Video Segmentation by Non-Local Consensus Voting

Alon Faktor
http://www.wisdom.weizmann.ac.il/~alonf/
Michal Irani
http://www.wisdom.weizmann.ac.il/~irani/

Dept. of Computer Science and Applied Math
The Weizmann Institute of Science
ISRAEL

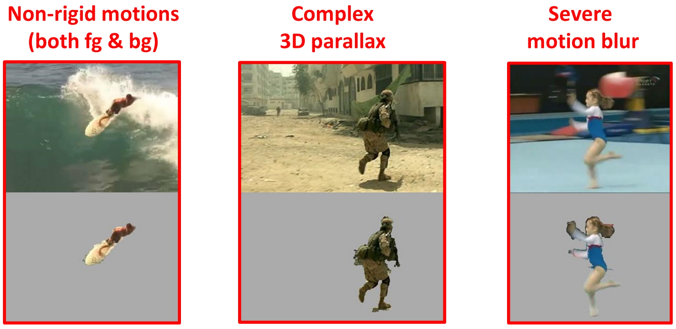

Figure 1: **A unified approach to foreground/background video segmentation in *unconstrained* videos.** Our algorithm can handle in a single framework video sequences which contain highly non-rigid foreground and background motions, complex 3D parallax as well as simple 2D motions and severe motion blur.

We address the problem of Foreground/Background (fg/bg) segmentation of "unconstrained" video. By "unconstrained" we mean that the moving objects and the background scene may be highly non-rigid (e.g., waves in the sea); the camera may undergo a complex motion with 3D parallax; moving objects may suffer from motion blur, large scale and illumination changes, etc. Fig. 1 shows a few such examples. Most existing segmentation methods fail on such unconstrained videos, especially in the presence of highly non-rigid motion and low resolution. Unconstrained video has thus become the focus of most recent video segmentation methods [5, 6, 9, 13].

In this paper, we suggest a simple yet general algorithm for performing fg/bg video segmentation, which handles complex unconstrained videos. We cast the video segmentation problem as a voting scheme on the graph of similar ("re-occurring") regions in the video sequence. 'Re-occurring' regions can be quite far both in space and in time, but are constrained to be close in the appearance feature space. We start from crude saliency votes at each pixel, and iteratively correct those votes by "consensus voting" of re-occurring regions across the video sequence. **The power of our consensus voting comes from the *non-locality* of the region re-occurrence, both in space and in time – enabling fast propagation of diverse and rich information across the entire video sequence.** This enables the correction of large errors in the initial fg/bg votes.

In contrast to trajectory-based methods [1, 2, 3, 4, 7, 8, 10, 11], we do not try to explicitly estimate long-term correspondences via flow estimation or tracking, but rather obtain long-term "probabilistic" correspondences using re-occurring regions across distant frames. This avoids the inherent uncertainties of explicit optical flow estimation, whose errors tend to accumulate over time. Similarly, MRF-based video segmentation methods [5, 6, 9, 13] tend to propagate information only *locally* in space-time. Their temporal links are based on optical-flow, whose rapidly accumulated errors induce weak (often zero) weights between related parts in faraway frames. The segmentation performance of video-MRF methods thus strongly depends on the quality of their initial fg/bg data term. However, fg/bg initializations tend to be very noisy, whether based on mining moving object proposals [5, 6, 13], or based on motion saliency maps [9] (especially in unconstrained low-quality videos). Therefore, current video segmentation methods encounter difficulties in such challenging videos. In contrast, our *non-local* consensus voting allows us to start with very 'noisy' fg/bg votes, and clean them rapidly according to 'consensus voting' of distant re-occurring regions.

Qualitative and quantitative experiments indicate that our approach outperforms current state-of-the-art methods. Some visual examples can be found in Fig. 2. **Full videos can be found on our project website www.wisdom.weizmann.ac.il/~vision/NonLocalVideoSegmentation.html.** Empirical comparisons on the SegTrack Dataset [12] can be found in the paper.

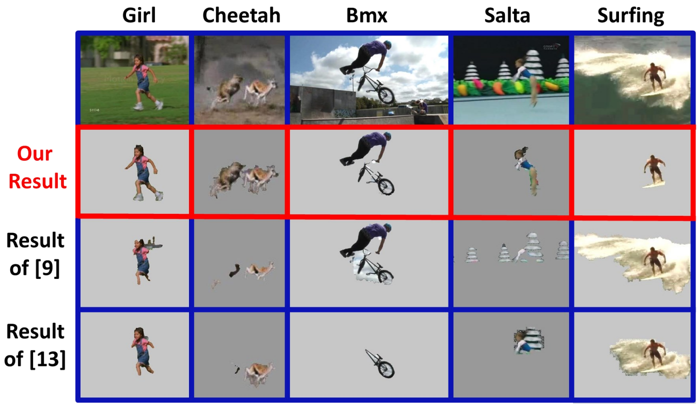

Figure 2: **Visual comparison of results.** Visual comparisons to [9, 13] using their publicly available code. The 3 first sequences are from the SegTrack dataset and the rest are new challenging sequences. For 'Bmx' and 'Salta', we show results of [13] using object selection without Grab-Cut (whereas for all other sequences with Grab-Cut), since these settings gave best results for [13]. See full videos on our Project Website (link in the text).

[1] T. Brox and J. Malik. Object segmentation by long term analysis of point trajectories. In *ECCV*, 2010.

[2] J. Costeira and T. Kanande. A multi-body factorization method for motion analysis. In *ICCV*, 1995.

[3] E. Elhamifar and R. Vidal. Sparse subspace clustering. In *CVPR*, 2009.

[4] K. Fragkiadaki and J. Shi. Video segmentation by tracing discontinuities in a trajectory embedding. In *CVPR*, 2012.

[5] Y. J. Lee, J. Kim, and K. Grauman. Key-segments for video object segmentation. In *ICCV*, 2011.

[6] T. Ma and L. J. Latecki. Maximum weight cliques with mutex constraints for video object segmentation. In *CVPR*, 2012.

[7] P. Ochs and T. Brox. Object segmentation in video: A hierarchical variational approach for turning point trajectories into dense regions. In *ICCV*, 2011.

[8] P. Ochs and T. Brox. Higher order motion models and spectral clustering. In *CVPR*, 2012.

[9] A. Papazoglou and V. Ferrari. Fast object segmentation in unconstrained video. In *ICCV*, 2013.

[10] S. R. Rao, R. Tron, R.Vidal, and Y. Ma. Motion segmentation via robust subspace separation in the presence of outlying, incomplete, or corrupted trajectories. In *CVPR*, 2008.

[11] J. Shi and J.Malik. Motion segmentation and tracking using normalized cuts. In *ICCV*, 1998.

[12] D. Tsai, M. Flagg, and J. Rehg. Motion coherent tracking with multi-label mrf optimization. In *BMVC*, 2010.

[13] Dong Zhang, Omar Javed, and Mubarak Shah. Video object segmentation through spatially accurate and temporally dense extraction of primary object regions. In *CVPR*, 2013.