

## A Supplementary Material

### A.1 Time-Awareness: PDE solutions

The proposed *time-aware flow* is given as the solution to (7). Letting the flow be  $\mathbf{v} = (v_x, v_y)^\top$ , the system of PDEs can be written as:

$$\begin{aligned} v_x \frac{\partial v_x}{\partial x} + v_y \frac{\partial v_x}{\partial y} + \frac{\partial v_x}{\partial t} &= 0, \\ v_x \frac{\partial v_y}{\partial x} + v_y \frac{\partial v_y}{\partial y} + \frac{\partial v_y}{\partial t} &= 0. \end{aligned} \quad (10)$$

Upwind and Burgers' schemes can be used to discretize and numerically solve the system of PDEs [11, 41].

**Discretization.** Let  $\mathbf{v}^n(x, y)$  be the flow vector at discretized space- (e.g., pixel) and time-indices  $(x, y, n)$ , with discretization steps  $\Delta x$ ,  $\Delta y$ , and  $\Delta t$ , respectively, and let the forward (+) and backward (−) differences of a scalar field  $w$  (e.g.,  $v_x^n$  or  $v_y^n$ ) be defined as

$$\begin{aligned} D_x^+ w &\equiv \frac{\partial w^+}{\partial x} = \frac{1}{\Delta x} (w(x + \Delta x, y) - w(x, y)), \\ D_y^+ w &\equiv \frac{\partial w^+}{\partial y} = \frac{1}{\Delta y} (w(x, y + \Delta y) - w(x, y)), \end{aligned} \quad (11)$$

and

$$\begin{aligned} D_x^- w &\equiv \frac{\partial w^-}{\partial x} = \frac{1}{\Delta x} (w(x, y) - w(x - \Delta x, y)), \\ D_y^- w &\equiv \frac{\partial w^-}{\partial y} = \frac{1}{\Delta y} (w(x, y) - w(x, y - \Delta y)). \end{aligned} \quad (12)$$

We discretize in time using forward differences,  $\frac{\partial w}{\partial t} \approx (w(t + \Delta t) - w(t))/\Delta t$ , to yield explicit update schemes:  $w(t + \Delta t) \approx w(t) + \Delta t \frac{\partial w}{\partial t}$ .

**Upwind scheme.** The first-order upwind scheme is an explicit scheme that updates the flow as follows, based on the sign of the variables: it uses  $D_x^+ v_x^n$  and  $D_x^- v_x^n$  for  $v_x^n > 0$  ( $D_x^- v_x^n$  and  $D_x^+ v_x^n$  otherwise), and  $D_y^+ v_y^n$  and  $D_y^- v_y^n$  for  $v_y^n > 0$  ( $D_y^- v_y^n$  and  $D_y^+ v_y^n$  otherwise). The scheme is stable if the flow satisfies  $\Delta t \max\{|v_x|/\Delta x + |v_y|/\Delta y\} < 1$  (CFL stability condition [22]). For example, in case that  $v_x^n > 0$  and  $v_y^n > 0$  at the current discretization time  $n$ :

$$\begin{aligned} v_x^{n+1} &= v_x^n - \Delta t (v_x^n D_x^+ v_x^n + v_y^n D_y^+ v_x^n), \\ v_y^{n+1} &= v_y^n - \Delta t (v_y^n D_y^+ v_y^n + v_x^n D_x^+ v_y^n). \end{aligned} \quad (13)$$

**Burgers' scheme.** The study of the inviscid Burgers' equation provides a more conservative scheme solution, especially at “shock” and “fan wave” cases [41]. In this explicit scheme, the product terms in the same variable (which

convey that the flow is transporting itself),  $v_x^n D_x^+ v_x^n$  and  $v_y^n D_y^+ v_y^n$  in (13), are replaced with  $U_x$  and  $U_y$  respectively, which are given by:

$$\begin{aligned} U_x &= \frac{1}{2} \left( \text{sgn}(v_x^n(x, y)) (v_x^n(x, y))^2 + F_x - B_x \right), \\ F_x &= \begin{cases} (v_x^n(x + \Delta x, y))^2, & \text{if } v_x^n(x + \Delta x, y) < 0 \\ 0, & \text{otherwise} \end{cases} \\ B_x &= \begin{cases} (v_x^n(x - \Delta x, y))^2, & \text{if } v_x^n(x - \Delta x, y) > 0 \\ 0, & \text{otherwise} \end{cases} \end{aligned} \quad (14)$$

