

Video Segmentation via Object Flow

Yi-Hsuan Tsai
UC Merced

ytsai2@ucmerced.edu

Ming-Hsuan Yang
UC Merced

mhyang@ucmerced.edu

Michael J. Black
MPI for Intelligent Systems

black@tuebingen.mpg.de

1. Model Analysis

We analyze the proposed segmentation model by evaluating the importance of appearance and location terms in Figure 1. For instance, in sequences such as *Penguin* and *Frog*, the object appearance is similar to the background with slow motions, and hence the location term in the model plays an important role to achieve better results. On the other hand, for non-rigid objects (*Soldier*, *Monkey*), the appearance term with an online updated model is able to handle the large appearance deformation. Note that with the combination of the location and appearance terms, our full model obtains better performance compared to only using one of them.

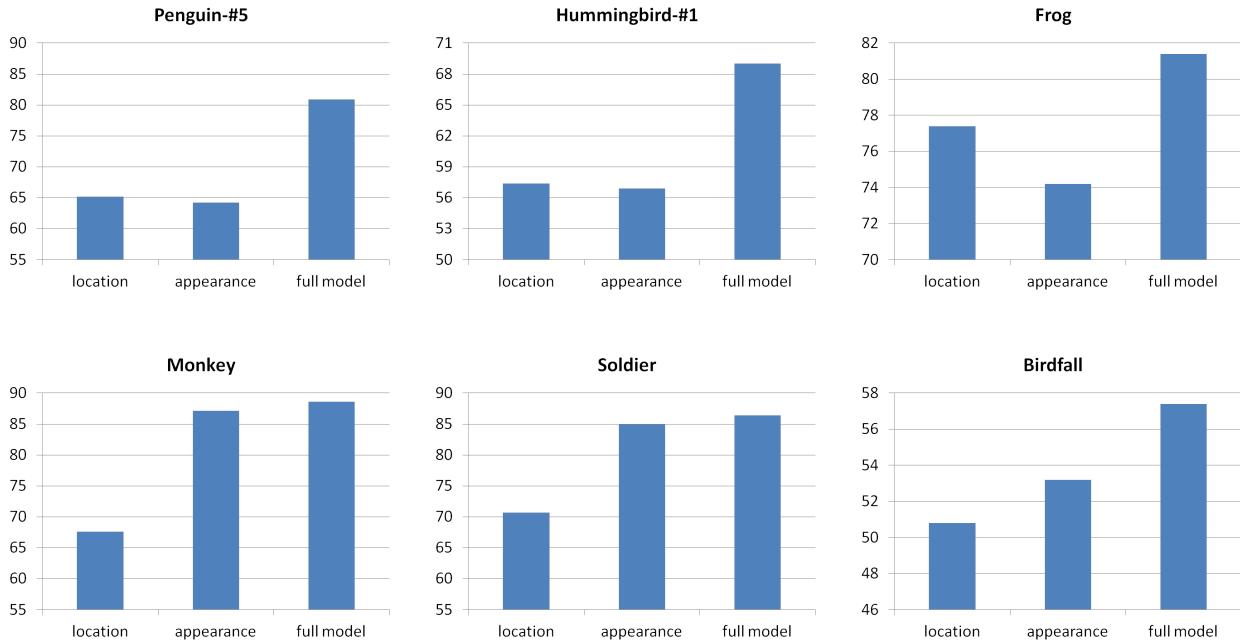


Figure 1. Model analysis with comparisons of the model only using location term, the model only using appearance term, and the full model combining both terms. The y-axis is the intersection-over-union ratio.

2. Effectiveness of the Multi-level Model

We present the comparison of multi-level and single-level models of the proposed algorithm on the SegTrack v2 dataset in Table 1. For the pixel output in the multi-level model, the accuracy is improved by a large margin in many sequences due to the help from superpixel level. Specifically, the superpixel term enhances temporal information so that the model can handle cases including fast movement and background noise in sequences such as *Drifting*, *Monkeydog-Monkey*, *Worm* and *Birdfall*. In addition, the superpixel output with multi-level performs much better than only considering the superpixel level, especially in sequences such as *Girl*, *Cheetah*, *BMX-Person* and *Hummingbird* that contain unclear object boundaries.



