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.

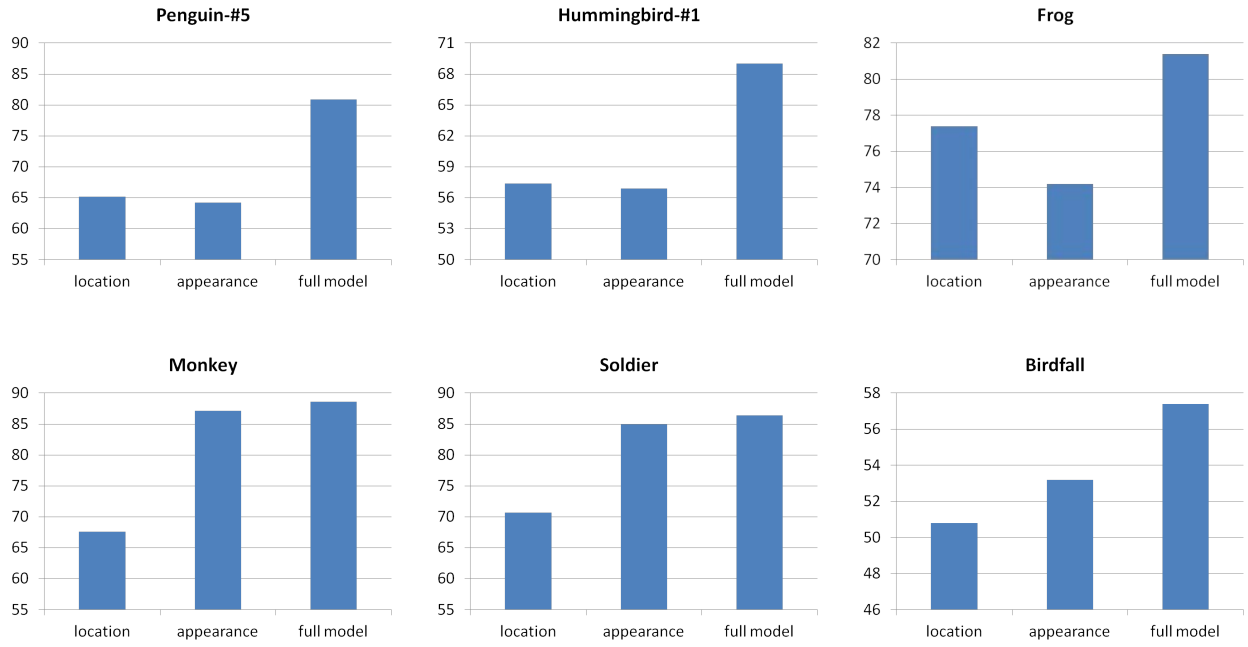


Figure 1. Model analysis with comparisons of the model only using location term, the model only using appearance term, and the full model combining both terms. The y-axis is the intersection-over-union ratio.

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.

Table 1. Segmentation results using multi-level and single-level models on the SegTrack v2 dataset with the intersection-over-union ratio.

Sequence/Object	Pixel multi	Pixel only	Superpixel multi	Superpixel only	Sequence/Object	Pixel multi	Pixel only	Superpixel multi	Superpixel only
Girl	**87.9**	86.1	74.5	60.0	Birdfall	**57.4**	48.4	47.4	51.6
Cheetah-Deer	**33.8**	30.4	22.6	3.6	Parachute	**94.5**	**94.5**	78.6	74.5
Cheetah-Cheetah	**70.4**	41.9	47.9	7.4	Monkeydog-Monkey	**54.4**	46.6	54.0	42.4
Penguin-#1	**93.9**	87.7	88.0	73.1	Monkeydog-Dog	53.3	**56.6**	39.6	2.5
Penguin-#2	**87.1**	84.9	79.9	70.0	BMX-Person	**88.0**	87.4	87.1	61.1
Penguin-#3	**89.3**	85.5	86.2	74.3	BMX-Bike	**7.0**	2.1	6.3	0.4
Penguin-#4	**88.6**	87.1	81.7	65.8	Drifting-#1	**84.3**	80.4	78.5	40.2
Penguin-#5	**80.9**	74.8	60.6	60.5	Drifting-#2	**39.0**	23.1	35.6	25.0
Penguin-#6	85.6	**88.1**	80.4	71.7	Hummingbird-#1	**69.0**	65.8	61.3	22.7
Frog	81.4	**86.9**	72.1	75.0	Hummingbird-#2	72.9	70.9	**73.3**	39.4
Worm	**89.8**	72.8	82.8	60.2	Soldier	**86.4**	85.4	70.1	62.9
Monkey	**88.6**	88.2	75.3	74.5	Bird of Paradise	95.2	**95.5**	90.7	89.0
Mean per Object	**74.1**	69.6	65.6	50.3	Mean per Sequence	**75.3**	71.2	66.0	52.8

