Optical Flow for Spike Camera with Hierarchical Spatial-Temporal Spike Fusion

Rui Zhao[1] Ruiqin Xiong[1]* Jian Zhang[1] Xinfeng Zhang[2] Zhaofei Yu[1] Tiejun Huang[1]

[1]Peking University [2]University of Chinese Academy of Sciences

AAAI 2024

北京大学 PEKING UNIVERSITY

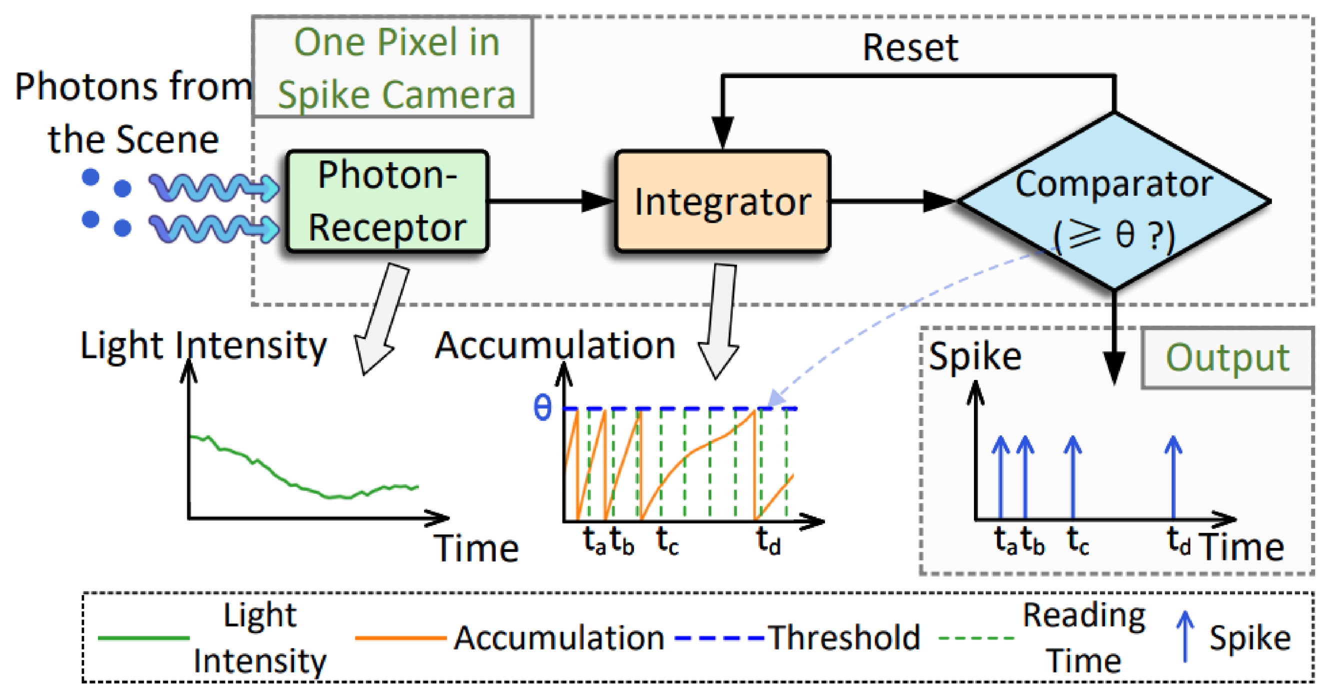
中国科学院大学 University of Chinese Academy of Sciences

NERC National Engineering Research Center of Visual Technology

Personal Homepage

1. Introduction

1.1 Spike Camera. Spike cameras are composed of an array of pixels working asynchronously. Each pixel of a spike camera is composed of three main components: photon-receptor, integrator, and comparator. The integrator accumulates the photoelectrons from the photon-receptor and transfers them to the voltage. The comparator compares the accumulation with the threshold continuously. Once the voltage of the integrator exceeds a certain threshold, the camera fires a spike and resets the accumulation.

