Self-Supervised Learning of Inlier Events for Event-based Optical Flow

Aoki Media Sensing Lab.
Keio University, Faculty of Science of Technology
DENSO IT LAB

Jun Nagata[1,2], Yoshimitsu Aoki[1] (1. Keio University, 2. Denso IT Laboratory)

BMVC 2022

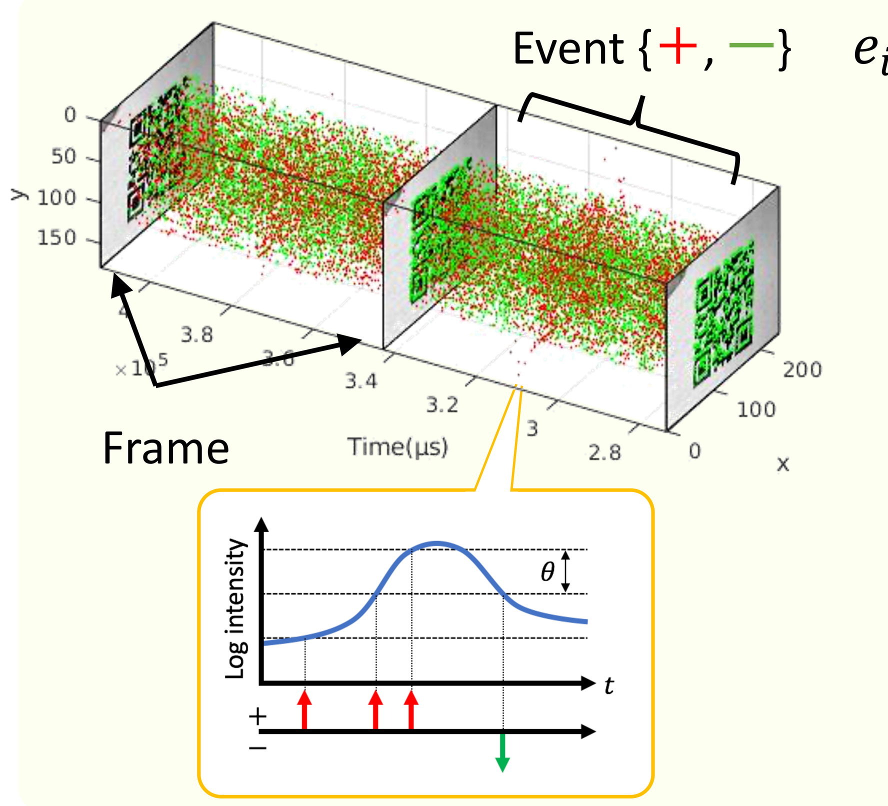
1. Introduction

The task is fast motion computing with event-based cameras.

☐ Event camera

Event-based cameras are bio-inspired sensors that asynchronously report per-pixel intensity changes at each pixel.

They are suitable devices for motion estimation because of their low-latency sensing mechanism.

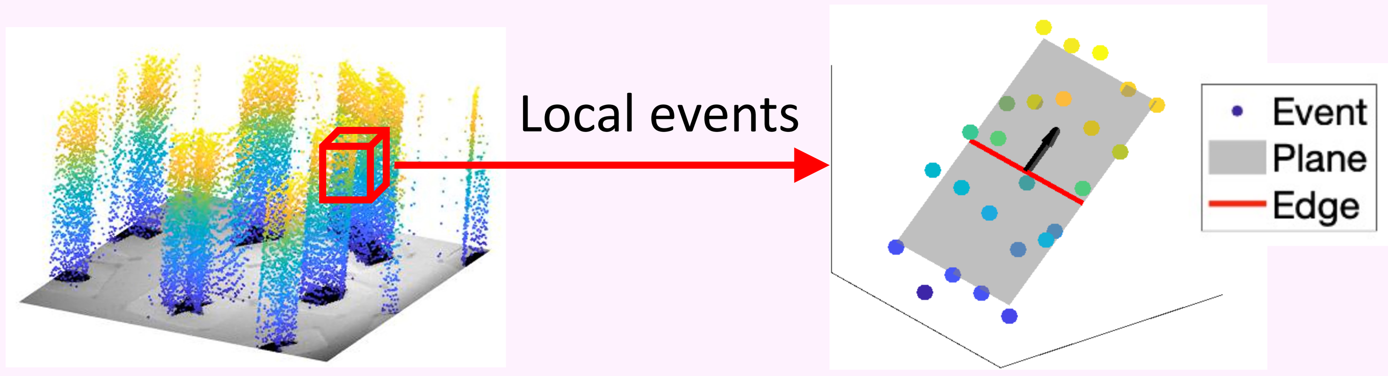


$e_i \doteq (\mathbf{x}_i, t_i, \rho_i)$

$\mathbf{x}_i$: pixel location
t_i : timestamp
ρ_i : polarity {+1, -1}

- No motion blur
- High dynamic range
- Low latency
- Low power consumption

☐ Normal flow by planar fitting



Local motion can be computed by planar fitting.