3. Video Segmentation

SegTrack v2. We present more qualitative results on the SegTrack v2 dataset in Figure 2 and 3. For sequences such as *Hummingbird, Soldier* and *BMX-Person*, the proposed multi-level model is able to deal with non-rigid objects that undergo large deformation. For the *Penguin* and *Frog* sequences that contain objects with slow motions and similar appearance to the background, our model achieves favorable segmentation results. More comparisons for the multi-level model are presented in the video.

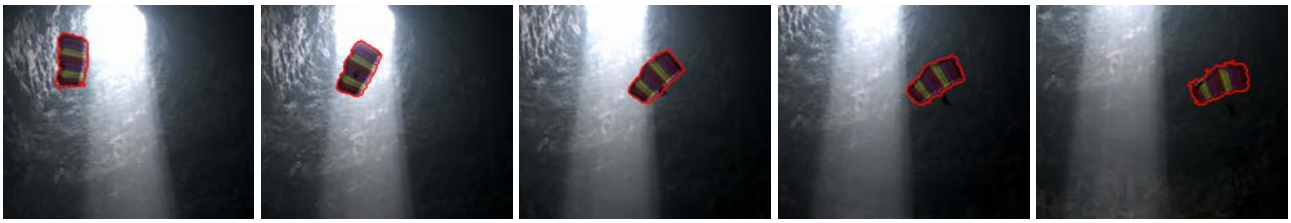
Youtube-Objects. We present more qualitative results on the Youtube-Objects dataset in Figure 4 and 5. The results show that our method is able to track and segment (multiple) objects under challenges such as occlusions (*aeroplane*), fast movements (*boat, motorbike*), deformed shapes (*dog, cow*) and cluttered backgrounds (*bird*).

In addition, we show segmentation results of the JOTS [10] along with other state-of-the-art methods in Table 2. However, this algorithm requires different parameter settings for challenging sequences, and it is not practical to evaluate on the large Youtube-Objects dataset with such assumption. Hence we use the code of [10] and fix all the parameters as the authors suggest (we also fix all the parameters and evaluate our algorithm). Note that with the fixed parameters, the JOTS fails to track objects or achieves low accuracy in 14 out of 126 sequences, which are excluded in measuring overlap ratios in Table 2.

Table 2. Segmentation results of [10] on the Youtube-Objects dataset with the intersection-over-union ratio.

Category	[4]	[3]	[9]	[2]	[6]	[5]	JOTS [10]	Ours
aeroplane	89.0	86.3	79.9	73.6	70.9	13.7	78.4	**89.9**
bird	81.6	81.0	78.4	56.1	70.6	12.2	57.2	**84.2**
boat	**74.2**	68.6	60.1	57.8	42.5	10.8	51.9	74.0
car	70.9	69.4	64.4	33.9	65.2	23.7	60.2	**80.9**
cat	67.7	58.9	50.4	30.5	52.1	18.6	57.5	**68.3**
cow	79.1	68.6	65.7	41.8	44.5	16.3	47.7	**79.8**
dog	70.3	61.8	54.2	36.8	65.3	18.0	42.7	**76.6**
horse	67.8	54.0	50.8	44.3	53.5	11.5	43.9	**72.6**
motorbike	61.5	60.9	58.3	48.9	44.2	10.6	32.5	**73.7**
train	**78.2**	66.3	62.4	39.2	29.6	19.6	43.8	76.3
Mean	74.0	67.6	62.5	46.3	53.8	15.5	51.6	**77.6**

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.