BMX-Person



Bird of Paradise



MonkeyDog-Monkey



Drifting-#1



Hummingbird-#1



Frog

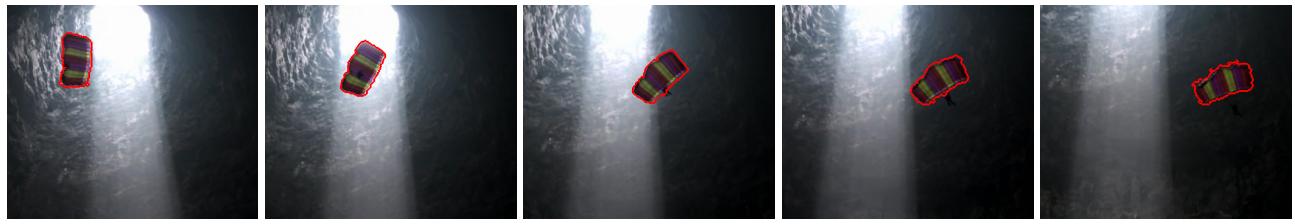


Worm



Soldier

Figure 2. Example results for segmentation in eight sequences on the SegTrack v2 dataset. The output on the pixel level of our multi-level model is indicated as the red contour. Best viewed in color.



Parachute



Monkey



Girl



Birdfall



Penguin-#4

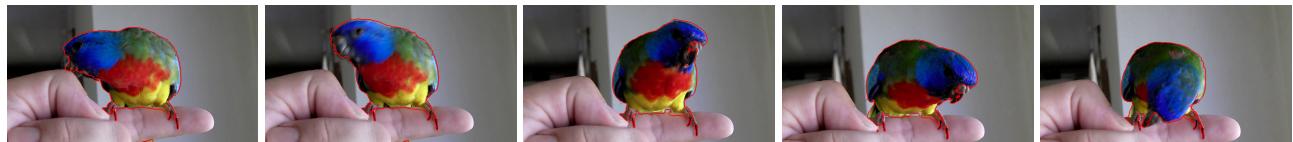


Penguin-#6

Figure 3. Example results for segmentation in five sequences on the SegTrack v2 dataset. The output on the pixel level of our multi-level model is indicated as the red contour. Best viewed in color.



aeroplane 0001



bird 0012



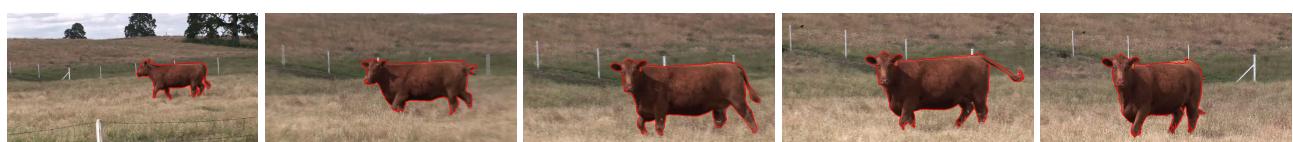
bird 0014



car 0004



cow 0011



cow 0016



boat 0007

Figure 4. Example results for segmentation in seven sequences on the Youtube-Objects dataset. The output on the pixel level of our multi-level model is indicated as the red contour. Best viewed in color.



dog 0022



dog 0028



cat 0020



train 0003



horse 0018



motorbike 0013

Figure 5. Example results for segmentation in six sequences on the Youtube-Objects dataset. The output on the pixel level of our multi-level model is indicated as the red contour. Best viewed in color.

