Spatio-Temporal Video Segmentation of Static Scenes and Its Applications

Supplementary Document

Hanqing Jiang, *Student Member, IEEE,* Guofeng Zhang*, *Member, IEEE,*
Huiyan Wang, *Member, IEEE,* and Hujun Bao, *Member, IEEE*

I. Influence Of Parameter Changing On Segmentation Results

Most of the parameters such as some thresholds are tuned through many experimental tests and can work well for all the experiments. The only parameters we have to tune are the mean shift parameters which can control the granularity of segmentation as shown in Fig. 1 of our manuscript. Besides, we have experimented the "Road" example with the most important parameters w_c, w_d and w_s changed to 0.9 respectively in the iterative optimization stage, as shown in Fig. 1. Since w_c, w_d and w_s control the Gaussian models of color, disparity and spatial distribution respectively without temporal statistics, they should not be set too large. Otherwise, the generated segmentation results cannot preserve consistent boundaries well, and will contain many segmentation noises, as shown in Fig. 1(b-d).

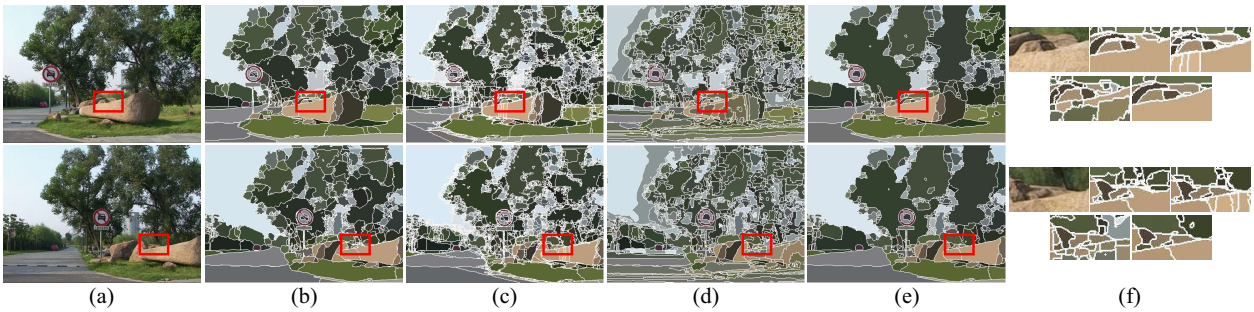

Fig. 1. Segmentation results of "Road" sequence with different parameter settings for iterative optimization. (a) Two selected original frames. (b) The segmentation results with $w_c = 0.9$, $w_h = 0.05$, $w_d = 0.01$ and $w_s = 0.04$. (c) The results with $w_c = 0.04$, $w_h = 0.05$, $w_d = 0.01$ and $w_s = 0.9$. (d) The results with $w_c = 0.04$, $w_h = 0.05$, $w_d = 0.9$ and $w_s = 0.01$. (e) The results with parameter setting in our manuscript. (f) The magnified regions of the red rectangles in (a-e).

(a) (b) (c) (d)

Fig. 2. Comparison with motion-based spatio-temporal segmentation on "Road" sequence. (a) Three selected original frames. (b) The motion-based segmentation results. (c) Our depth-based segmentation results. (d) The magnified regions of the red rectangles in (a-c).

II. Comparison With Motion-Based Spatio-Temporal Segmentation Results

Although temporal correspondences and statistics can also be collected using motion information across consecutive frames, the results of motion-based spatio-temporal segmentation are unsatisfactory, as shown in Fig. 2. We experiment the "Road" example using optical flows [1]. Compared to our depth-based segmentation (Fig. 2(c)), the motion-based segmentation results are inconsistent in both spatial and temporal domains, with a variety of segmentation errors along object boundaries, as shown in Fig. 2(b). The projection overlapping rate of the motion-based segmentation results is 77.39%, which indicates a lower temporal stability compared to the overlapping rate of our depth-based segmentation (92.43%). The reason is that motion information is unreliable especially along the depth-discontinuous boundaries due to occlusions, and can therefore only be tracked within no more than 5 consecutive neighboring frames, which are not enough for collecting robust temporal correspondences and statistics to ensure temporal consistency.

III. Segmentation Results of Forward Camera Motion

Our spatio-temporal segmentation can handle kinds of camera motion. Fig 3 shows the segmentation results of a sequence where the camera moves forward. As can be seen, the segmented volume segments

Fig. 3. The segmentation results of “Forward” sequence. Top: selected frames. Bottom:: the segmentation results.

are very temporally consistent among different frames, which demonstrates the robustness of the proposed method.