However, least-squares fitting suffers from heavy event noise.

☐ Related work

1. Iterative outlier rejection [1]
   ☺ Exploring a plane ☹ iterative
2. Greedy selection of the neighbor largest timestamp [2]
   ☺ Non iterative ☹ Not exploring a plane

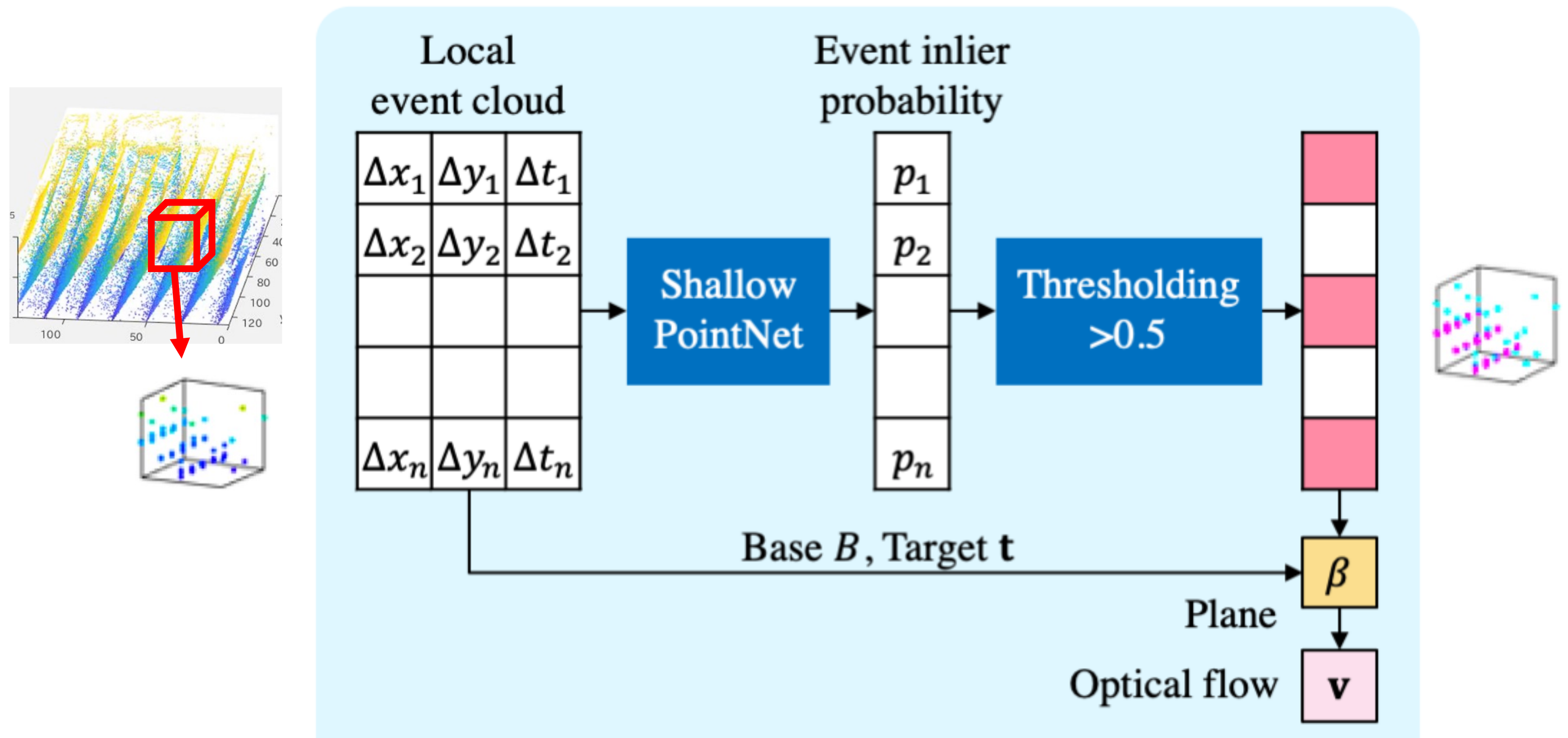
Our purpose is non-iterative selection of inlier events while capturing a plane structure.