4. Optical Flow

We present more qualitative results of updated optical flow on the SegTrack v2 dataset in Figure 6 and 7. Compared to the initial flow [7] and the other two methods [1, 8], the optical flow fields generated by the proposed algorithm have clearer object boundaries corresponding to the segmented areas.

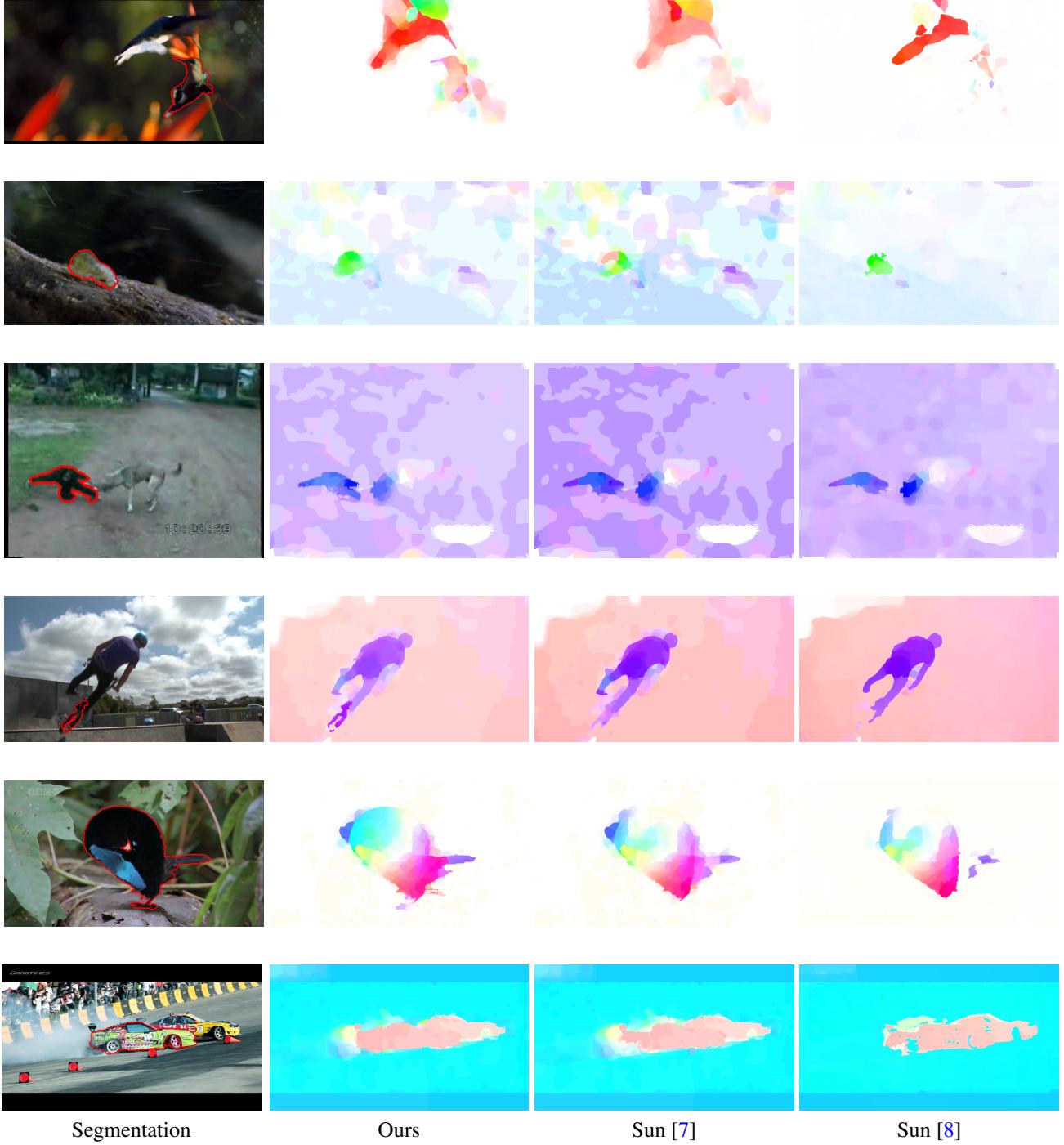


Figure 6. Results by updated optical flow on the SegTrack v2 dataset. For each sequence, we present our updated optical flow compared to the other two methods. Our results show clearer object boundaries guided by the segmented objects marked with the red contours, while the layered model [8] usually generates incomplete flows inside objects. Best viewed in color with enlarged images.