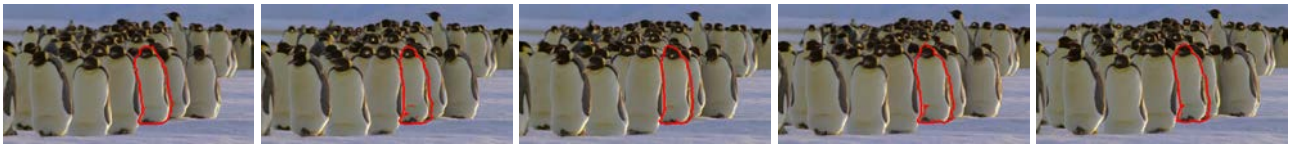


Parachute

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.

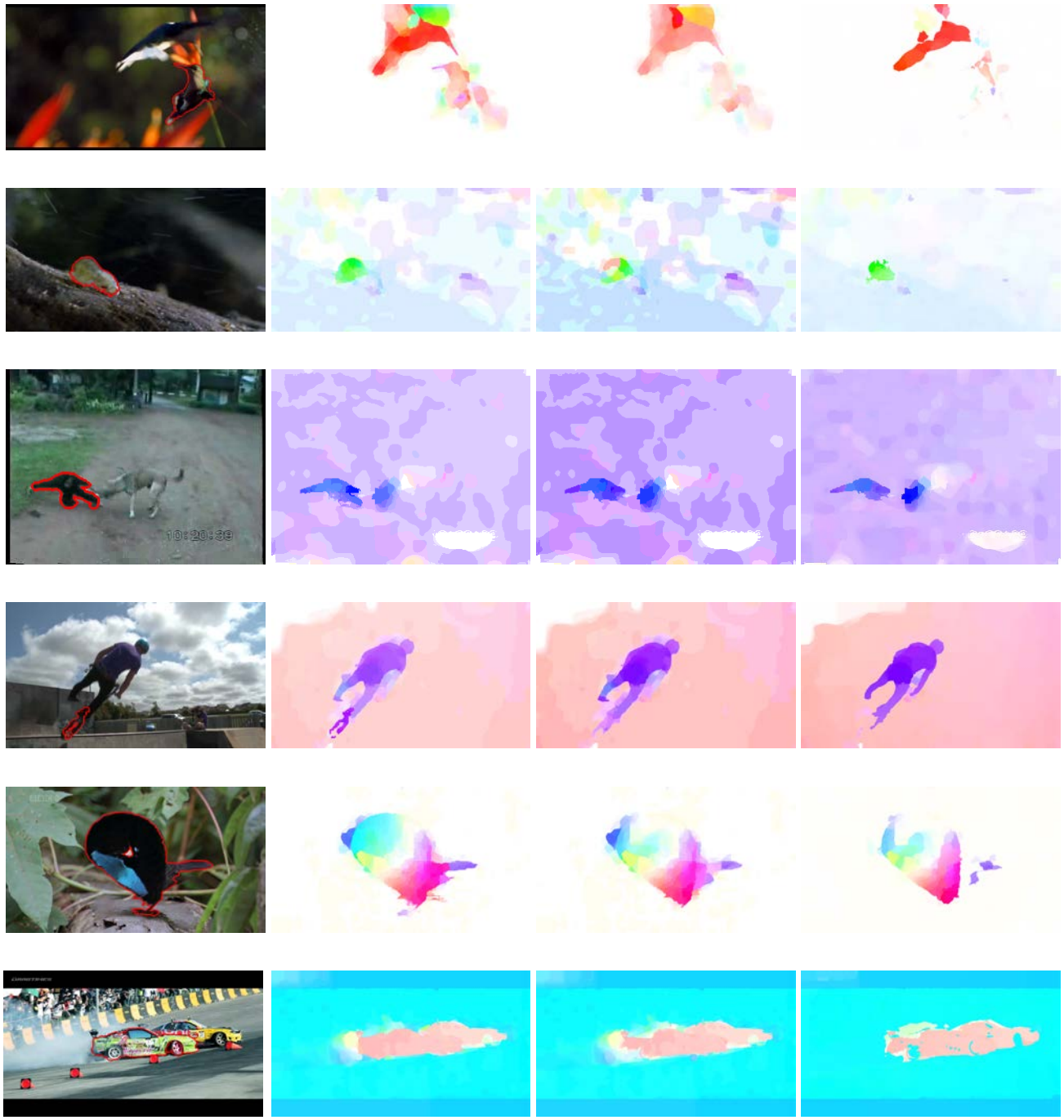


Figure 6. Results by updated optical flow on the SegTrack v2 dataset. For each sequence, we present our updated optical flow compared to the other two methods. Our results show clearer object boundaries guided by the segmented objects marked with the red contours, while the layered model [8] usually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

