# Unsupervised Flow Refinement near Motion Boundaries

Shuzhi Yu
shuzhiyu@cs.duke.edu

Hannah Halin Kim
hannah@cs.duke.edu

Shuai Yuan
shuai@cs.duke.edu

Carlo Tomasi
tomasi@cs.duke.edu

Duke University
Durham, NC, USA

## A    Pre-trained Models

All the pre-trained models used in our experiments, namely SMURF [3], ARFlow [2], and RAFT [4] (used in the following Table 4) are provided by their authors. Specifically, for experiments on the Sintel dataset, the ARFlow model is (unsupervisedly) fine-tuned on the Sintel training set; neither the SMURF model nor RAFT model are fine-tuned on the Sintel training set. The pre-trained SMURF and ARFlow models used in the KITTI experiments are (unsupervisedly) fine-tuned on the KITTI-2015 training set.

We keep the default parameters of the classical flow estimator LDOF [1] in our experiments.

## B    Impact of Hyperparameters

A larger $\theta_{ism}$ reduces the candidates of invalid smooth motion, and thus, reduces the number of predicted MB points. This results in a smaller improvement in motion boundary detection (Table 1). Nevertheless, our method still outperforms the baseline method over a large range of $\theta_{ism}$.

| $\theta_{ism}$ | 0.1 | 0.2 | 0.4 | 0.6 | 0.8 | Baseline |
|---|---|---|---|---|---|---|
| clean | 74.6 | 74.5 | 73.5 | 72.2 | 71.2 | 70.3 |
| final | 67.8 | 67.4 | 65.9 | 65.0 | 64.4 | 63.5 |

Table 1: Impact of the threshold $\theta_{ism}$ for $M_{ism}$ on MB detection in F1 score (%), with flow estimates by SMURF.

Table 2 shows the improvements of the flow estimates of the points in set $P$ by our refinement method with different $\alpha$ and $\tau$. Increasing $\alpha$ (left panel in Table 2) sets a stricter

criterion for Equation 5 (in the main paper), and thus, less points are selected into set $P$. For instance, with $\tau = 0.2$ and $\alpha = 1.0$, the number of replaced points is reduced from 61% to 44% of MB points, compared with setting both $\tau$ and $\alpha$ as 0.2. Those selected points are largely affected by the flow estimates across the motion boundaries since the estimated flows of point $\mathbf{q}$ and $\mathbf{q}'$ on the two sides of the motion boundary are large. Replacing flow for such points likely leads to better rewards.

Similarly, increasing $\tau$ (right panel in Table 2) would decrease the smallest safe distance $d^*$, and thus, less points are selected into the set $P$. In addition, those selected points in set $P$ are very close to motion boundaries (MBs) where the estimates are poorer, resulting in a generally increasing improvement by the replacement method.

| Sintel | different $\alpha$ ($\tau = 0.2$) | | | | different $\tau$ ($\alpha = 0.2$) | | | |
|---|---|---|---|---|---|---|---|---|
| | 0.2 | 0.5 | 0.8 | 1.0 | 0.2 | 0.5 | 0.8 | 1.0 |
| Clean | 5.4 | 6.0 | 6.0 | 6.0 | 5.4 | 6.5 | 7.0 | 7.2 |
| Final | 17.4 | 19.0 | 19.3 | 19.3 | 17.4 | 18.2 | 17.7 | 17.3 |

Table 2: Effect of $\tau$ (Eq. 4 in the main paper) and $\alpha$ (Eq. 5 in the main paper) on flow estimates in terms of the percentage (%) of reduction in EPE (higher = better).

# C  Impact of Bi-directional Flow on MB Detection

The forward and backward flow complement each other since one may be more accurate than the other in different regions. Thus, using the bi-directional flow as inputs benefits both the baseline and our detection method on motion boundaries. As shown in Table 3, using bidirectional flows as inputs to the detection methods (rightmost column in each panel) consistently outperforms those with only input flow of a single direction, over both the clean and final passes of the Sintel dataset. This benefit is also demonstrated by the two examples (from the clean pass of Sintel training set) in Figure 1. Particularly, the red rectangle marks the motion boundaries that are better detected using only the forward flow for the detection than using only the backward flow, and the blue rectangle indicates the opposite situation. The predictions in the last column combine the advantages of the second and third columns.

