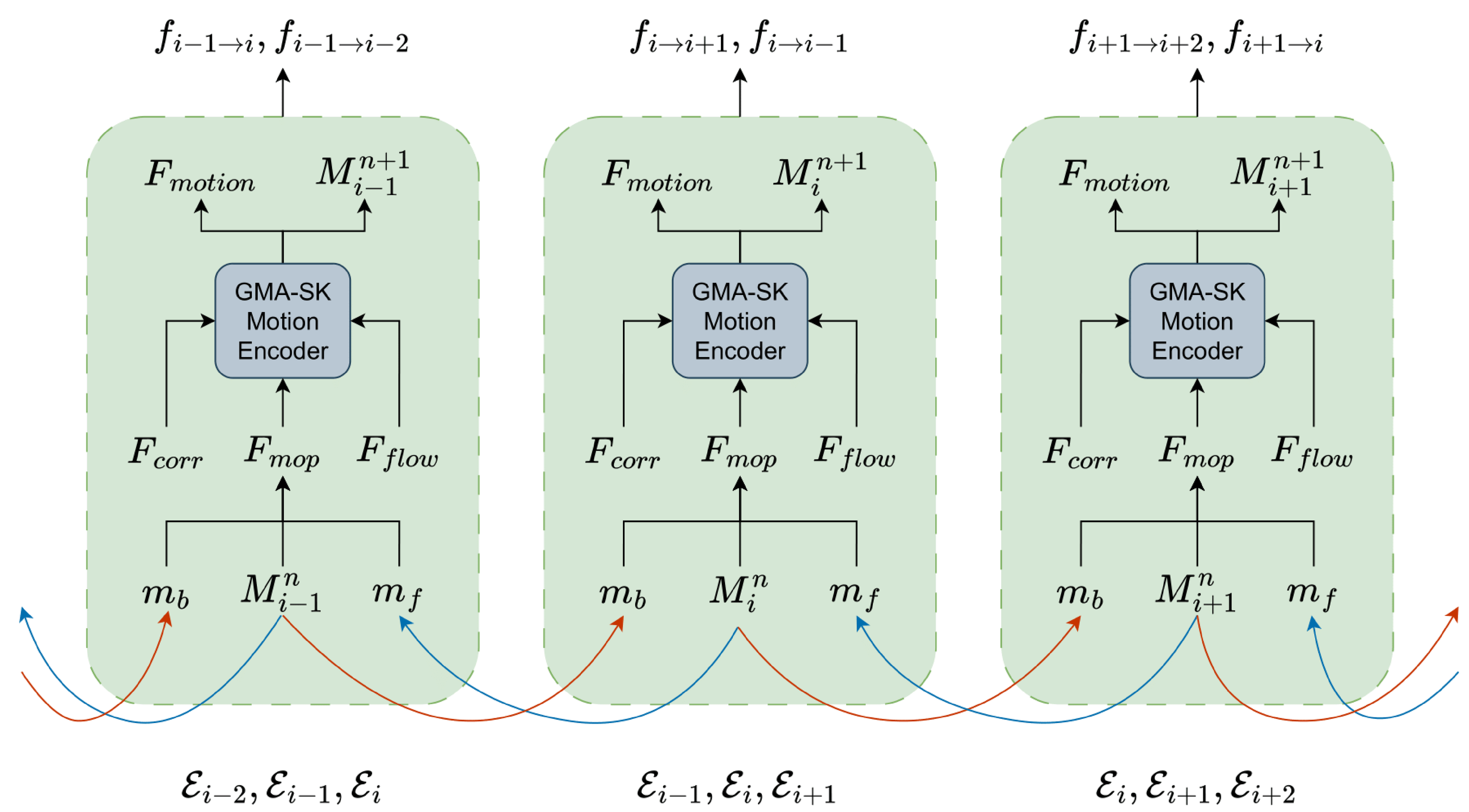


Diagram of the E-MOP model, which integrates 3 E-TROFs for 5 consecutive event representations to predict and refine bidirectional optical flows by sharing dynamic temporal motion information among adjacent TROFs. Integrating P-SSE for Multi-Event Optical Flow. E-TROF processes triplets of event representations. E-MOP extends temporal context across 5 consecutive event representations and propagates motion state information between adjacent E-TROF modules. This enables iterative refinement and richer temporal reasoning.

