Segmentation of Dynamic Scenes with Distributions of Spatiotemporally Oriented Energies

Damien TENEY
Matthew BROWN
University of Bath, UK

Motivation

Video segmentation: disambiguate appearance with motion cues

Identification of motion non-trivial

Camera motion, non-rigid objects, dynamic textures (smoke, water, fire, …)

Usual approach: ~~optical flow~~ + parametric ~~motion models~~

Restrictions: brightness constancy, rigidly moving objects, …

Comptationally expensive

Unnecessary intermediate goal ?

This work: low-level motion features w/ existing video segm. framework

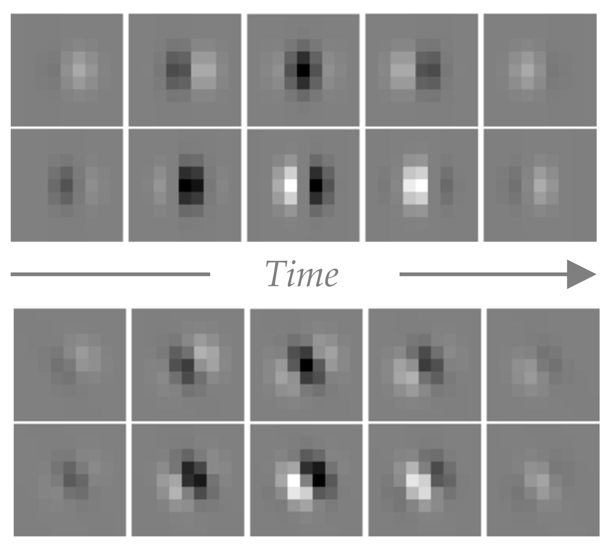
- → Capture wide range of image dynamics: non-rigid motion, brightness changes, flickering effects, …
- → Model-free, unsupervised
- → Convolution-based features: inexpensive nowadays with GPUs

Can filter-based motion features compete w/ optical flow ?

Steerable 3D spatiotemporal filters

Like 2D filters identify oriented structures (edges) in 2D images
3D filters are applied on the video volume of stacked frames

Steered in 3D to particular orientations / velocities



Vertical pattern moving at 0.5 px/frame

Gaussian 2nd derivative
(responds to lines)

and its Hilbert transform
(transfer fct. phase shifted by 90°: responds to edges)

Quadrature pair, for phase-independent response

Other example orientation:
oblique pattern moving up/rightward

Banks of filters and histograms of motion energies

Convolution of video volume with pairs of quadrature filters [1,2]

$$E_{\hat{\theta}}(x,y,t) \;=\; (G2_{\hat{\theta}} * \mathcal{V})^2 + (H2_{\hat{\theta}} * \mathcal{V})^2$$

Phase-independent energy measure

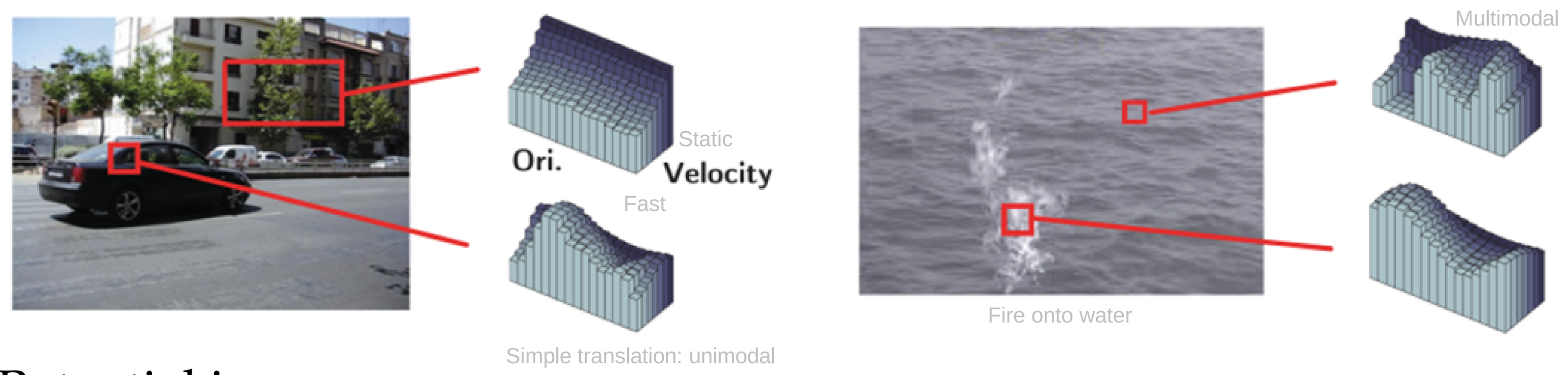
Filter space-time orientation — Video volume

