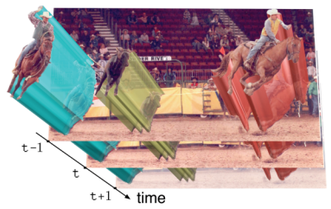
VIDEO OBJECT SEGMENTATION BY TRACKING REGIONS

William Brendel
brendelw@onid.orst.edu

Sinisa Todorovic
sinisa@eecs.oregonstate.edu

ICCV 09 京都

PROBLEM STATEMENT



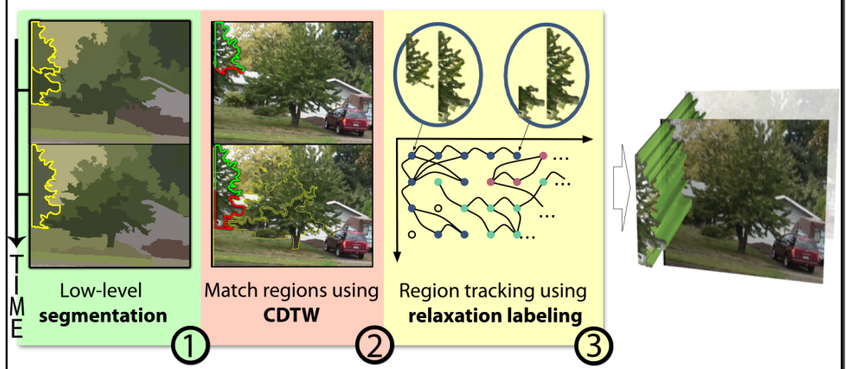
Given a video, delineate the contours of all moving and static objects present.

RATIONALE

Objects in the 3D scene	Objects in the video
are spatially cohesive, and	occupy regions in each frame,
have locally smooth motions.	have small shape/location variations from frame to frame.

Video object segmentation <=> Tracking regions, such that the resulting spatiotemporal tubes are locally smooth

OVERVIEW OF OUR APPROACH



CONTRIBUTIONS

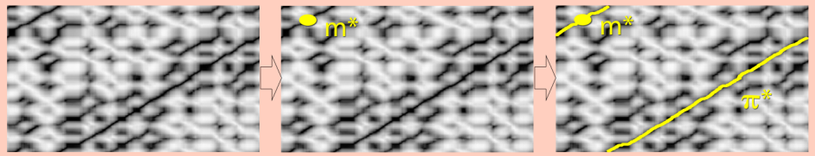
We address region merging and splitting in a low-level segmentation:

② CDTW -- Matching regions by finding parts of their boundaries unaffected by the splits and merges

③ Relaxation labeling -- Many-to-many region matching to find correspondences between splits and merges

② CYCLIC DYNAMIC TIME WARPING (CDTW)

Given two regions, identify only parts of region boundaries that match.



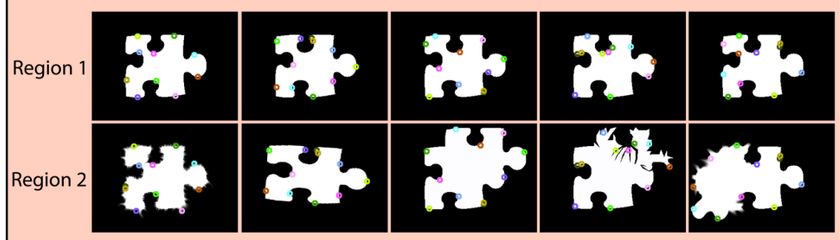
cost matrix C | find starting point m* | DTW from m*

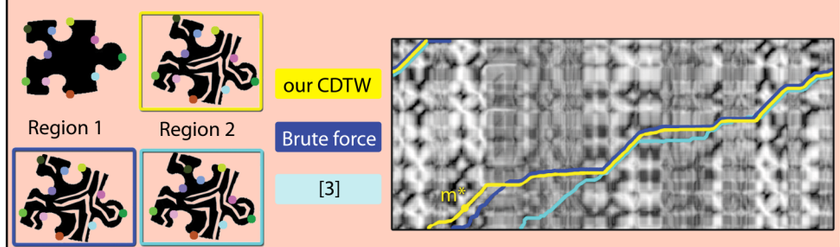
π^* -- optimal path
π_m -- path through point m
C_m -- vicinity of m
$c(\pi_m)$ -- cost along π_m

$$P(\{\pi_m \sim \pi^*\}) \propto \exp(-\mu c(\pi_m)),$$
$$P(\{\pi_n \sim \pi_m\}|\{\pi_m \sim \pi^*\}) \propto \exp(-\lambda |c(\pi_n) - c(\pi_m)|)$$

$$\begin{aligned} m^* &= \max_{m \in C} P(m \in \pi^*) \\ &= \max_{m \in C} P(\{\pi_m \sim \pi^*\}, \{\forall n \in C_m, \pi_n \sim \pi_m\}) \\ &= \max_{m \in C} P(\{\pi_m \sim \pi^*\}) \prod_{n \in C_m} P(\{\pi_n \sim \pi_m\}|\{\pi_m \sim \pi^*\}) \end{aligned}$$