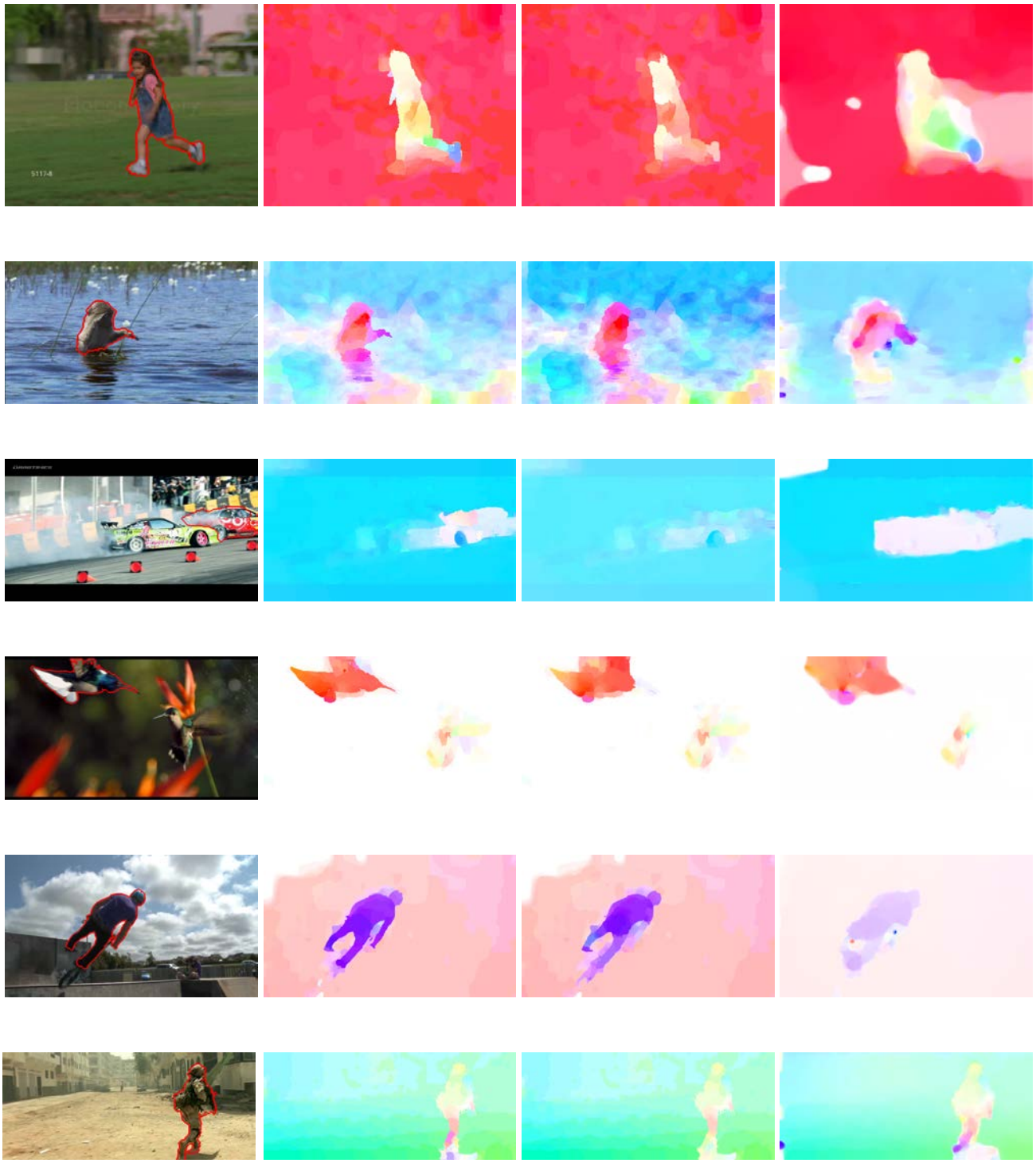


Figure 7. Results by updated optical flow on the SegTrack v2 dataset. For each sequence, we present our updated optical flow compared to the other two methods. Our results show clearer object boundaries guided by the segmented objects marked with the red contours, while the results from [1] are usually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

Drifting-#1 78.3 update: 83.6

Girl 82.5 update: 88.3

Worm 68.1 update: 74.3

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.

