Supplementary Materials
S2F: Slow-To-Fast Interpolator Flow

Yanchao Yang Stefano Soatto
UCLA Vision Lab, University of California, Los Angeles, CA 90095
{yanchao.yang, soatto}@cs.ucla.edu

Implementation Details and Thresholds

In this section we list all the threshold and parameters, and their numerical values, kept constant throughout each experiment. Together with the main body of the paper, this enables independent implementation and reproduction of the results. More importantly, we plan to release our implementation, so independent validation is straightforward.

The three key threshold in our algorithm are: ε_r on residual, ε_c on compatibility check, and increment of radii δ. $\varepsilon_r = 30$ for all datasets (notice that we don't calculate residuals based on RGB color space, instead we use Lab color space for more invariance to illumination change). And $\varepsilon_c = 5$, $\delta = 2$ for both Sintel and Middlebury datasets; $\varepsilon_c = 1$, $\delta = 1$ for Kitti dataset, which are inline with the parameter setting in [1].

Q&A

In this section we answer questions likely to be raised during the review process.

Q0: *The paper is not ranked #1 on Sintel!*

Answer: Indeed, since we posted our entry, and just before the deadline, someone else outperformed us on Sintel, so our paper is now ranked #3. We believe this does not diminish the value of our approach, which requires no learning and focuses on a specific aspect of the general flow problem. Combining our approach with some learning, and reclaiming the top spot, is our current focus.

Q1: *Why start with a small radius and then increase it? Isn't it better to pick a single, large, radius?*

Answer: Small regions are less discriminative but yield more accurate localization. Large regions are more discriminative, but less accurate especially when they straddle occluding boundaries (Fig.1). Notice that more matches are found on the hand and sleeve of the girl, as well as the background, but less accurate (brighter = larger error). Also see Tab.1 for a quantitative comparison on the pair of images in Fig.1.

Also, larger regions take longer to match (for quantitative analysis on the time cost of matching using different radii, please refer to [1]).

Q2: *How do you deal with occlusions?*

Answer: With each iteration, we also update the estimated occlusion O, and black out the occlusion from the unmatched region. We
1. Detect pixels O_{out} that exit the image domain if warped using the current flow.
2. Detect pixels $O_{collide}$ that collide if warped using the updated flow f. Note that this set of pixels contains both occluded pixels and occluding pixels. Then we need to identify pixels O_{high} that have high residual. Since occluding pixels are covisible, and their residuals are always low, so the occluded pixels form the set $O_{occ} = O_{collide} \cap O_{high}$.

Radius(pixels)	EPE≤1	EPE=2	EPE=3	EPE=4	EPE=5	EPE>5
6	0.917	0.035	0.023	0.014	0.005	0.006
12	0.889	0.047	0.029	0.021	0.007	0.007

Table 1. This table shows the percentage of erroneous matches with certain endpoint-error, found with two different radii on the images shown in Fig.1. The percentage of matches with endpoint-error less than or equal to 1 pixel is higher when the radius is smaller.

3. The final estimation of occlusion is $O = O_{out} \cup O_{occ}$.

Q3: *If larger radii yields less accurate matches, what to do?*

Answer: At the end of the estimation procedure, we apply the following steps to remove possible outliers in the matching list:
1. For each match (m_i^1, m_i^2), we can find the nearest N(200) matches, M_i^N, by looking at the geodesic distance defined in Sec. 3.3 of the main paper.
2. Given the interpolated flow f, we can calculate the discrepancy between m_j^2 and $m_j^1 + f(m_j^1)$ for $j \in M_i^N$, then we can have the standard deviation σ of these N discrepancies.
3. If $|m_i^1 - m_i^2| > 2 * \sigma$, we remove (m_i^1, m_i^2) from the matching list.

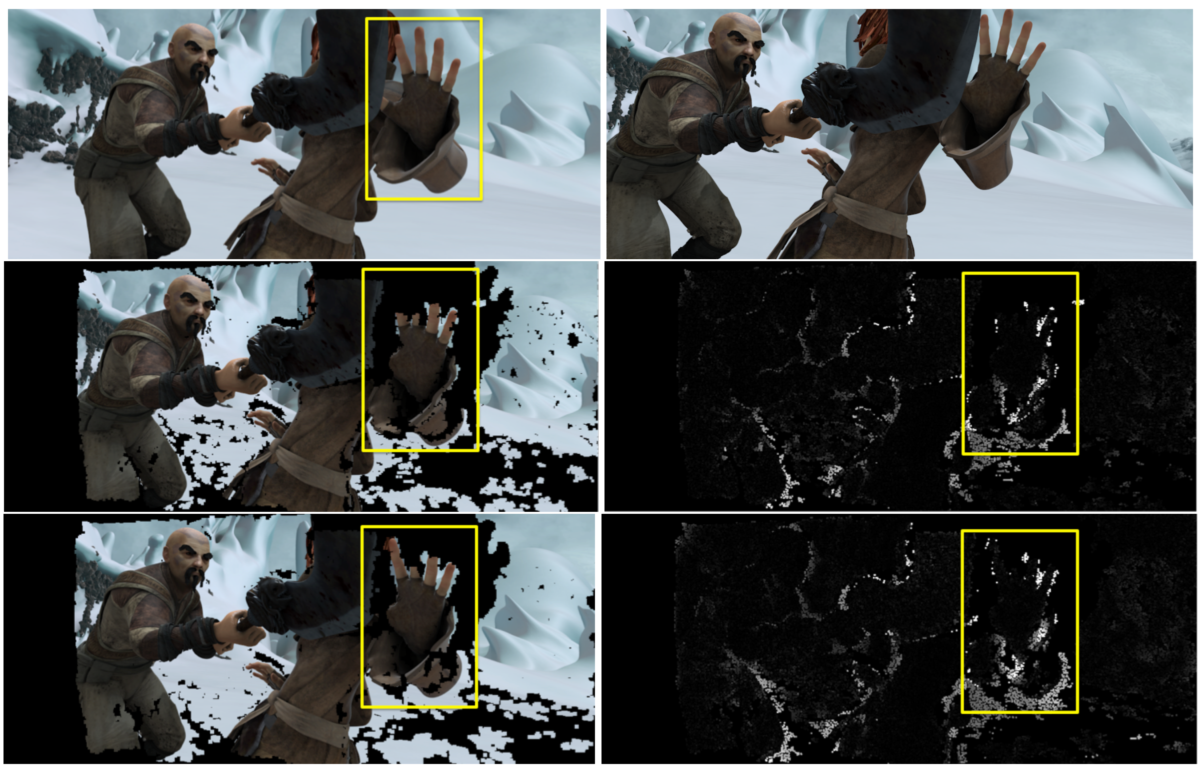

Figure 1. First row: test image and target image from Sintel training set, with known ground-truth flow. Second row: highlighted region, matched area on the test image, found using radius 6(pix), pixels within half of the radius around a match are also shown as matched; endpoint-error map of each match, brighter means higher error. Third row: similar to second row, with radius 12(pix).

Additional Qualitative results

See Fig.2 for more qualitative comparisons to [1].

Figure 2. More comparisons to [1]. Left to right: overlapped image pairs, results of [1], results of the proposed method. Details are highlighted using yellow rectangles. Kitti: 1st-4th row; Sintel: 5th-9th row.

References

[1] Christian Bailer, Bertram Taetz, and Didier Stricker. Flow fields: Dense correspondence fields for highly accurate large displacement optical flow estimation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4015–4023, 2015.