Learning to Segment Moving Objects in Videos

Katerina Fragkiadaki, Pablo Arbelaez, Panna Felsen, Jitendra Malik

Berkeley UNIVERSITY OF CALIFORNIA

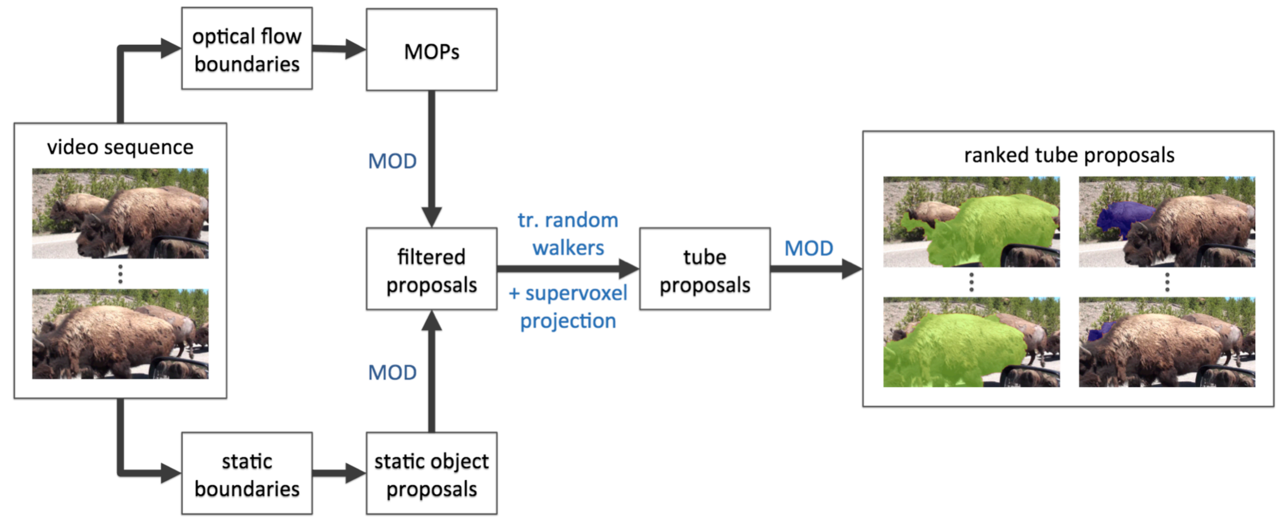
IEEE 2015 Conference on Computer Vision and Pattern Recognition

CVPR 2015 BOSTON JUNE 7-12

Summary

We present a learning-based approach for motion segmentation. Our method:

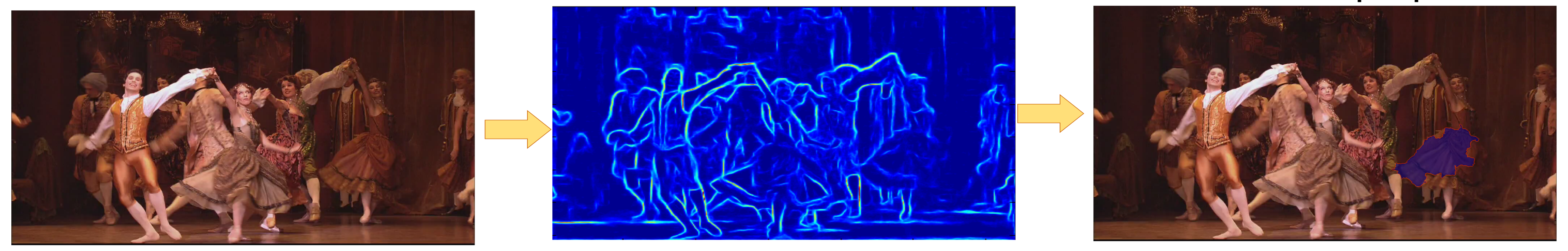
- computes per frame segment proposals from flow and image boundaries,
- ranks them with a multilayer ``moving objectsness" detector (MOD), and
- extends per frame segments into space–time tubes using random walkers on dense point trajectory motion affinities.