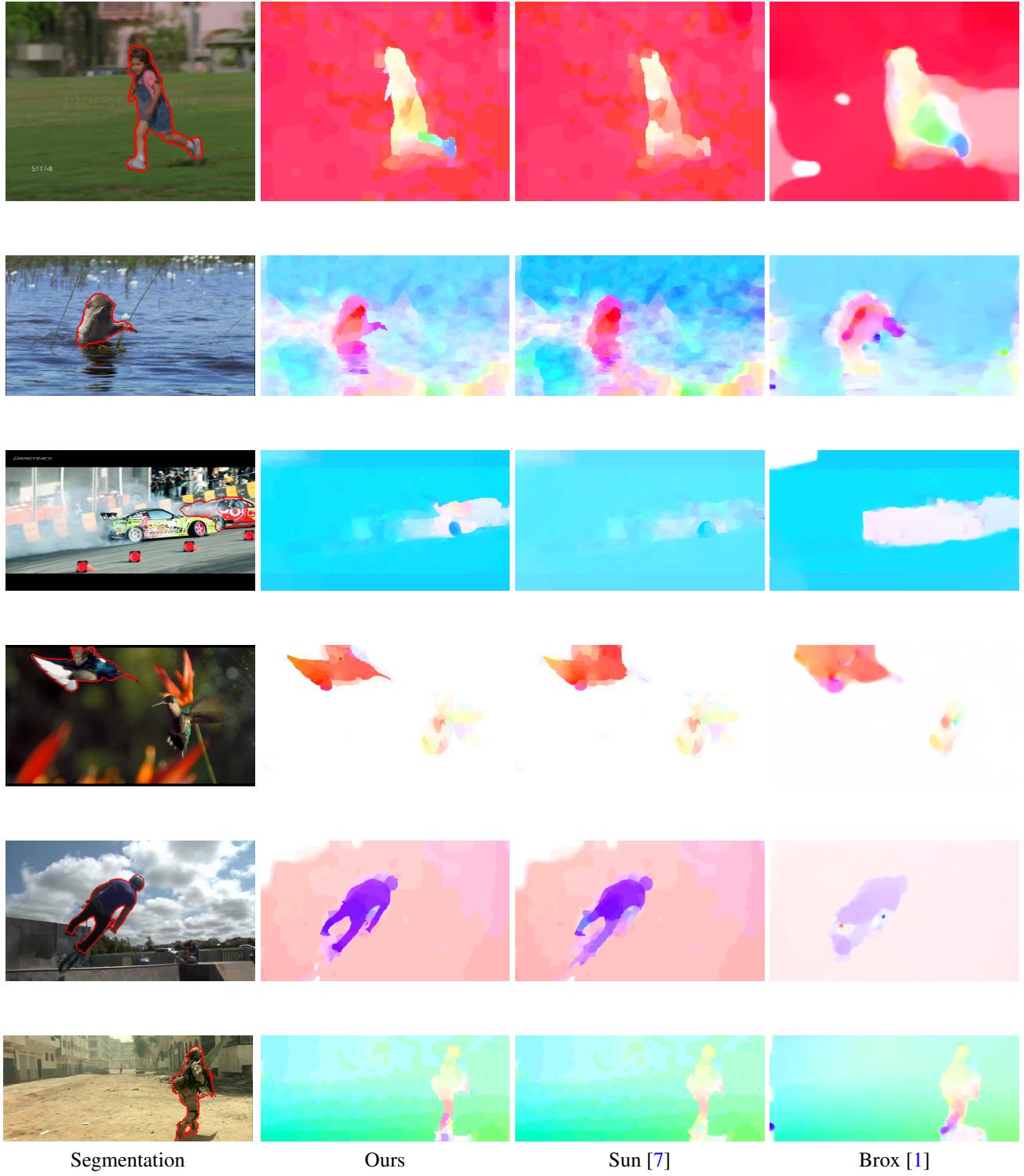


Figure 7. Results by updated optical flow on the SegTrack v2 dataset. For each sequence, we present our updated optical flow compared to the other two methods. Our results show clearer object boundaries guided by the segmented objects marked with the red contours, while the results from [1] are usually oversmoothed. Best viewed in color with enlarged images.

5. Object Flow

We first show that the updated optical flow improves segmentation accuracy on the SegTrack v2 dataset in Table 3, and then we present results of the proposed object flow algorithm in Figure 8 and 9. In each row of the figures, we show that both segmentation and optical flow results are improved after updating both models iteratively (see also Figure 6 and 7). The leftmost column is the updated optical flow, and the rightmost column is the updated segmentation result. The number indicates the overlap ratio, and we show that updated results are better than the initial segmentations in the middle column. In the figures, updated segmentations often recover parts of the object and refine the object boundary.

Table 3. Intersection-over-union ratio for updated segmentation using updated optical flow estimations on the SegTrack v2 dataset. The performance is evaluated on sequences that rely on the optical flow.

Sequence	Girl	Monkeydog	Worm	BMX	Drifting	Hummingbird	Monkey
Initial Result	86.3	45.3	78.8	19.3	73.1	70.5	88.2
Updated Result	87.6	47.3	81.7	26.3	75.5	72.3	89.3



Figure 8. Results by the proposed object flow on the SegTrack v2 dataset. For each sequence, we present the segmented object with updated optical flow compared to the initial segmentation result. We show that updated segmentations achieve better overlap ratios. Best viewed in color.