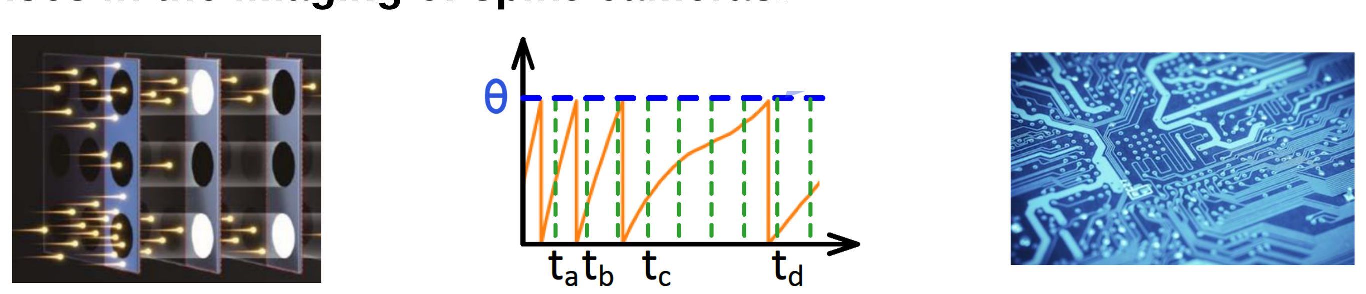


Key components of a pixel in spike camera

$$A(\mathbf{x}, t) = \int_0^t \alpha \cdot I(\mathbf{x}, \tau) \mathrm{d}\tau \mod \theta$$

1.2 Challenges of Spike-Based Optical Flow.

Noises in the imaging of spike cameras.



(a) Poisson Effect of Photons' Arrival

(b) Quantitative Effect from Spike Reading

(c) Thermal Noises in the Circuits

Fluctuations and Randomness in Spikes

⇩

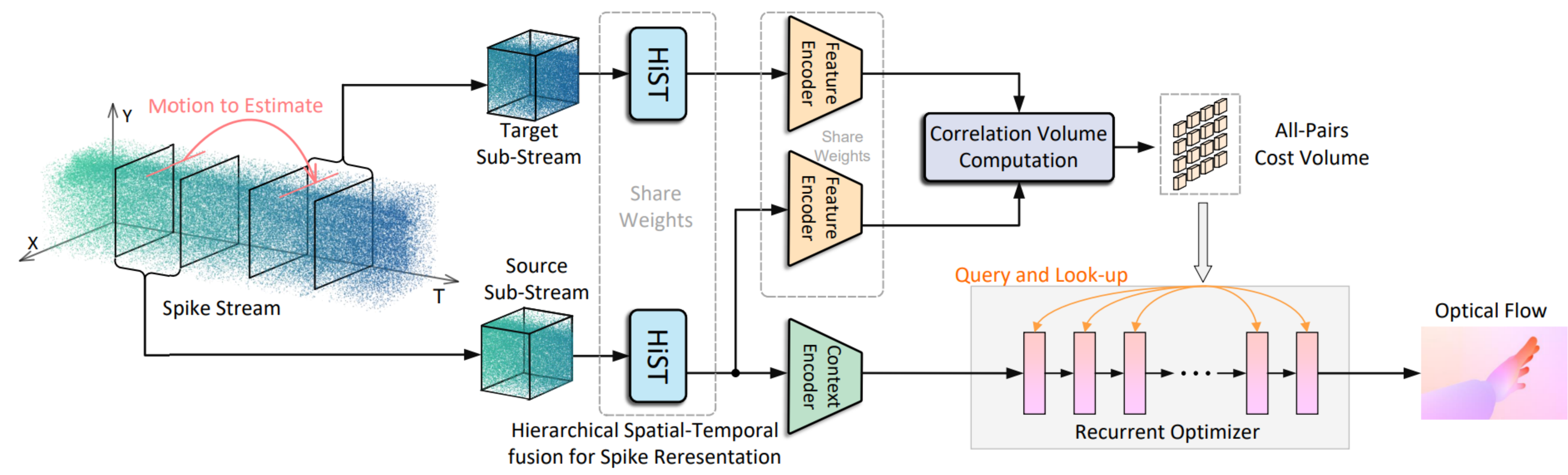
Ambiguities in correlation → Inaccurate feature matching