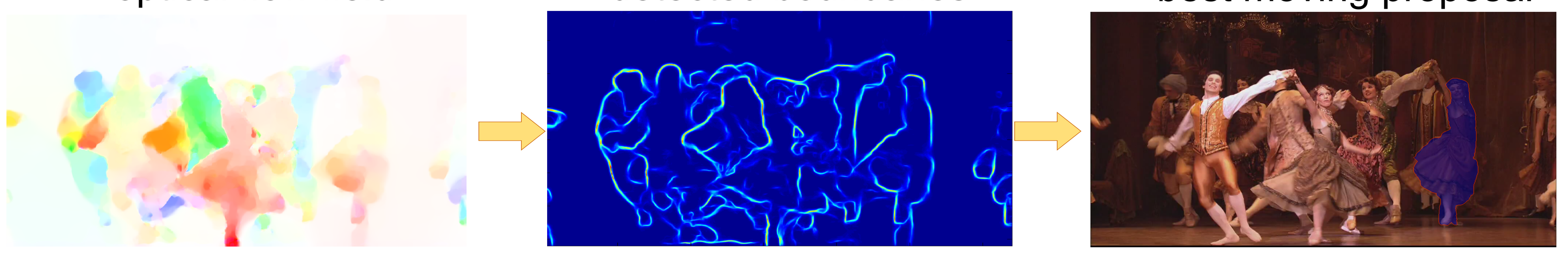
We achieve higher intersection over union with ground-truth tubes than competing approaches on Moseg and VSB100 video benchmarks, under any number of proposals.

Bottom-up Moving Object Proposals

In cluttered scenes, image boundary detectors fire at internal (fake) boundaries. Segmentations then tends to over-fragment the objects.



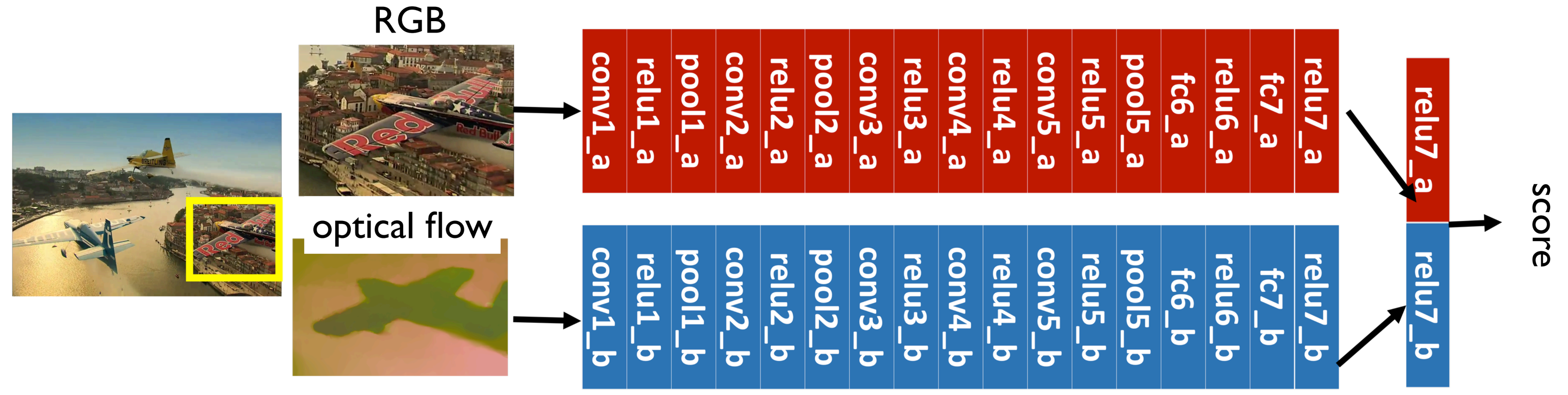
Motion, in contrast to appearance, is smooth at object interiors. Boundaries detected on the optical flow field suffer less from internal clutter.