Aggregate responses of filters consistent w/ same direction of motion

$$ME_{\hat{n}}(x,y,t) \;=\; \sum_{i=0}^{N} E_{\hat{\theta}_i}(x,y,t)$$

« Motion energies » [1]
Maginalization over appearance: individual filters only captured normal flow wrt. local orientation

Build histogram for a number of motion orientations / velocities



Potential issues

Sensitivity to contrast

$$ME'_{\hat{n}}(x,y,t) \;=\; ME_{\hat{n}}(x,y,t) \;/\; \max_{\hat{n}} ME_{\hat{n}}(x,y,t)$$

Normalize wrt. strongest local orientation

Correlations at nearby orientations

$$ME''_{\hat{n}}(x,y,t) \;=\; e^{\alpha(ME'_{\hat{n}}(x,y,t)-1)}$$

Emphasize peaks

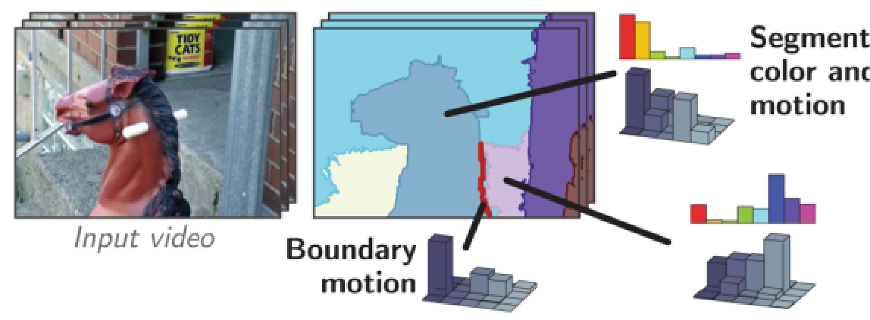
Segmentation framework

1 Graph-based segmentation[3], regions described by color + motion histograms

2 Assign each boundary to either of its adjacent segments = depth ordering

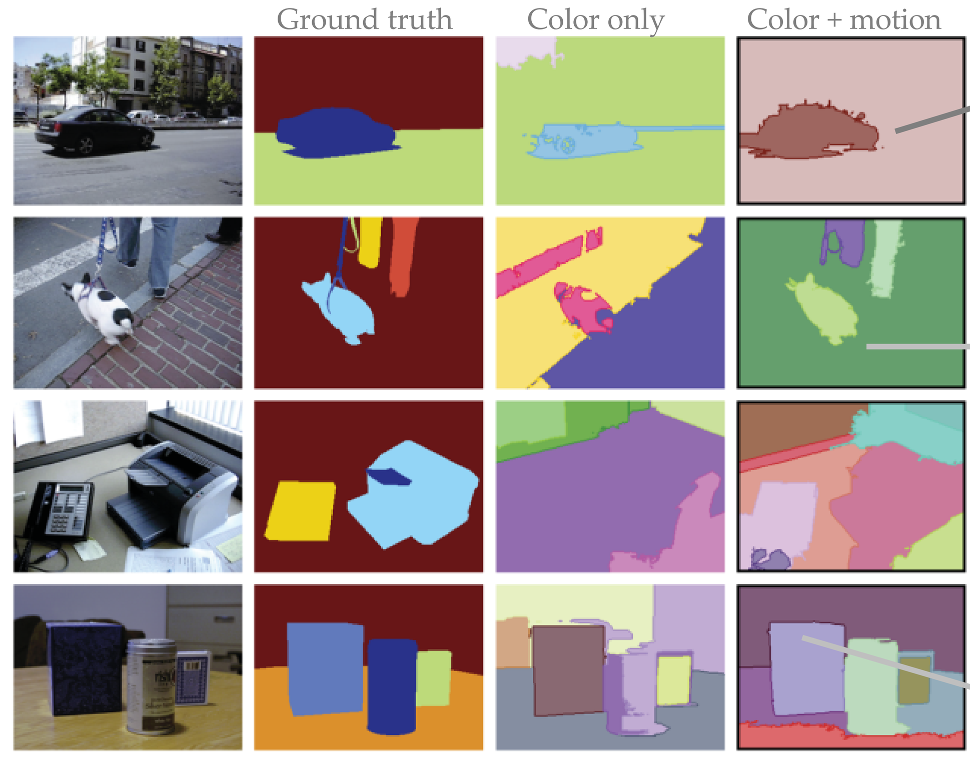
Intuition: *occlusion boundaries move together with the occluding segment*

→ Build similar motion histograms for boundaries
then assign boundary to the most similar of its two adjacent segments



