# TMA: Temporal Motion Aggregation for Event-based Optical Flow — Supplementary Material

Haotian Liu[1] Guang Chen[1]* Sanqing Qu[1] Yanping Zhang[1]
Zhijun Li[1] Alois Knoll[2] Changjun Jiang[1]
[1]Tongji University [2]Technical University of Munich

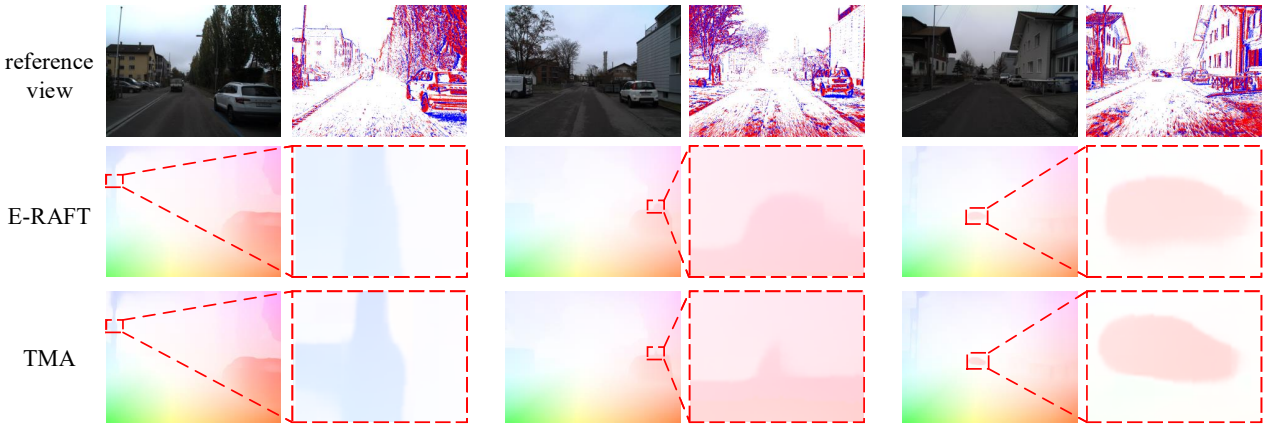


Figure 1: **More qualitative results on DSEC-Flow.** For each two columns, we show the reference image and event frame and compare our method with the state-of-the-art E-RAFT [2]. Significant improvements are highlighted by red boxes.

## 1. More Results on DSEC-Flow

In the main paper, we have visualized some qualitative examples of DSEC-Flow. Here we provide more results to support the superiority of TMA. The results are shown in Figure 1. These examples further demonstrate the effectiveness of TMA when addressing challenging areas (e.g., telegraph pole in the first example and automotive antenna in the second example). Compared with E-RAFT [2], TMA can better generate clear boundaries in flow predictions.

## 2. More Inter-domain Results on MVSEC

We notice that compared with MVSEC [3], the sensors in DSEC-Flow [1] record event data in a denser manner, and the variants of flow magnitudes normalized by the number of pixels in image width are higher, contributing to a more universal driving dataset. We supplement the inter-domain evaluation by only pretraining TMA on DSEC-Flow and evaluating indoor_flying sequences without fine-tuning. As shown in Figure 2, we compare the results on indoor_flying sequences trained on outdoor_day2 or DSEC-Flow. Training on DSEC-Flow provides a great improvement in both EPE and % Outlier. In detail, EPE decreases exceeding 1 and % Outlier decreases exceeding 20% at most.