We compute boundaries using the structured forest boundary detector of Dollar *et al* [1]. Given the boundaries, we compute segment proposals using multiple figure-ground segmentations of Krahenbrul and Koltun [2].

Moving Objectness Detector

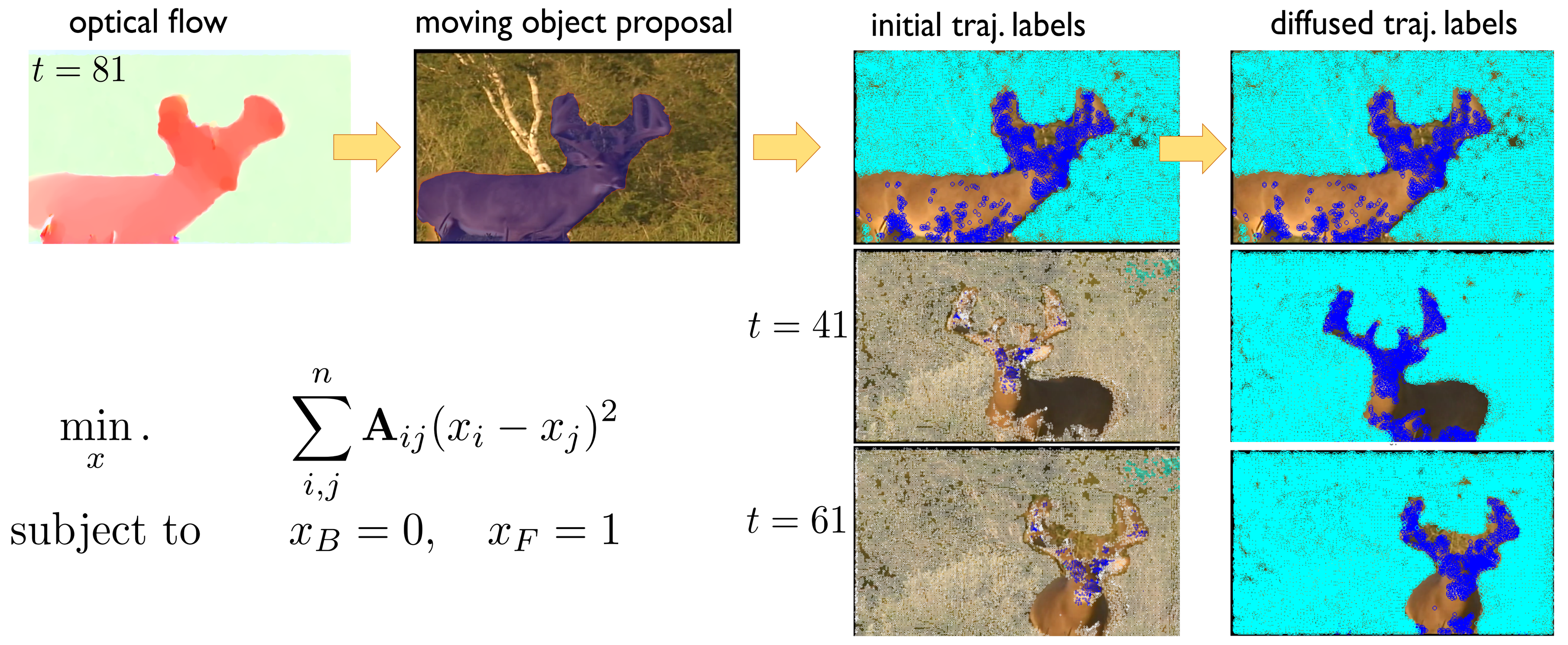
We train a moving objectness detector from optical flow and RGB to regress to the intersection over union score of the input bounding box of the region proposal.