Experiments

Motion segmentation, MIT dataset

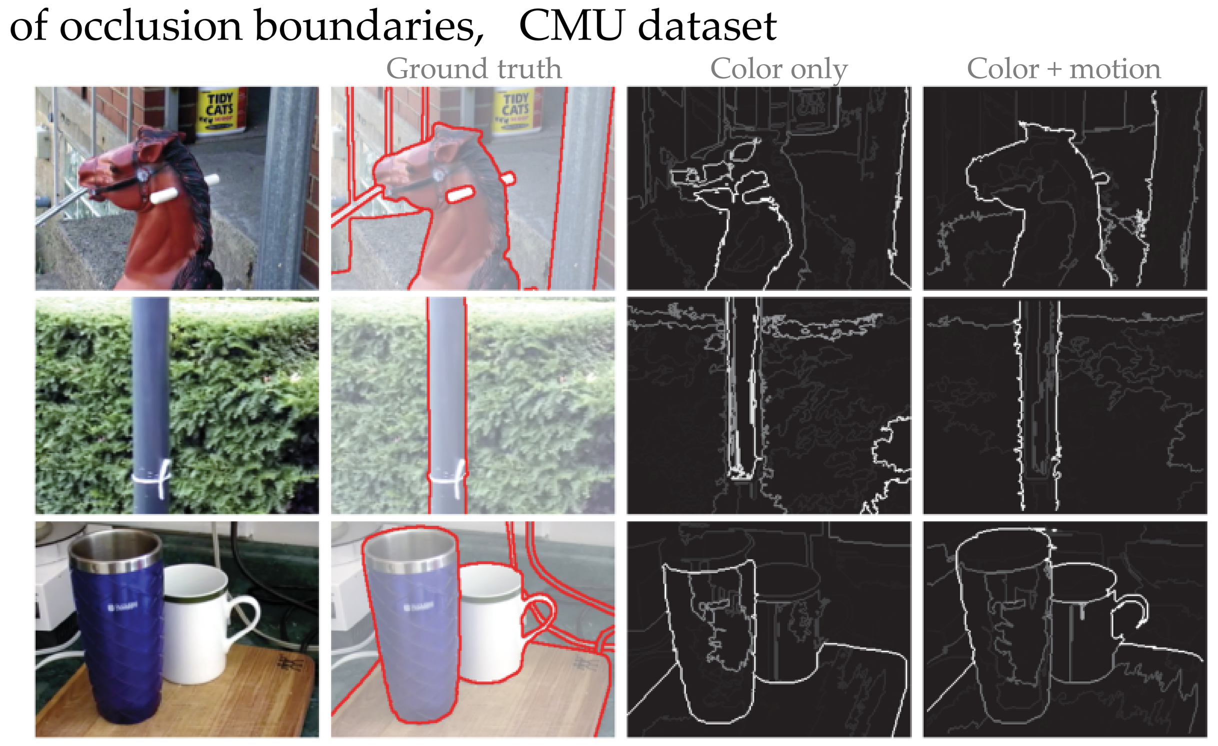


Visualization of results:
Boundary colored similarly as the adjacent segment with the most similar motion histogram
= most foreground segment

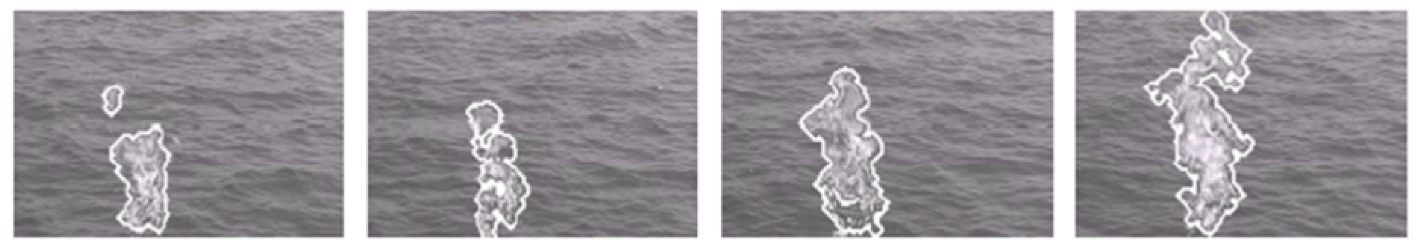
Correct boundary assignment

Incorrect boundary assignment

Detection of occlusion boundaries, CMU dataset



Dynamic texture segmentation (fire over water; see paper for many more examples !)



[1] K. G. Derpanis and R. P. Wildes. Spacetime texture representation and recognition based on a spatiotemporal orientation analysis. IEEE Trans. PAMI, 2012.
[2] W. T. Freeman and E. H. Adelson. The design and use of steerable filters. IEEE Trans. PAMI, 1991.
[3] M. Grundmann, V. Kwatra, M. Han, and I. A. Essa. Efficient hierarchical graph-based video segmentation. CVPR, 2010.

 http://montefiore.ulg.ac.be/~dteney/ Work supported by EPSRC Grant EP/K02339X/1.