Learning to Segment Moving Objects in Videos
Katerina Fragkiadaki, Pablo Arbelaez, Panna Felsen, Jitendra Malik

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 objectness" 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 tend to over-segment 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:
• static background
• under-segmentations
• over-segmentations

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.

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.

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