2. Contributions

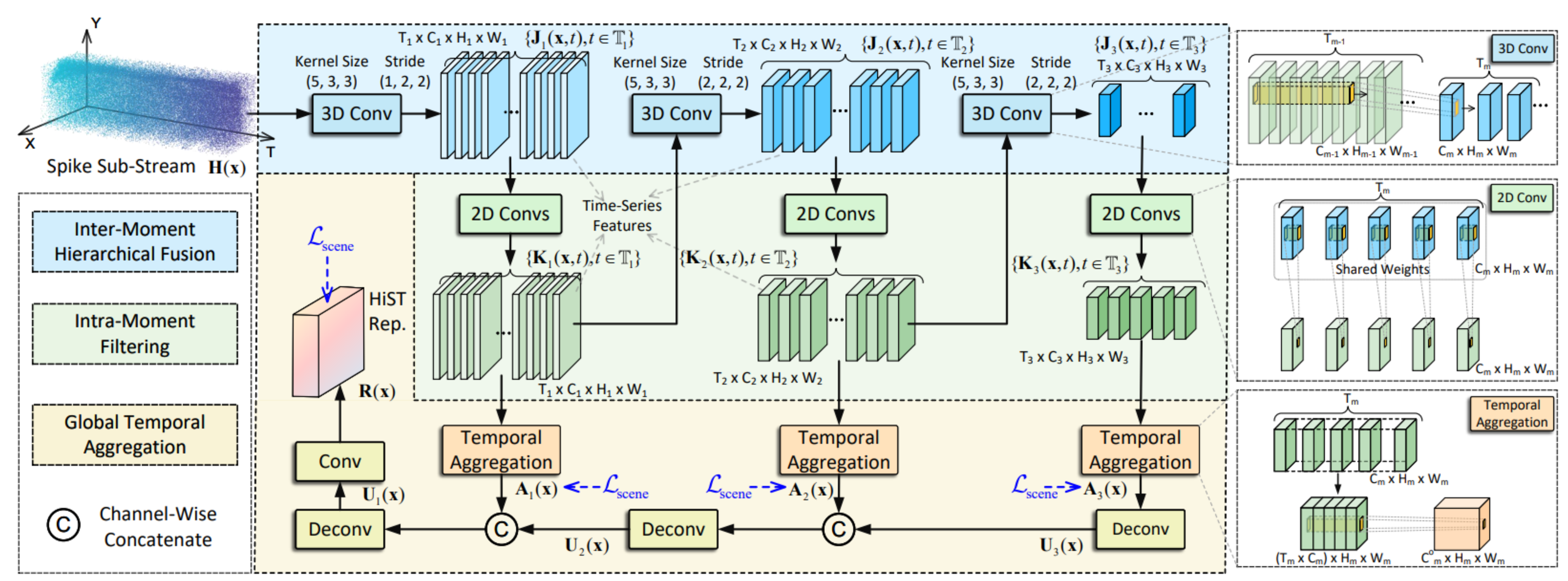
- A HiST-SFlow is proposed for spike-based optical flow. In HiST-SFlow, the spikes are represented by the HiST module and extracted to features for correlation. The optical flow is estimated by a recurrent optimizer.
- An inter-moment hierarchical fusion (InterF) module and an intra-moment filtering (IntraF) module are proposed to suppress the randomness in the spikes. A scene loss is proposed to constrain high-fidelity representation to contain the brightness information of the scene.

3. Approaches

3.1 Overall Architecture of HiST-Sflow.



3.2 Hierarchical Spatial-Temporal (HiST) Fusion.



(a) Inter-Moment Hierarchical Fusion (InterF).

- Fuse features at different moments while retaining the time information in features.

$$\mathbf{J}_m(\mathbf{x}, t) = \mathscr{J}_m \left[\{ \mathbf{K}_{m-1}(\mathbf{x}, \tau) \mid \tau \in \mathbb{T}_{m-1} \} \right]$$

$$\mathbb{T}_{m-1} = \{ T_c - T_{m-1}^{\text{half}}, \dots, T_c, \dots T_c + T_{m-1}^{\text{half}} \}$$

(b) Intra-Moment Filtering (IntraF).

- Reduce the influence of spikes' fluctuations for each moment through the feature at the current moment.
- The InterF and IntraF are implemented alternatively in each level of the pyramid.

$$\mathbf{K}_m(\mathbf{x}, t) = \mathscr{K}_m \left[\mathbf{J}_m(\mathbf{x}, t) \right], \ t \in \mathbb{T}_m$$

(c) Global Temporal Aggregation (GTA).