2. Method

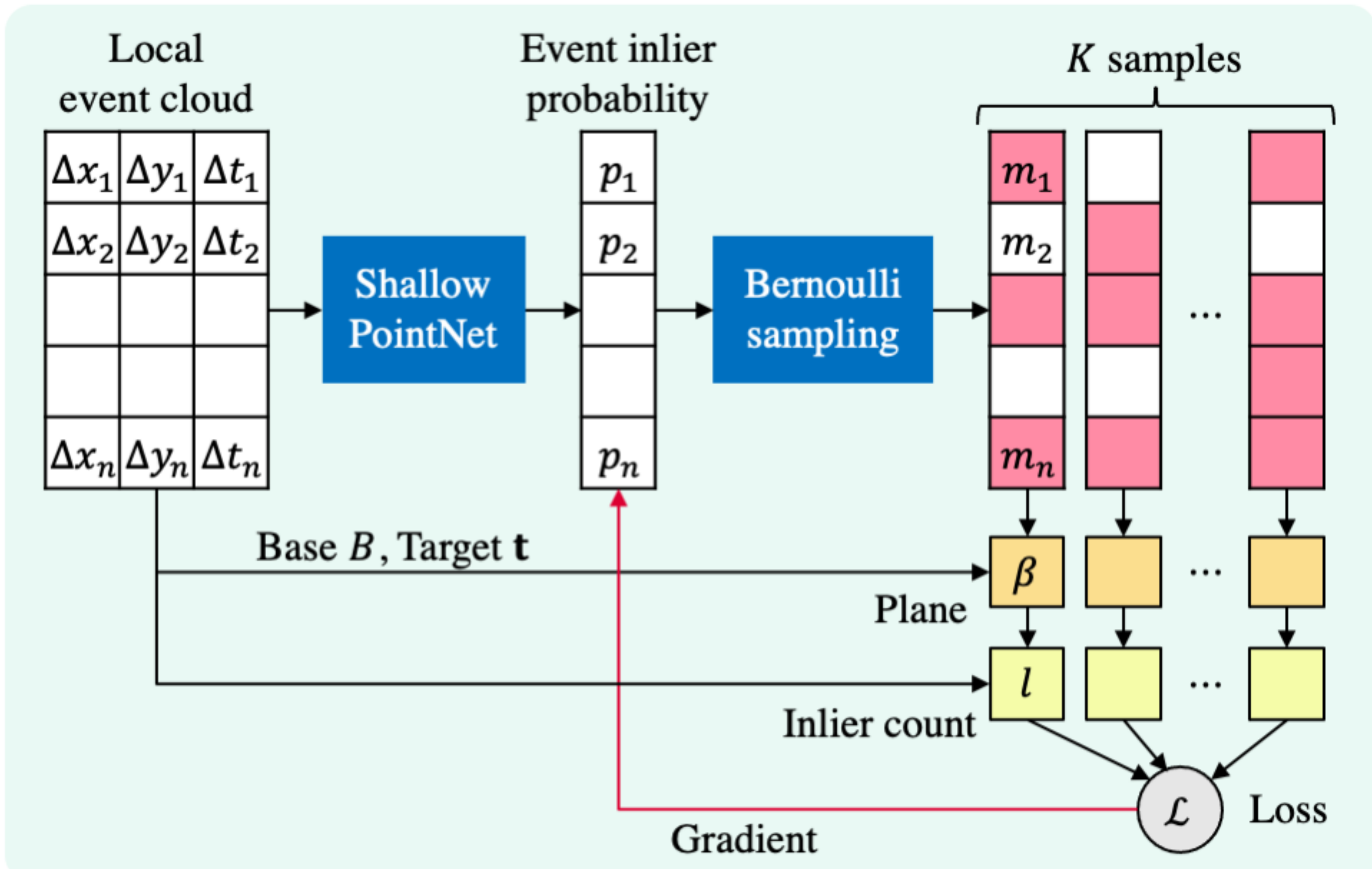
☐ Key ideas

- Outputting the event inlier probability for each event by PointNet, and using only events with higher probabilities
- Training the network in a self-supervised manner by sampling.