and

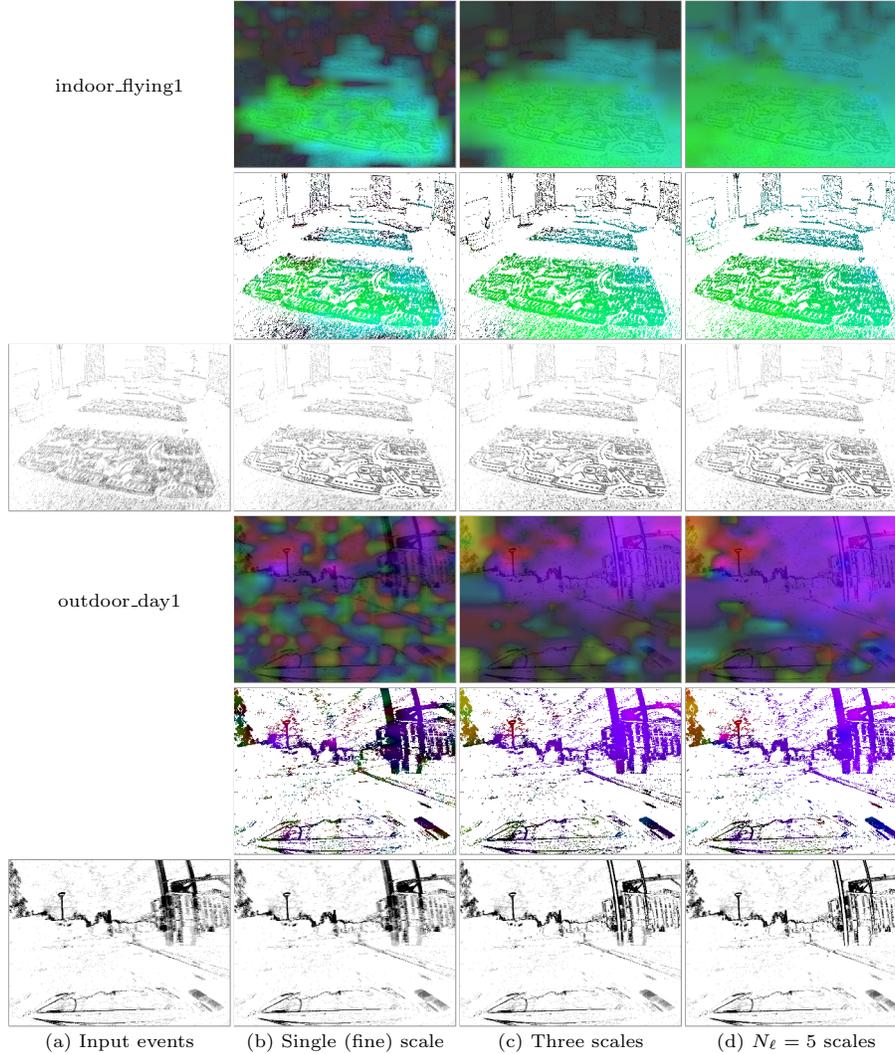
$$\begin{aligned} U_y &= \frac{1}{2} \left( \text{sgn}(v_y^n(x, y)) (v_y^n(x, y))^2 + F_y - B_y \right), \\ F_y &= \begin{cases} (v_y^n(x, y + \Delta y))^2, & \text{if } v_y^n(x, y + \Delta y) < 0 \\ 0, & \text{otherwise} \end{cases} \\ B_y &= \begin{cases} (v_y^n(x, y - \Delta y))^2, & \text{if } v_y^n(x, y - \Delta y) > 0 \\ 0, & \text{otherwise} \end{cases} \end{aligned} \quad (15)$$

## A.2 Effect of the Multi-scale Approach

The effect of the proposed multi-scale approach (Fig. 4) is shown in Fig. 9. This experiment compares the results of using multi-scale approaches (in a coarse-to-fine fashion) versus using a single (finest) scale. With a single scale, the optimizer gets stuck in a local extremal, yielding an irregular flow field (see the optical flow rows), which may produce a blurry IWE (e.g., outdoor\_day1 scene). With three scales (finest tile and two downsampled ones), the flow becomes less irregular than with one single scale, but there may be regions with few events where the flow is difficult to estimate. With five scales the flow becomes smoother, more coherent over the whole image domain, while still being able to produce sharp IWEs.