- Fuse features of all the moments at each level of the pyramid to represent the central moment of input spike sub-stream.

$$\mathbf{A}_m(\mathbf{x}) = \mathscr{A}_m \left[\mathrm{Cat} \left\{ \mathbf{K}_m(\mathbf{x}, \tau) \mid \tau \in \mathbb{T}_m \right\} \right]$$

(d) Scene Loss.

- Ensure the spike representation contain the scene's brightness information
- The $\{\mathscr{P}_m\}_{m=0}^{3}$ are 3-layer convolution layers, which are used only during training and not in inference.

$$\mathcal{L}_{\text{scene}} = \|\mathbf{I}_{\text{scene}}(\mathbf{x}, T_c) - \mathscr{P}_0(\mathbf{R}_{T_c}(\mathbf{x}))\|_1 + \sum_{m=1}^{3} \lambda_m \|\sigma_m(\mathbf{I}_{\text{scene}}(\mathbf{x}, T_c)) - \mathscr{P}_m(\mathbf{A}_m(\mathbf{x}))\|_1$$

3.3 Loss Function.

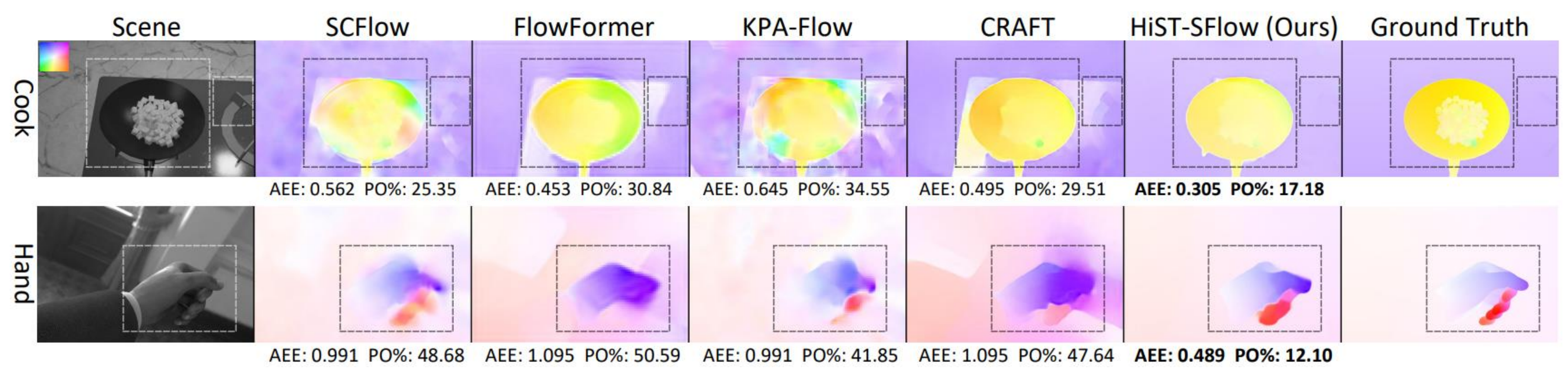
$$\mathcal{L}_{\text{flow}} = \sum_{i=1}^{N} \gamma^{N-i} \|\mathbf{w}_i(\mathbf{x}) - \mathbf{w}_{\text{gt}}(\mathbf{x})\|_1 \qquad \mathcal{L} = \mathcal{L}_{\text{flow}} + \lambda(\mathcal{L}_{\text{scene}}^{\text{src}} + \mathcal{L}_{\text{scene}}^{\text{tgt}})$$