Our detector learns to suppress:

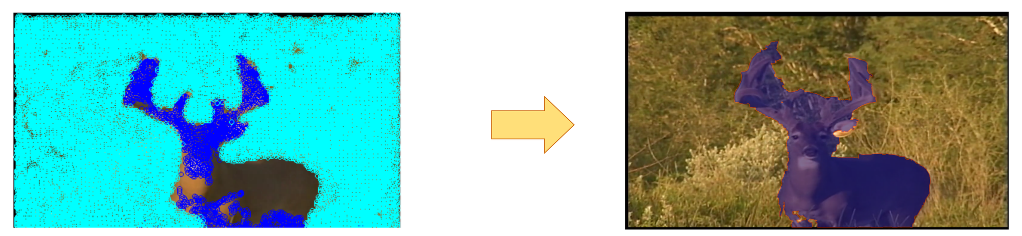
From per frame segments to space-time tubes

We map high-scoring per frame segments to space time tubes using random walkers on point trajectory motion affinities.



$$\min_{x} \ \sum_{i,j}^{n} \mathbf{A}_{ij}(x_i - x_j)^2$$
$$\text{subject to} \quad x_B = 0, \quad x_F = 1$$

We map trajectory clusters to pixel tubes using a weighted sum of supervoxels.

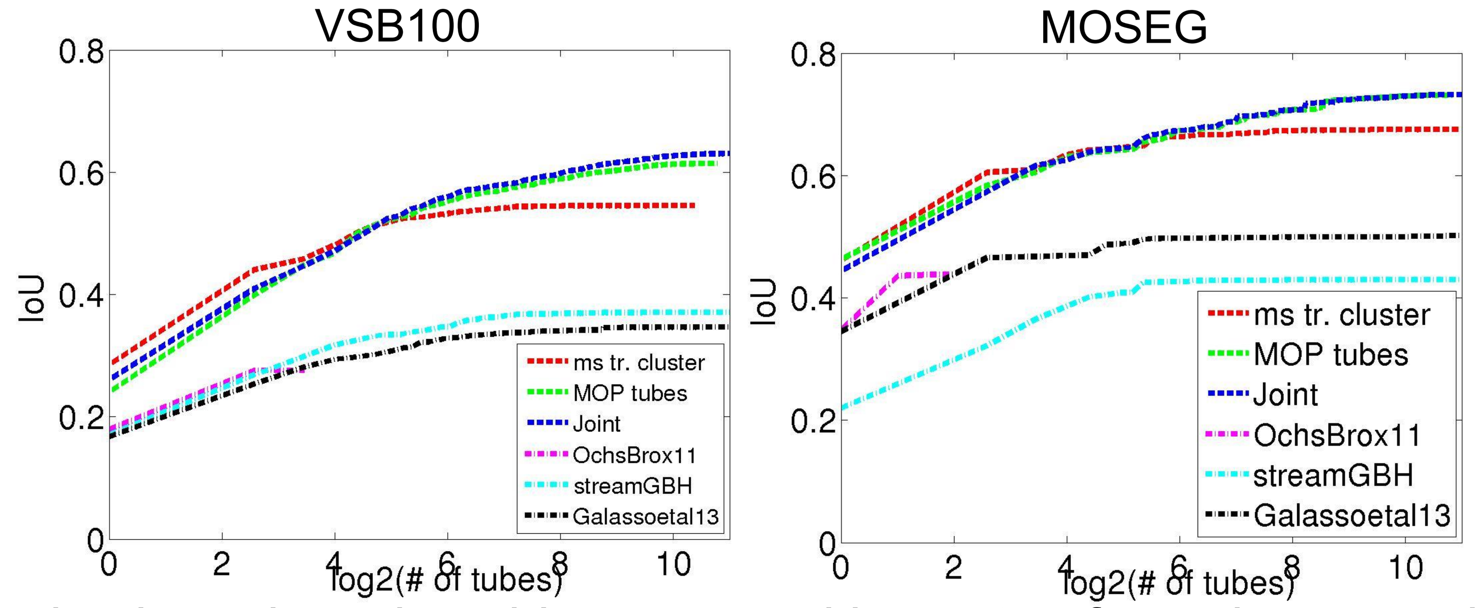


Experiments

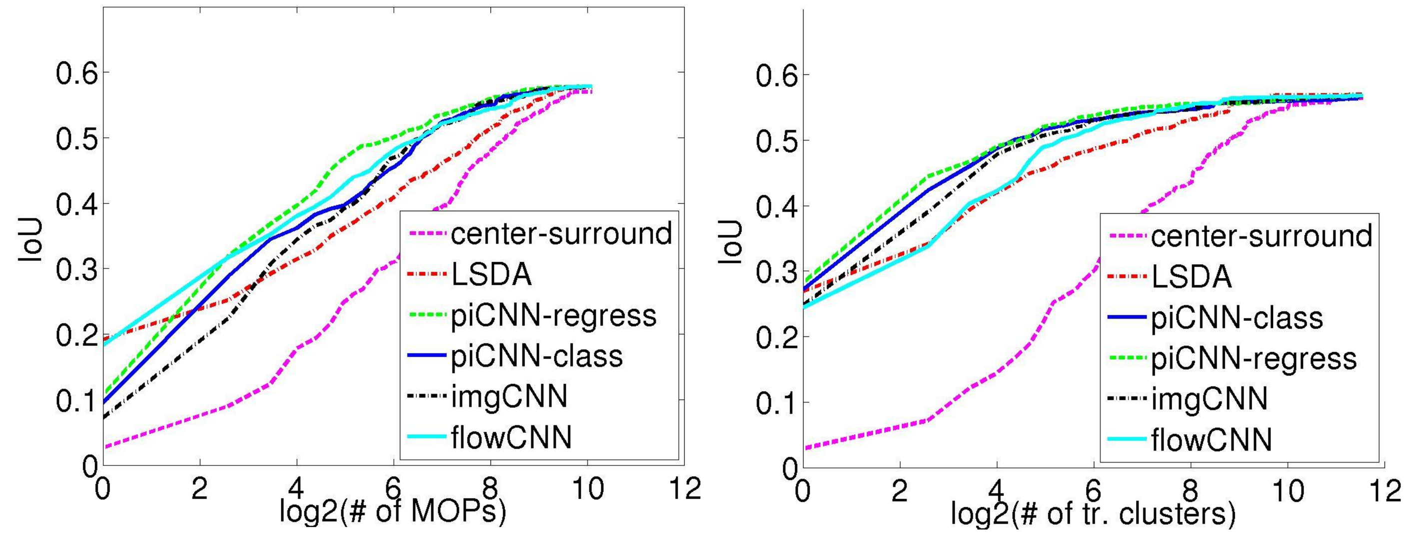
Per frame segmentation. Moving object proposals (MOPs) complement static object proposals (GOPs). The combined performance goes beyond the saturation level of each method.

		avg best ol	coverage	det 50%	det 70%	avg best ol ab	det 50% ab	det 70% ab
VSB 100	GOP (2715)	53.74	66.84	60.34	26.12	65.08	82.6	48.08
	MOP (873)	46.47	61.3	47.25	13.85	57.92	73.75	29.79
	GOP+MOP (2659=1786+873)	**56.17**	**69.85**	**66.48**	**31.50**	**67.15**	**86.14**	**51.92**
MOSEG	GOP (2500)	68.47	76.56	87.59	64.54	74.72	91.94	79.03
	MOP(839)	57.74	68.49	70.57	37.94	66.42	83.87	59.68
	GOP+MOP (2512=1673+839)	**69.65**	**78.29**	87.59	**70.21**	**75.38**	91.94	**83.87**

Motion Segmentation. We compare our method with multiscale trajectory clustering and video segmentation methods in the literature. We outperform previous methods by a margin at any number of proposals.



Ranking. Learning based moving objectness ranking outperforms by a margin hand-designed center-surround motion saliency. Combining image and optical flow gives better performance than using each one in isolation.



References

1. Structured forests for fast edge detection, P. Dollar, C. L. Zitnick, In *ICCV* 2013
2. Geodesic Object proposals, P. Krahenbrul, V. Koltun, In *CVPR* 2014
3. A unified video segmentation benchmark: Annotation, metrics and analysis. F. Galasso, N. S. Nagaraja, T. J. Cardenas, T. Brox, and B. Schiele, In *ICCV*, 2013
4. Evaluation of super-voxel methods for early video processing. C.Xuand, J.J.Corso, In *CVPR*, 2012.