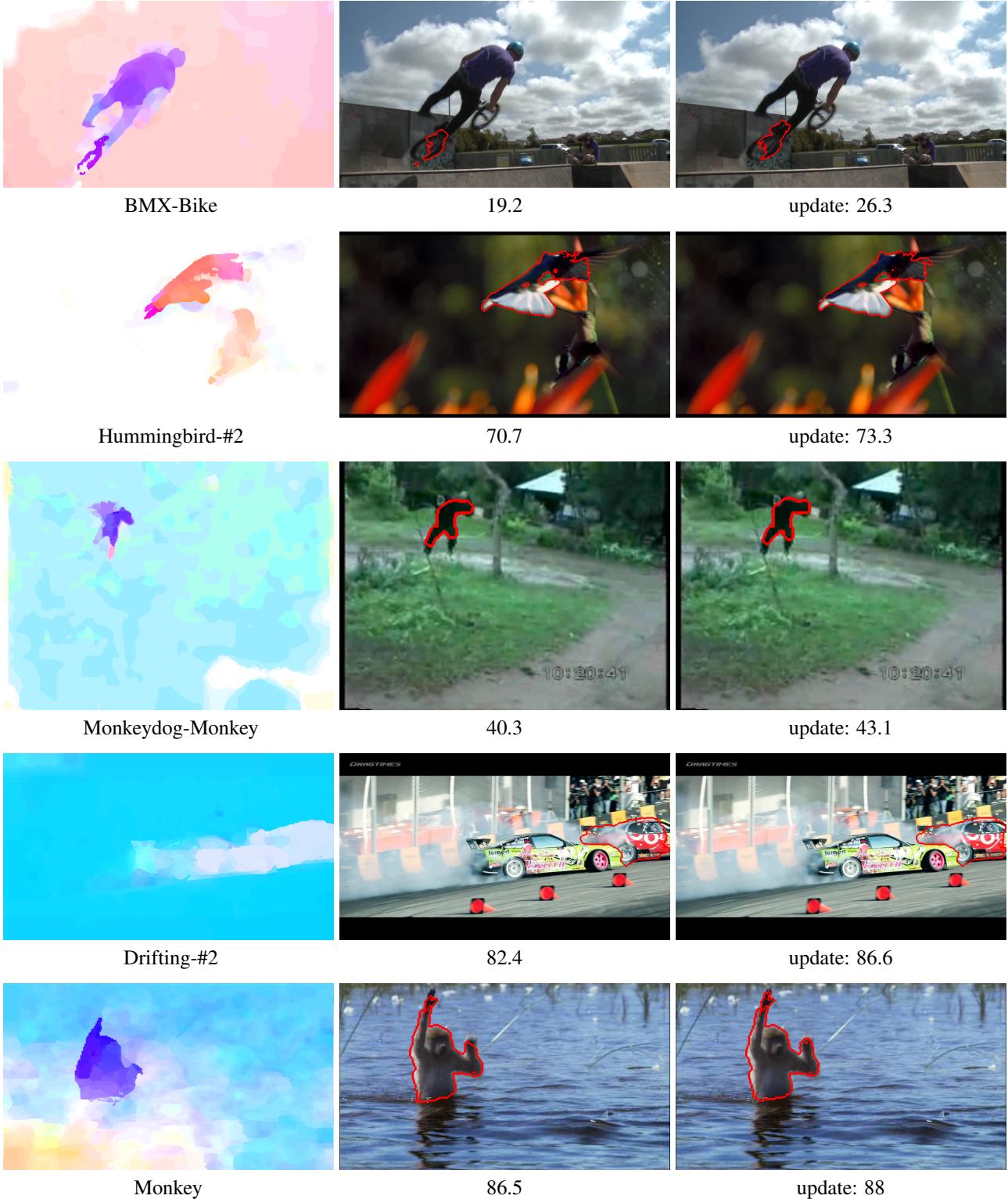


Figure 9. Results by the proposed object flow on the SegTrack v2 dataset. For each sequence, we present the segmented object with updated optical flow compared to the initial segmentation result. We show that updated segmentations achieve better overlap ratios. Best viewed in color.

References

- [1] T. Brox and J. Malik. Large displacement optical flow: descriptor matching in variational motion estimation. *PAMI*, 33(3):500–13, 2011. [7](#), [8](#)
- [2] M. Godec, P. M. Roth, and H. Bischof. Hough-based tracking of non-rigid objects. In *ICCV*, 2011. [2](#)
- [3] S. D. Jain and K. Grauman. Supervoxel-consistent foreground propagation in video. In *ECCV*, 2014. [2](#)
- [4] N. S. Nagaraja, F. Schmidt, and T. Brox. Video segmentation with just a few strokes. In *ICCV*, 2015. [2](#)
- [5] P. Ochs, J. Malik, and T. Brox. Segmentation of moving objects by long term video analysis. *PAMI*, 36(6):1187–1200, 2014. [2](#)
- [6] A. Papazoglou and V. Ferrari. Fast object segmentation in unconstrained video. In *ICCV*, 2013. [2](#)
- [7] D. Sun, S. Roth, and M. J. Black. A quantitative analysis of current practices in optical flow estimation and the principles behind them. *IJCV*, 106(2):115–137, 2014. [7](#), [8](#)
- [8] D. Sun, J. Wulff, E. B. Sudderth, H. Pfister, and M. J. Black. A fully-connected layered model of foreground and background flow. In *CVPR*, 2013. [7](#)
- [9] S. Vijayanarasimhan and K. Grauman. Active frame selection for label propagation in videos. In *ECCV*, 2012. [2](#)
- [10] L. Wen, D. Du, Z. Lei, S. Z. Li, and M.-H. Yang. Jots: Joint online tracking and segmentation. In *CVPR*, 2015. [2](#)