4. Experimental Results

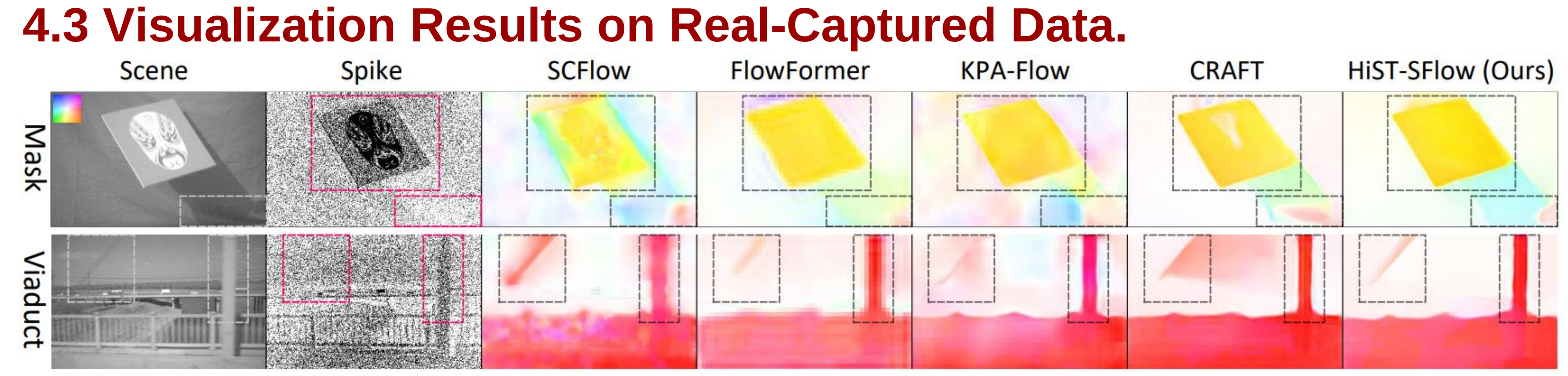
4.1 Quantitative Results on PHM Dataset (AEPE / PO%).

	Architecture	Ball	Cook	Dice	Doll	Fan	Hand	Jump	Poker	Top	**Average**
$\Delta t = 10$	SCFlow	0.51 / 20.3	1.34 / 38.6	1.10 / 30.7	0.22 / 5.6	**0.24** / 10.7	1.30 / 57.3	0.11 / 3.0	0.80 / 41.1	2.14 / 17.7	0.863 / 25.00
	RAFT	0.46 / 12.5	1.32 / 43.7	0.95 / 29.3	0.24 / 6.7	0.28 / 12.7	1.11 / 45.1	0.11 / 3.0	0.67 / 37.1	2.19 / 19.7	0.813 / 23.30
	GMA	0.61 / 21.7	1.84 / 74.7	1.13 / 34.2	0.39 / 9.4	0.36 / 12.1	2.13 / 80.6	0.17 / 2.8	0.88 / 43.5	2.29 / 23.6	1.087 / 33.63
	Flow1D	0.79 / 51.4	1.28 / 50.8	1.15 / 47.9	0.27 / 6.3	0.28 / 11.0	1.86 / 83.1	0.13 / 3.4	0.85 / 50.1	2.19 / 17.7	0.979 / 35.76
	KPA-Flow	0.47 / 14.9	1.41 / 45.9	0.87 / 29.9	0.27 / 7.1	0.29 / 12.7	1.19 / 47.7	0.12 / 3.0	0.65 / 36.6	2.19 / 19.4	0.827 / 24.12
	GMFlow	0.76 / 42.4	1.29 / 61.0	1.54 / 81.7	0.31 / 8.4	0.43 / 14.1	1.83 / 65.0	0.30 / 3.7	0.95 / 54.2	2.29 / 23.3	1.077 / 39.33
	GMFlowNet	0.45 / 12.1	1.22 / 43.8	1.02 / 32.9	0.35 / 7.8	0.25 / 10.7	1.53 / 65.3	0.12 / 3.2	0.65 / 31.5	2.18 / 17.5	0.863 / 24.98
	CRAFT	0.61 / 15.0	1.28 / 43.5	0.93 / 27.6	**0.19** / **5.0**	0.25 / **10.2**	1.67 / 73.3	0.10 / 2.6	0.56 / **23.1**	2.15 / 15.1	0.860 / 23.94
	FlowFormer	0.52 / 13.5	1.48 / 58.7	0.98 / 31.0	0.25 / 6.7	0.29 / 11.5	1.82 / 84.5	0.14 / 3.6	0.94 / 54.9	2.22 / 19.5	0.959 / 31.54
	HiST-SFlow	**0.28 / 7.8**	**0.80 / 27.4**	**0.85 / 23.3**	0.20 / 5.6	0.27 / 12.8	**0.64 / 21.7**	**0.08 / 2.5**	**0.53** / 23.9	**2.11 / 14.8**	**0.640 / 15.54**
$\Delta t = 20$	SCFlow	0.94 / 27.1	3.00 / 50.6	1.72 / 33.2	0.41 / 8.1	**0.46 / 13.6**	3.71 / 71.3	0.19 / 5.9	1.57 / 53.7	4.25 / 18.9	1.804 / 31.37
	RAFT	0.78 / 18.6	2.75 / 54.4	1.57 / 30.1	0.43 / 9.3	0.50 / 14.6	2.81 / 59.9	0.21 / 5.8	1.31 / 46.7	4.30 / 21.2	1.628 / 28.94
	GMA	1.01 / 22.1	4.95 / 96.4	1.52 / 35.9	1.00 / 59.6	1.19 / 98.4	6.66 / 99.5	0.81 / 84.4	1.39 / 45.2	4.64 / 64.9	2.575 / 67.38
	Flow1D	1.19 / 51.6	4.52 / 96.3	1.58 / 50.7	0.78 / 53.3	1.01 / 82.1	6.65 / 99.2	0.72 / 73.1	1.39 / 52.3	4.75 / 79.7	2.510 / 70.90
	KPA-Flow	0.80 / 20.9	2.93 / 55.6	1.48 / 31.4	0.45 / 9.6	0.52 / 14.5	2.86 / 62.5	0.22 / 5.6	1.31 / 48.4	4.28 / 19.7	1.649 / 29.81
	GMFlow	1.49 / 80.3	2.64 / 80.1	2.72 / 91.8	0.54 / 15.3	0.77 / 22.0	3.79 / 81.5	0.55 / 27.8	1.78 / 75.3	4.45 / 32.5	2.080 / 56.28
	GMFlowNet	0.92 / 31.4	2.61 / 70.4	2.17 / 42.7	0.61 / 27.5	0.56 / 13.9	3.30 / 93.2	0.21 / 4.5	1.33 / 53.4	4.33 / 25.3	1.782 / 40.25
	CRAFT	1.16 / 85.5	2.68 / 61.0	1.99 / 46.8	0.39 / 7.8	0.48 / 12.5	3.53 / 87.1	0.20 / **3.6**	**1.23** / 38.9	4.31 / 22.0	1.775 / 40.57
	FlowFormer	0.91 / 13.8	4.41 / 96.3	**1.40** / 32.6	0.80 / 54.8	1.03 / 90.0	6.54 / 99.3	0.74 / 75.8	1.47 / 57.4	4.59 / 61.9	2.432 / 64.67
	HiST-SFlow	**0.55 / 8.8**	**2.04 / 33.6**	1.64 / **26.3**	**0.38 / 7.2**	0.51 / 13.9	**2.00 / 34.7**	**0.17** / 5.0	1.28 / **33.1**	**4.18 / 15.1**	**1.417 / 19.73**