In addition, our method consistently outperforms the baseline method regardless what input flow is used.

| | | Flow (EPE) | Baseline | | | Ours | | |
|---|---|---|---|---|---|---|---|---|
| | | | Fward (F1) | Bward (F1) | Bidirect (F1) | Fward (F1) | Bward (F1) | Bidirect (F1) |
| Sintel | Clean | 2.01 | 68.1 | 67.1 | 70.3 | 69.7 | 68.7 | **74.5** |
| | Final | 2.87 | 61.7 | 60.6 | 63.5 | 63.0 | 61.9 | **67.4** |

Table 3: $F_1$-score for our **MB estimation** with input flow estimates (by SMURF [◻]) of different directions, namely forward (Fward), backward (Bward), and bidirectional (Bidirect), compared with the baseline method.

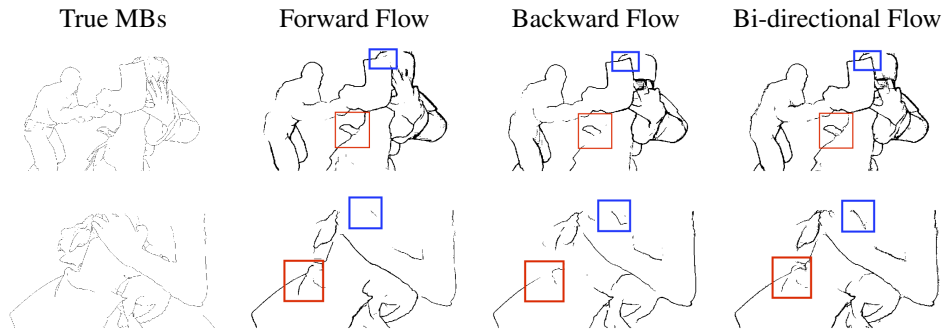| True MBs | Forward Flow | Backward Flow | Bi-directional Flow |
| --- | --- | --- | --- |



Figure 1: Two MB detection samples from the Sintel (clean) training set with our method using input flow of different directions. The red box marks the superior predictions of using only forward flow for detection than using only the backward flow, and the opposite situation is indicated by the blue box. The input flow is estimated by SMURF [3].

# D   Replacement on Supervised Flow Estimator

Although our proposed method is under the unsupervised setting, it can also take as input the flow estimates by supervised flow estimators. Table 4 shows that our method can also correct some flow estimates by a top supervised flow estimator RAFT [4] near our detected motion boundaries.

| Input Flow | Dataset | Input Flow AEPE | Replaced Points | | | |
| --- | --- | --- | --- | --- | --- | --- |
| | | | % of MB points | Init AEPE | Our AEPE | ↓ |
| RAFT | Clean | 1.48 | 45.17 | 3.57 | **3.46** | 3.08% |
| | Final | 2.83 | 34.14 | 4.77 | **3.96** | 16.98% |

Table 4: Average EPE and average EPE improvement with our replacement method near our estimated MBs of the flow estimates by RAFT [4]. About 1% of all MPI Sintel pixels are true MB points.

# References

[1] Thomas Brox and Jitendra Malik. Large displacement optical flow: descriptor matching in variational motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(3):500–513, 2011. URL http://lmb.informatik.uni-freiburg.de/Publications/2011/Bro11a.

[2] Liang Liu, Jiangning Zhang, Ruifei He, Yong Liu, Yabiao Wang, Ying Tai, Donghao Luo, Chengjie Wang, Jilin Li, and Feiyue Huang. Learning by analogy: Reliable supervision from transformations for unsupervised optical flow estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6489–6498, 2020.

[3] Austin Stone, Daniel Maurer, Alper Ayvaci, Anelia Angelova, and Rico Jonschkowski. Smurf: Self-teaching multi-frame unsupervised raft with full-image warping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3887–3896, 2021.

[4] Zachary Teed and Jia Deng. Raft: Recurrent all-pairs field transforms for optical flow. In *European Conference on Computer Vision*, pages 402–419. Springer, 2020.