## A.3 Sensitivity Analysis

**A.3.1 The choice of loss function.** Table 5 shows the results on the MVSEC benchmark for different loss functions. We compare the (squared) gradient magnitude, image variance, average timestamp [58], and normalized average timestamp [21]. The gradient magnitude and image variance losses produce the best accuracy compared with the two average timestamp losses. Quantitatively, the image variance loss gives competitive results with respect to the gradient magnitude. However, for the reasons described in Sec. 3.2, and because the image variance sometimes overfits, we use gradient magnitude. Both average timestamp losses are trapped in the global optima which pushes all events out of the image



**Fig. 9.** *Effect of the multi-scale approach.* For each sequence, the top row shows the estimated flow, the middle row shows the estimated flow masked by the events, and the bottom row shows the IWEs.

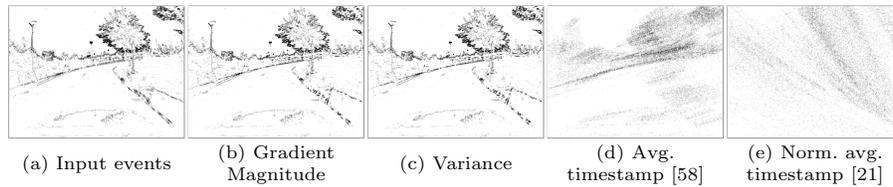
plane, hence, they provide very large errors (marked as “> 99” in Tab. 5). This effect is visualized in Fig. 10.

*Remark:* Maximization of (6) does not suffer from the problem mentioned in [21] that affects the average timestamp loss function, namely that the optimal flow warps all events outside the image so as to minimize the loss (undesired global optima shown in Fig. 10d-10e). If most events were warped outside of

the image, then (6) would be smaller than the identity warp, which contradicts maximization.

**Table 5.** Sensitivity analysis on the choice of loss function (MVSEC,  $dt = 4$ ). The contrast and gradient magnitude functions provide notably better results than the losses based on average timestamps.

|                           | indoor_flying1 |              | indoor_flying2 |              | indoor_flying3 |              | outdoor_day1 |             |
|---------------------------|----------------|--------------|----------------|--------------|----------------|--------------|--------------|-------------|
|                           | AEE ↓          | %Out ↓       | AEE ↓          | %Out ↓       | AEE ↓          | %Out ↓       | AEE ↓        | %Out ↓      |
| Gradient magnitude [13]   | <b>1.68</b>    | 12.79        | 2.49           | 26.31        | 2.06           | 18.93        | <b>1.25</b>  | <b>9.19</b> |
| Image variance [15]       | 1.70           | <b>11.25</b> | <b>2.18</b>    | <b>21.91</b> | <b>1.93</b>    | <b>15.84</b> | 1.82         | 15.89       |
| Avg. timestamp [58]       | >99            | >99          | >99            | >99          | >99            | >99          | >99          | >99         |
| Norm. avg. timestamp [21] | >99            | >99          | >99            | >99          | >99            | >99          | >99          | >99         |



**Fig. 10.** *IWEs for different loss functions.* Average timestamp losses overfit to undesired global optima, which pushes most events out of the image plane.

**A.3.2 The regularizer weight.** Table 6 shows the sensitivity analysis on the regularizer weight  $\lambda$  in (9).  $\lambda = 0.0025$  provides the best accuracy in the outdoor sequence, while  $\lambda = 0.025$  provides slightly better accuracy in the indoor sequences. Comparing their accuracy differences, we use the former because it has a higher accuracy gain.

**Table 6.** Sensitivity analysis on the regularizer weight (MVSEC data,  $dt = 4$ ).

|                    | indoor_flying1 |        | indoor_flying2 |        | indoor_flying3 |        | outdoor_day1 |        |
|--------------------|----------------|--------|----------------|--------|----------------|--------|--------------|--------|
|                    | AEE ↓          | %Out ↓ | AEE ↓          | %Out ↓ | AEE ↓          | %Out ↓ | AEE ↓        | %Out ↓ |
| $\lambda = 0.0025$ | 1.68           | 12.79  | 2.49           | 26.31  | 2.06           | 18.93  | 1.25         | 9.19   |
| $\lambda = 0.025$  | 1.52           | 9.07   | 2.39           | 26.26  | 1.94           | 18.44  | 1.86         | 17.11  |
| $\lambda = 0.25$   | 1.89           | 16.54  | 3.19           | 36.95  | 2.91           | 30.85  | 2.57         | 27.86  |