Robustness of our CDTW under various region transformations

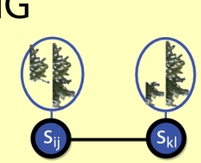




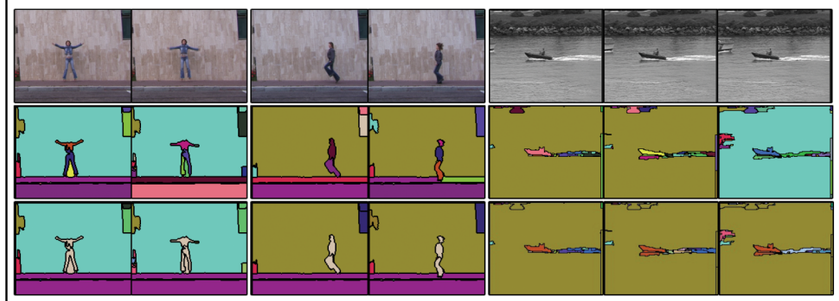
③ RELAXATION LABELING

Given CDTW similarities of two region pairs, s_{ij} and s_{kl}, cluster them if they move similarly.

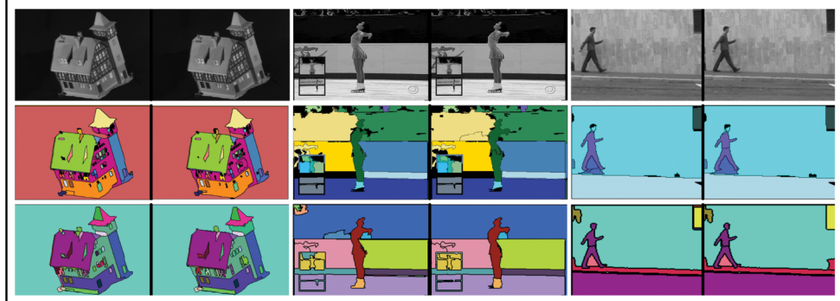
correlation $r = 2\,[[s_{ij} \cdot s_{kl}] \text{ XOR } \overline{s_{ijkl}}] - 1$



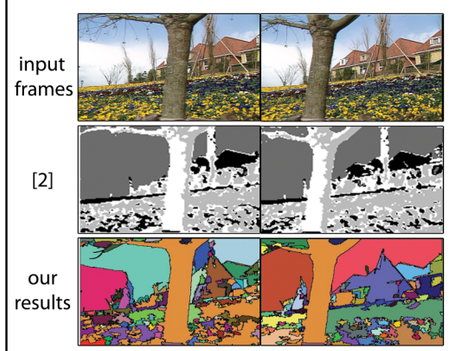
RESULTS

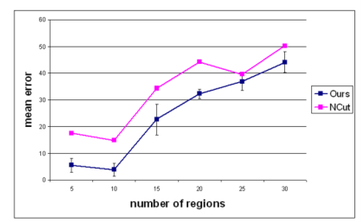


top: input frames; middle: without tracking; bottom: with tracking



top: input frames; middle: [1]; bottom: our results





Segmentation error with and without region tracking as a function of the number of regions per frame.

	Background		Foreground	
Videos	MeanShift	Ours	MeanShift	Ours
Jack	14.03%	0.98%	59.11%	0.51%
Run	30.70%	0.35%	73.10%	5.39%
Skip	14.59%	0.53%	73.21%	5.52%
Walk	8.18%	0.68%	54.76%	2.51%
10 activities	**16.88%**	**0.64%**	**65.04%**	**3.48%**

	Background		Foreground	
Videos	NCut	Ours	NCut	Ours
Bend	14.67%	3.92%	18.52%	0.03%
Jump	20.57%	9.24%	16.34%	0.05%
PJump	10.93%	2.27%	0.30%	0.30%
Side	21.93%	7.09%	12.92%	0.73%
Wave-1	15.95%	7.57%	3.89%	0.42%
Wave-2	12.71%	7.36%	23.14%	0.28%
10 activities	**16.13%**	**6.24%**	**12.52%**	**0.30%**

Segmentation error with and without region tracking on Wiezmann activity videos.

[1] V. Hedau, H. Arora, and N. Ahuja. Matching images under unstable segmentations. In CVPR, 2008
[2] A. Torsello, M. Pavan, and M. Pelillo. Spatio-temporal segmentation using dominant sets. In EMMCVPR, 2005
[3] N. Arica. Cyclic sequence comparison using dynamic warping. in CIVR, 2005