☐ Inference



☐ Training



$$L = \mathbb{E}_{\mathbf{m}\sim f(\cdot;\mathbf{p})}[l(\mathbf{m})]$$ Expected negative inlier count loss

$$m_j \sim f(\cdot; p_j) = p_j^{m_j}(1-p_j)^{1-m_j}$$ Bernoulli sampling whether to use

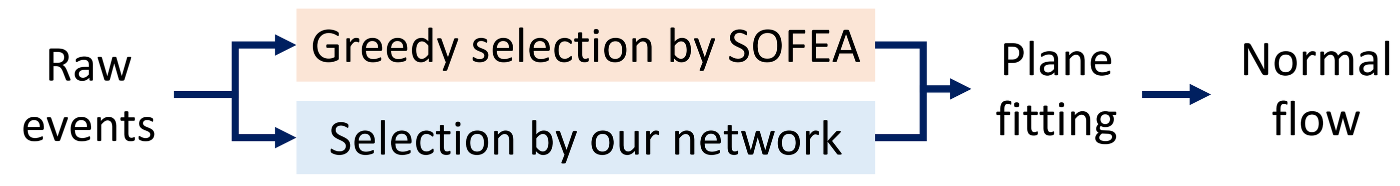
$$\frac{\partial L}{\partial \mathbf{z}} = \mathbb{E}_{\mathbf{m}}\left[l(\mathbf{m})\frac{\partial}{\partial \mathbf{z}}\log f(\cdot;\mathbf{p})\right]$$ The gradient w.r.t. the logit $p = \sigma(z)$

$$\frac{\partial L}{\partial z_j}\log f = \begin{cases} 1-\sigma(z_j) & (m_j = 1) \\ -\sigma(z_j) & (m_j = 0) \end{cases}$$

3. Experiment

☐ Setting

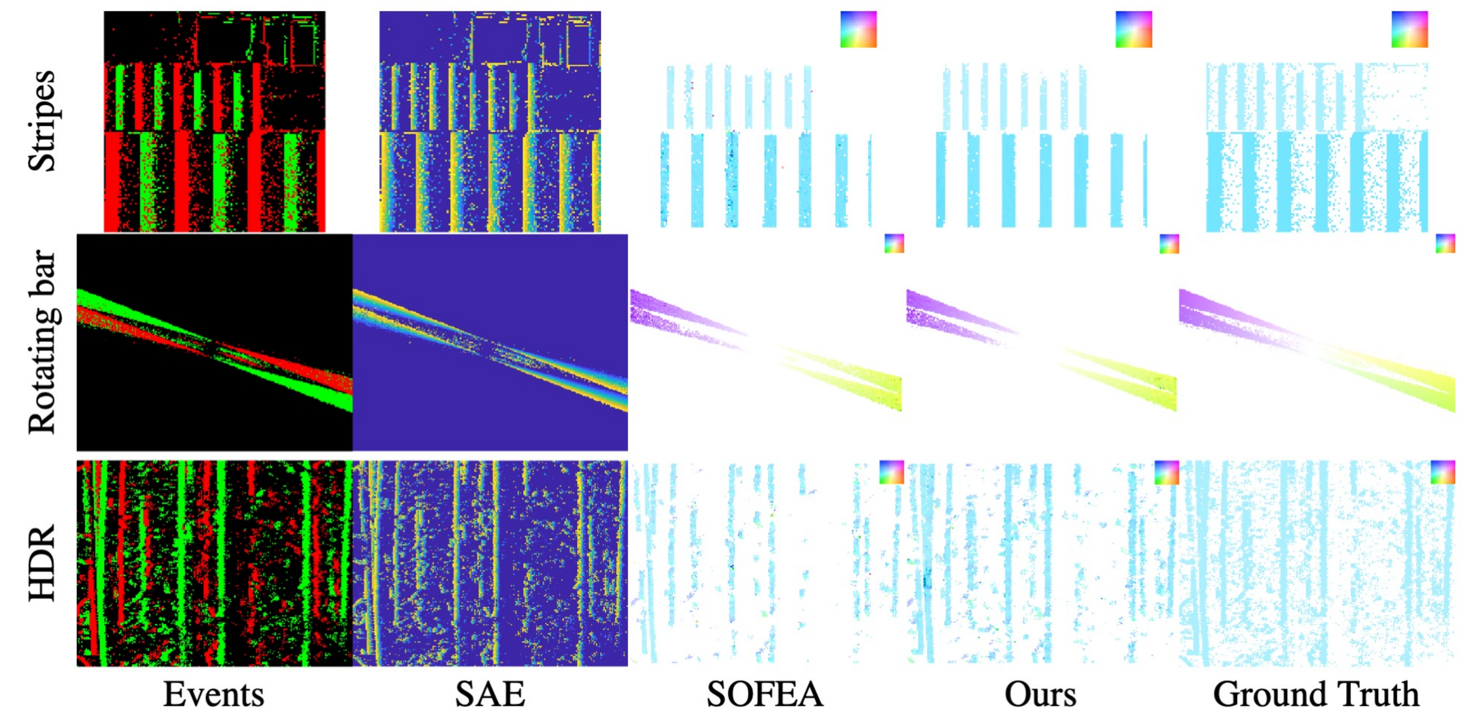
The experiments aimed to confirm the effectiveness of self-supervised learned inlier points within each scene, rather than generalization between sequences, compared with a greedy selection, SOFEA [2]