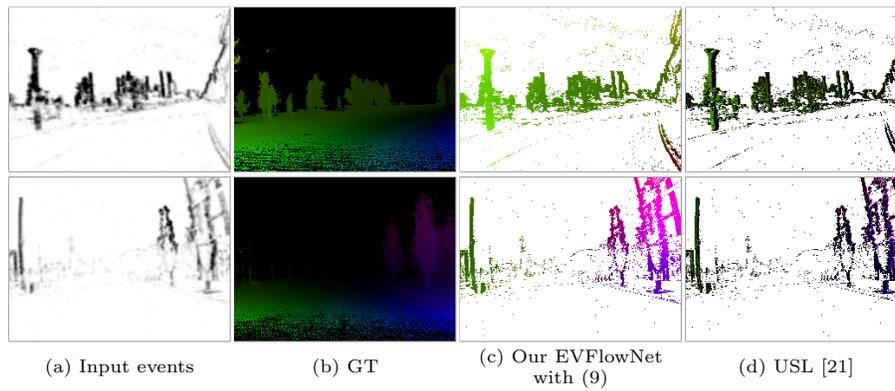
## A.4 Additional Results

**A.4.1 Full results on DSEC test sequences.** For completeness, Tab. 7 reports the results on all DSEC test sequences. No GT flow is available for these sequences. As mentioned in Sec. 4.3, only one competing method is available at the time of publication, which is supervised learning. Our method consistently produces better FWL (i.e., sharper IWEs) than the supervised learning method, which suffers from GT issues to produce sharp IWEs. The sharpness differences are most significant in IMOs and on the road close to the vehicle (see Fig. 6). The FWL is computed using the same 100ms intervals used for the accuracy benchmark calculation. Since the FWL is sensitive to the number of events, the previous convention is consistent with the benchmark.

**Table 7.** Results on the DSEC optical flow benchmark [18].

|             | All       |        |             | interlaken_00.b  |        |             | interlaken_01.a  |        |             | thun_01.a        |        |             |
|-------------|-----------|--------|-------------|------------------|--------|-------------|------------------|--------|-------------|------------------|--------|-------------|
|             | AEE ↓     | %Out ↓ | FWL ↑       | AEE ↓            | %Out ↓ | FWL ↑       | AEE ↓            | %Out ↓ | FWL ↑       | AEE ↓            | %Out ↓ | FWL ↑       |
| E-RAFT [18] | 0.79      | 2.68   | 1.29        | 1.39             | 6.19   | 1.32        | 0.90             | 3.91   | 1.42        | 0.65             | 1.87   | 1.20        |
| Ours        | 3.47      | 30.86  | <b>1.37</b> | 5.74             | 38.93  | <b>1.50</b> | 3.74             | 31.37  | <b>1.51</b> | 2.12             | 17.68  | <b>1.24</b> |
|             | thun_01.b |        |             | zurich_city_12.a |        |             | zurich_city_14.c |        |             | zurich_city_15.a |        |             |
|             | AEE ↓     | %Out ↓ | FWL ↑       | AEE ↓            | %Out ↓ | FWL ↑       | AEE ↓            | %Out ↓ | FWL ↑       | AEE ↓            | %Out ↓ | FWL ↑       |
| E-RAFT [18] | 0.58      | 1.52   | 1.18        | 0.61             | 1.06   | 1.12        | 0.71             | 1.91   | 1.47        | 0.59             | 1.30   | 1.34        |
| Ours        | 2.48      | 23.56  | <b>1.24</b> | 3.86             | 43.96  | <b>1.14</b> | 2.72             | 30.53  | <b>1.50</b> | 2.35             | 20.99  | <b>1.41</b> |

**A.4.2 Qualitative results for DNN.** Additional qualitative results of our unsupervised learning setting (Sec. 4.6) are shown in Fig. 11. We compare our method with the state-of-the-art unsupervised learning [21]. Our results resemble the GT flow. See Tab. 4 for the quantitative result.



**Fig. 11.** *Result of our DNN on the MVSEC outdoor sequence.* Our DNN (EV-FlowNet architecture) trained with (9) produces better result than the state-of-the-art unsupervised learning method [21]. For a quantitative comparison, see Table 4.