4.2 Visualization Results on PHM Dataset.



4.3 Visualization Results on Real-Captured Data.



4.4 Ablation Studies.

Ablations on Proposed Modules

Index	Settings			$\Delta t = 10$		$\Delta t = 20$	
	InterF	IntraF	$\mathcal{L}_{\text{scene}}$	AEPE	PO%	AEPE	PO%
(A)	✗	✗	✗	0.986	33.17	2.095	56.56
(B)	✓	✗	✗	0.694	18.18	1.449	21.99
(C)	✓	✓	✗	0.676	17.34	1.433	22.79
(D)	✓	✗	✓	0.675	16.63	1.448	21.40
(E)	✓	✓	✓	**0.640**	**15.54**	**1.417**	**19.73**

Ablations on Different Representations

Representation	$\Delta t = 10$		$\Delta t = 20$	
	AEPE	PO%	AEPE	PO%
Window-Based	0.868	25.72	1.757	34.19
Interval-Based	0.880	29.77	1.824	37.91
Multi-Window	0.799	21.10	1.703	34.58
Flow-Guided Window	0.696	16.99	1.533	23.36
HiST (Ours)	**0.640**	**15.54**	**1.417**	**19.73**

4.5 Using HiST for Other Baselines.

Architecture	with HiST	$\Delta t = 10$		$\Delta t = 20$	
		AEPE	PO%	AEPE	PO%
GMA	No	1.087	33.63	2.575	67.38
	Yes	**0.666**	**16.91**	**1.391**	**21.20**
KPA-Flow	No	0.827	24.12	1.649	29.81
	Yes	**0.659**	**16.99**	**1.363**	**22.27**
GMFlowNet	No	0.863	24.98	1.782	40.25
	Yes	**0.730**	**21.22**	**1.452**	**24.93**