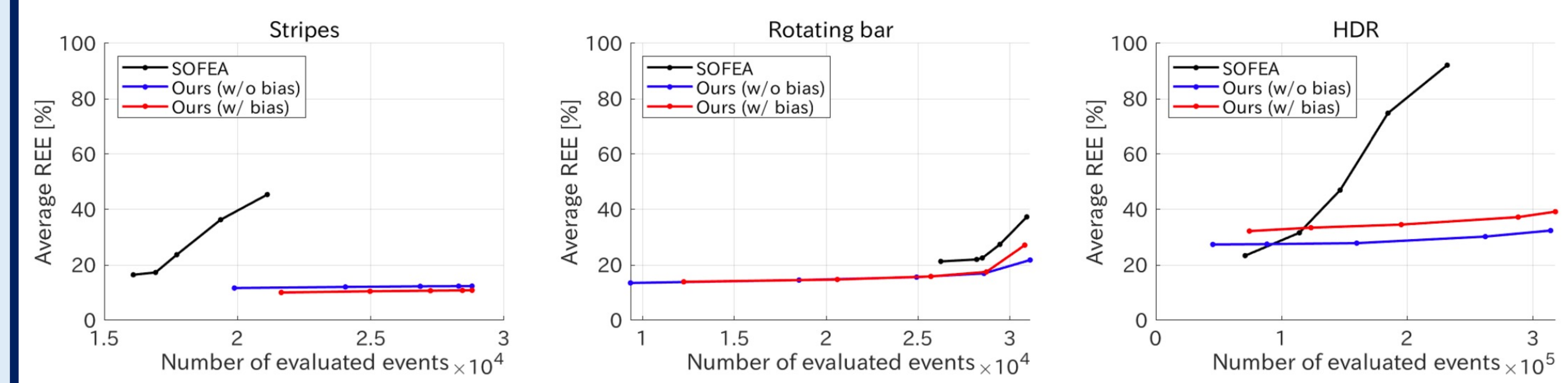


☐ Results

Scene	Stripes		Rotating bar		HDR	
	REE [%]	AE [°]	REE [%]	AE [°]	REE [%]	AE [°]
SOFEA	17.3±14.4	2.06±7.72	22.0±31.1	6.15±6.71	31.6±45.5	8.82±10.4
Ours (w/o bias)	12.3±6.56	0.67±0.57	**16.9±12.3**	**5.72±4.30**	**30.2±33.7**	**5.79±4.76**
Ours (w/ bias)	**10.8±6.09**	**0.65±0.55**	17.5±15.4	5.79±4.76	37.3±48.8	7.16±7.68

REE: Relative endpoint error, AE: Angular error





Ours keeps low errors regardless of the tightness of the rejection because our event selection is able to exclude outliers in advance of fitting by capturing the planar structure with the neural network.

☐ Future work

- Test in realistic and complex scenarios

☐ Reference

[1] Benosman *et al.* "Event-based visual flow", IEEE Transactions on Neural Networks and Learning Systems, 2014
[2] Low *et al.* "SOFEA: A non-iterative and robust optical flow estimation algorithm for dynamic vision sensors", CVPR Workshop, 2020