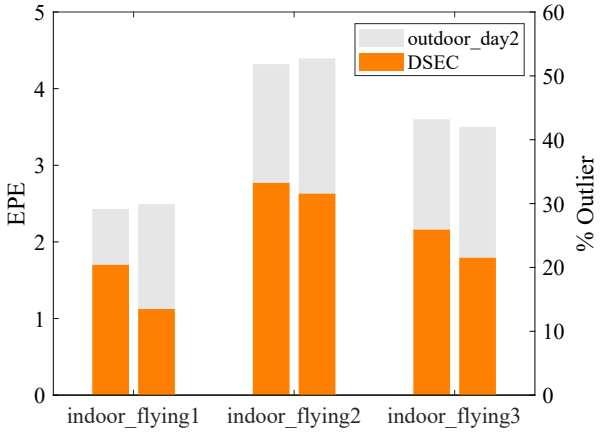


Figure 2: **DSEC to MVSEC.** The figure exhibits the prediction results on MVSEC indoor_flying sequences trained on outdoor_day2 sequences (gray) or DSEC-Flow (orange). In each group, the left bar represents EPE and the right bar represents % Outlier.

*corresponding author: guangchen@tongii.edu.cn

| $dt = 1$ | Train Set + | indoor_flying1 | | indoor_flying2 | | indoor_flying3 | | outdoor_day1 | |
|---|---|---|---|---|---|---|---|---|---|
| | | EPE | % Outlier | EPE | % Outlier | EPE | % Outlier | EPE | % Outlier |
| E-RAFT [2] | indoor_flying1 | - | - | 0.80 | 6.10 | 0.61 | 2.47 | **0.25** | **0.11** |
| Ours | | - | - | **0.76** | **5.84** | **0.59** | **2.37** | 0.26 | 0.09 |
| E-RAFT [2] | indoor_flying2 | 0.48 | 0.07 | - | - | 0.55 | 1.03 | 0.25 | 0.08 |
| Ours | | **0.47** | **0.03** | - | - | **0.54** | **0.38** | **0.24** | **0.03** |
| E-RAFT [2] | indoor_flying3 | 0.49 | 0.08 | **0.60** | **0.78** | - | - | 0.26 | 0.15 |
| Ours | | **0.47** | **0.06** | 0.62 | 1.17 | - | - | **0.25** | **0.06** |

Table 1: **Evaluation results on MVSEC [3] training with outdoor_day2 and one indoor_flying sequence.** + means the indoor_flying sequence to be included into training set.

## 3. More Intra-domain Results on MVSEC

In addition, we have conducted intra-domain evaluation on MVSEC [3] in the main paper. Especially, to boost accuracy of our method under a supervised learning setting, we include one indoor_flying sequence into training set ($dt = 4$ grayscale frames). Here we provide more results to support the superiority of TMA for $dt = 1$ grayscale frames in Table 1. Similar to the results corresponding to $dt = 4$ grayscale frames, TMA achieves a significant accuracy improvement for $dt = 1$ on indoor_flying sequences. In detail, on indoor_flying1, EPE decreases at most 0.59 (1.06 $\rightarrow$ 0.47); on indoor_flying2, EPE decreases at most 1.19 (1.81 $\rightarrow$ 0.62); and on indoor_flying3, EPE decreases at most 1.04 (1.58 $\rightarrow$ 0.54); The results confirm the presence of domain gap between outdoor day and indoor flying data. By including more training data to reduce the domain gap, the accuracy of our method can be further enhanced.

## References


[1] Mathias Gehrig, Willem Aarents, Daniel Gehrig, and Davide Scaramuzza. Dsec: A stereo event camera dataset for driving scenarios. *IEEE Robotics and Automation Letters*, 6(3):4947–4954, 2021. 1

[2] Mathias Gehrig, Mario Millhäusler, Daniel Gehrig, and Davide Scaramuzza. E-raft: Dense optical flow from event cameras. In *2021 International Conference on 3D Vision (3DV)*, pages 197–206. IEEE, 2021. 1, 2

[3] Alex Zihao Zhu, Dinesh Thakur, Tolga Özaslan, Bernd Pfrommer, Vijay Kumar, and Kostas Daniilidis. The multivehicle stereo event camera dataset: An event camera dataset for 3d perception. *IEEE Robotics and Automation Letters*, 3(3):2032–2039, 2018. 1, 2