IV. Segmentation Results of Challenging Wide-baseline Images

Given an extremely small number of wide-baseline images, the collected statistics may be degraded and handling the problems of large occlusions or out-of-view will become more difficult, which may cause our method to produce unsatisfactory segmentation results. Fig. 4 shows a challenging example [1], which only contains 3 wide-baseline images. Fig. 4(b-f) show the depth maps, mean-shift segmentation results, computed probabilistic boundary maps, the spatial segmentation results, and our final spatio-temporal segmentation results, respectively. Even in this extreme case, our segmentation results still preserve most accurate and temporally consistent object boundaries, except for the individual regions as highlighted by yellow rectangles in Fig. 4(f). As shown in Fig. 4(e), some regions in different frames correspond to the same object but could not be matched due to low projection overlapping rate caused by the problem of out-of-view. It eventually results in unsatisfactory segmentation results, as shown in Fig. 4(l): a unified region are separated into different regions, however, which have rather similar colors and depths, as shown in Fig. 4(g-h).

[1]This dataset is downloaded from the Middlebury stereo evaluation website [2]:
http://vision.middlebury.edu/stereo/data/

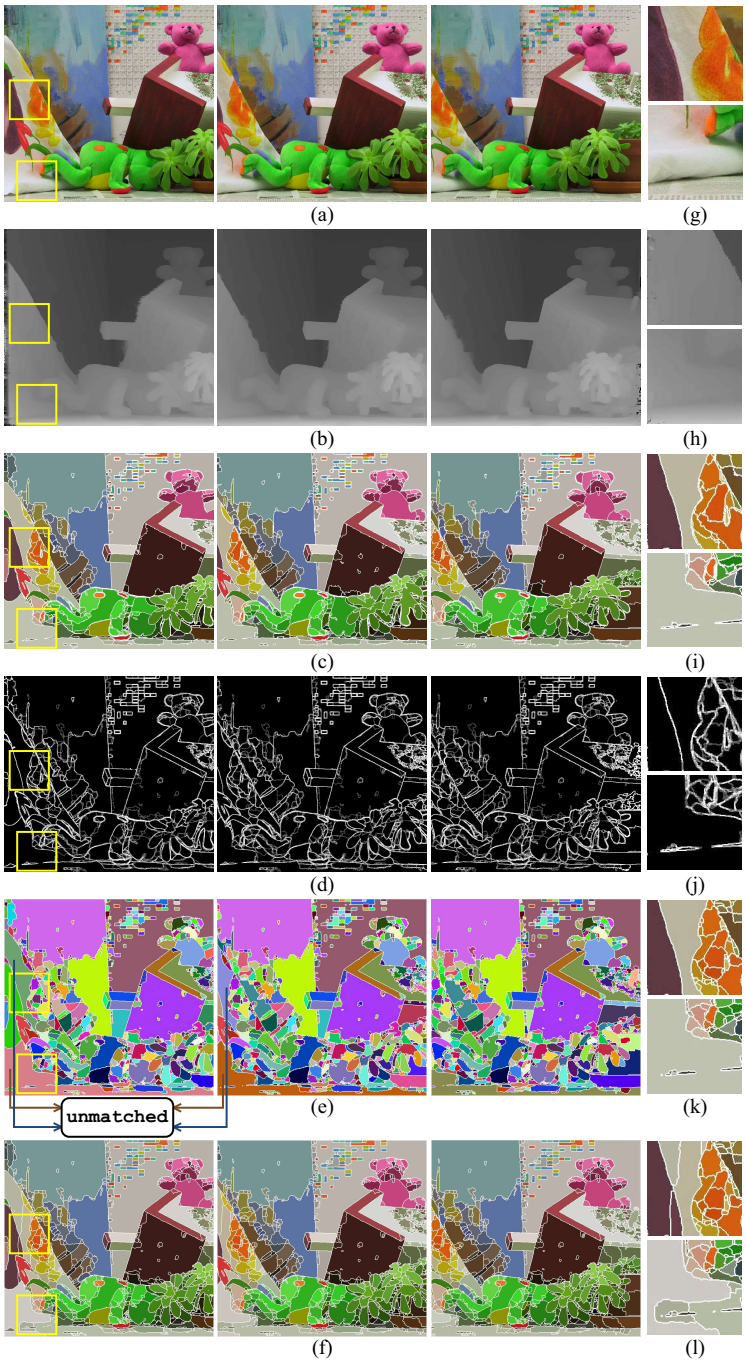

Fig. 4. The segmentation results of "Teddy" sequence. (a) The whole sequence. (b) The depth maps. (c) The mean shift segmentation results. (d) The computed probabilistic boundary maps. (e) Our spatial segmentation results. After segment matching, the matched segments are represented with unique color. (f) Our final spatio-temporal segmentation results. (g-l) The magnified regions of (a-f).

REFERENCES

[1] C. Liu, “Beyond pixels: exploring new representations and applications for motion analysis,” Ph.D. dissertation, Massachusetts Institute of Technology, May 2009.

[2] D. Scharstein and R. Szeliski, “High-accuracy stereo depth maps using structured light,” in *CVPR*, vol. 1, 2003, pp. 195–202.