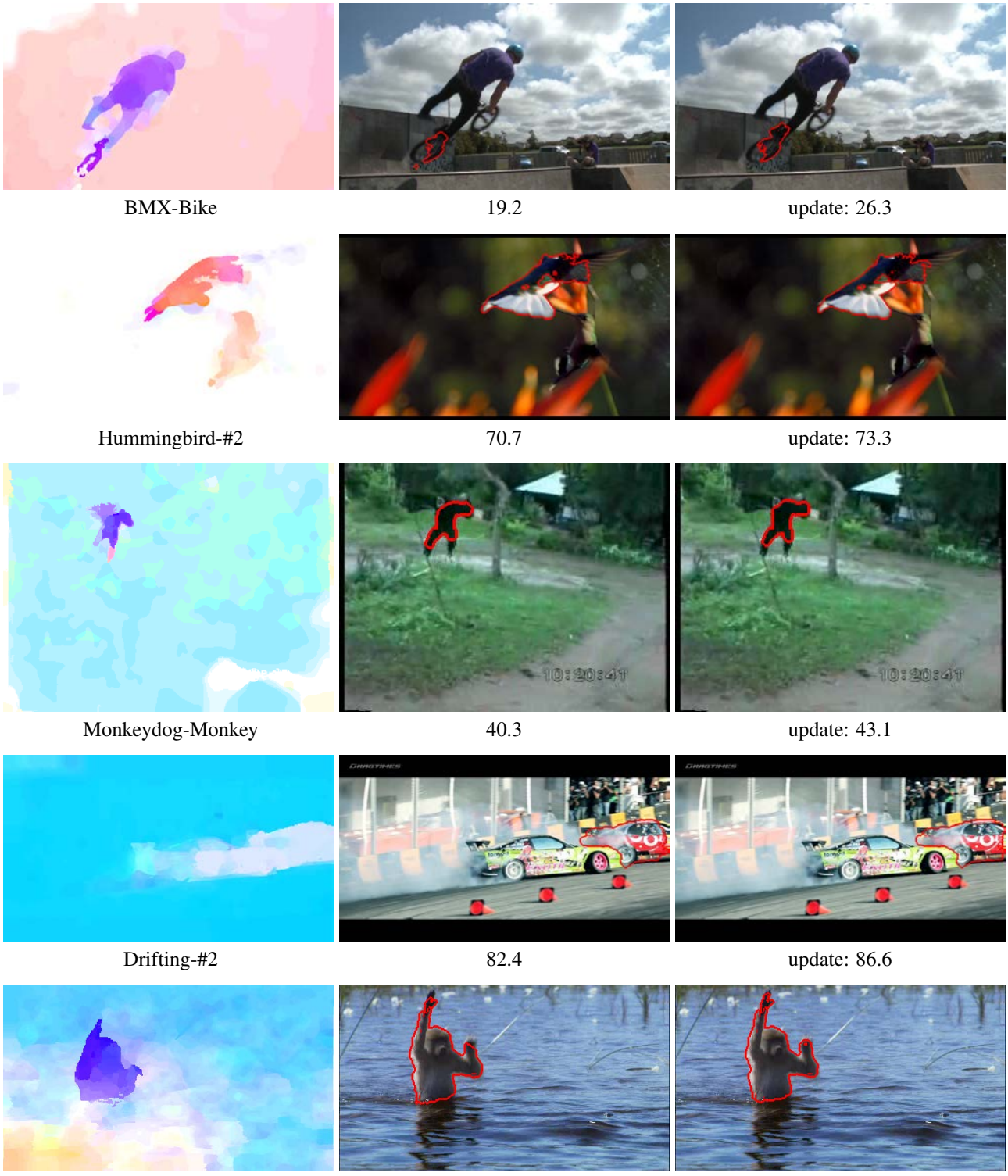


Figure 9. Results by the proposed object flow on the SegTrack v2 dataset. For each sequence, we present the segmented object with updated optical flow compared to the initial segmentation result. We show that updated segmentations achieve better overlap ratios. Best viewed in color.

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