Experiments & Results

Superior Performance on Benchmarks

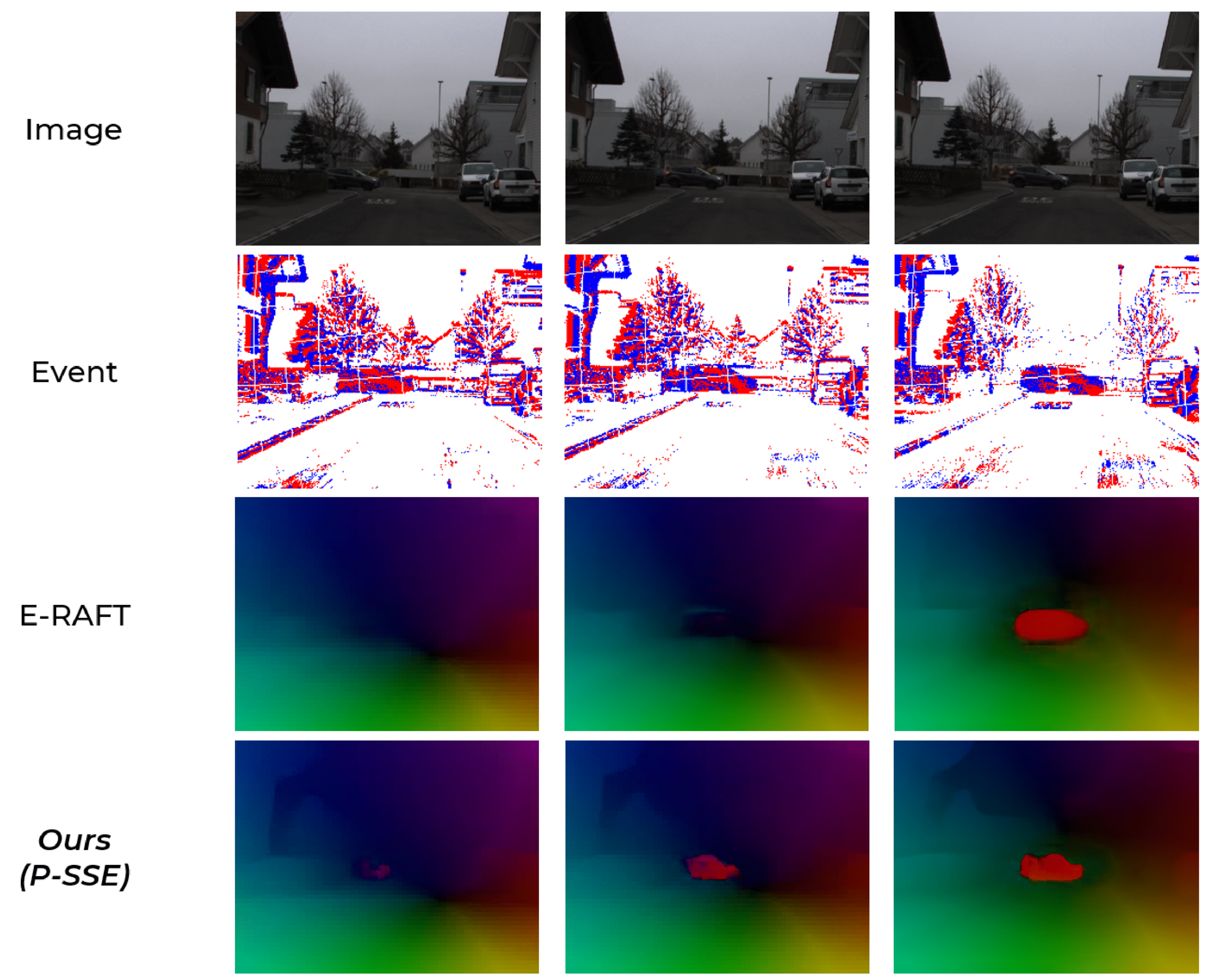
- **Datasets:** DSEC-Flow (train & test), MVSEC (zero-shot test)
- **Event Representation:** ERGO-12
- **Metrics:** End-Point Error (EPE), Angular Error (AE), N-Pixel Error (NPE)

Method	1PE	2PE	3PE	EPE	AE
E-RAFT	12.742	4.740	2.684	0.788	2.851
ADMFlow	12.522	4.673	2.647	0.779	2.838
E-FlowFormer	11.225	4.102	2.446	0.759	2.676
BFlow	11.901	4.411	2.440	0.750	2.680
TMA	10.863	3.972	2.301	0.743	2.684
IDNet	10.111	3.523	2.018	0.723	2.724
ECDDP	**8.887**	**3.199**	1.958	0.697	2.575
Ours (P-SSE)	**9.144**	**3.232**	**1.816**	**0.680**	**2.560**

P-SSE significantly outperforms prior methods on DSEC-Flow, achieving a top EPE of 0.680

Method	Avg. EPE
Stoffregen et al.	0.60
Li et al.	0.62
STE-FlowNet	0.69
DCEIFlow	0.59
Ours (P-SSE, ERGO-12, Base)	**0.52**

P-SSE (0.52 EPE) shows an 11.86% improvement over previous best (0.59 EPE) on MVSEC (zero-shot)



Demonstration of P-SSE's capability in managing partially occluded scenes, compared with E-RAFT. The sequence progresses from left to right, showcasing frames from a DSEC-Flow test sequence.