What more can we do with videos?

Occlusion and Motion Reasoning for Tracking & Human Pose Estimation

Karteek Alahari

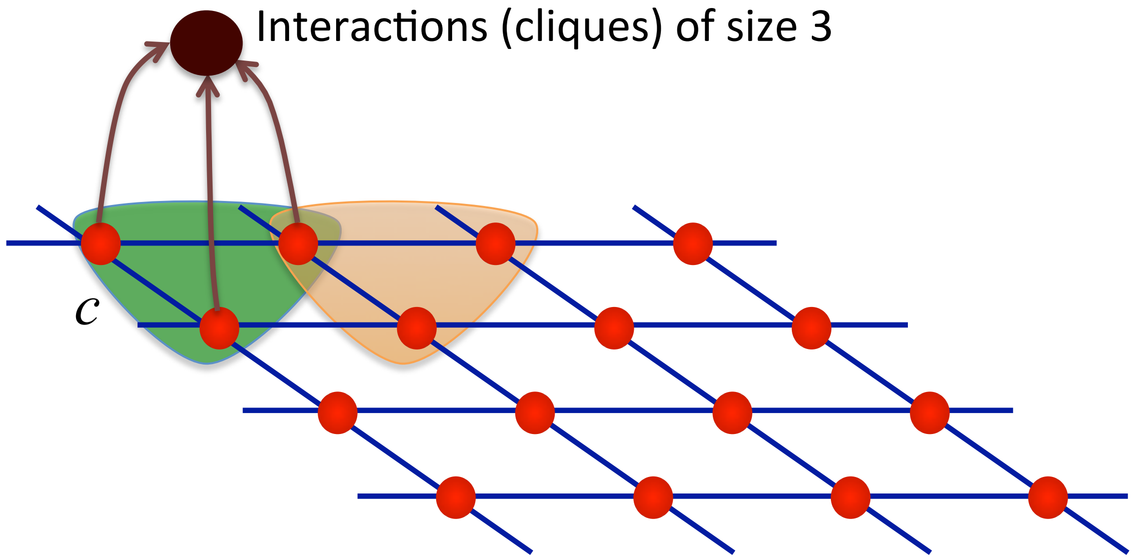
Inria Grenoble – Rhone-Alpes

Joint work with Anoop Cherian, Yang Hua, Julien Mairal, Cordelia Schmid

But before that… a blast from the past

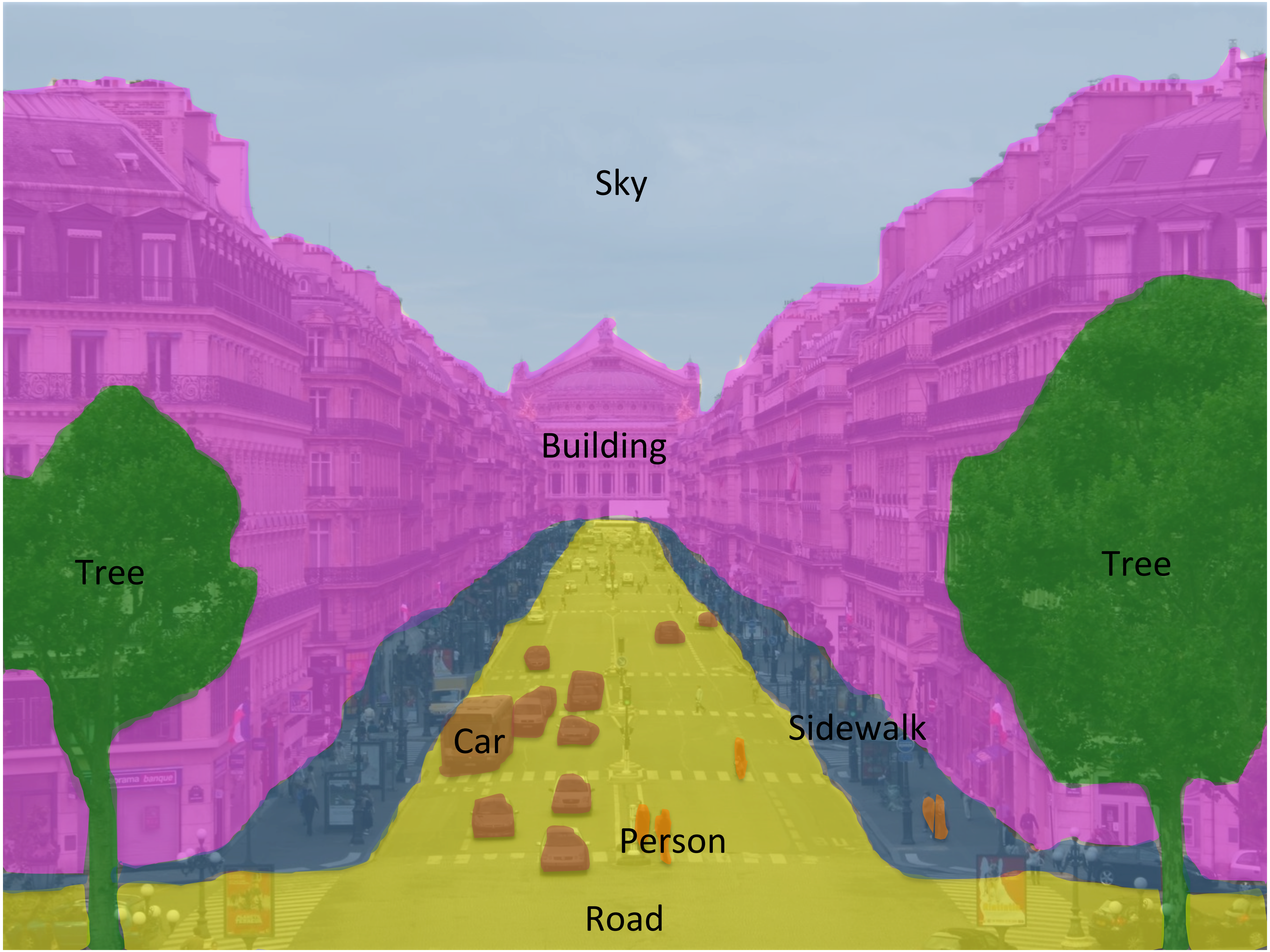
Scene Understanding

$$E(\mathbf{x}) = \sum_{i \in \mathcal{V}} \psi_i(x_i) + \sum_{(i,j) \in \mathcal{E}} \psi_{ij}(x_i, x_j) + \underbrace{\sum_{c \in \mathcal{S}} \psi_c(\mathbf{x}_c)}_{\text{New higher order potentials}}$$



Joint work with L. Ladicky, C. Russell, P. Sturgess, P. H. S. Torr


sorama banque

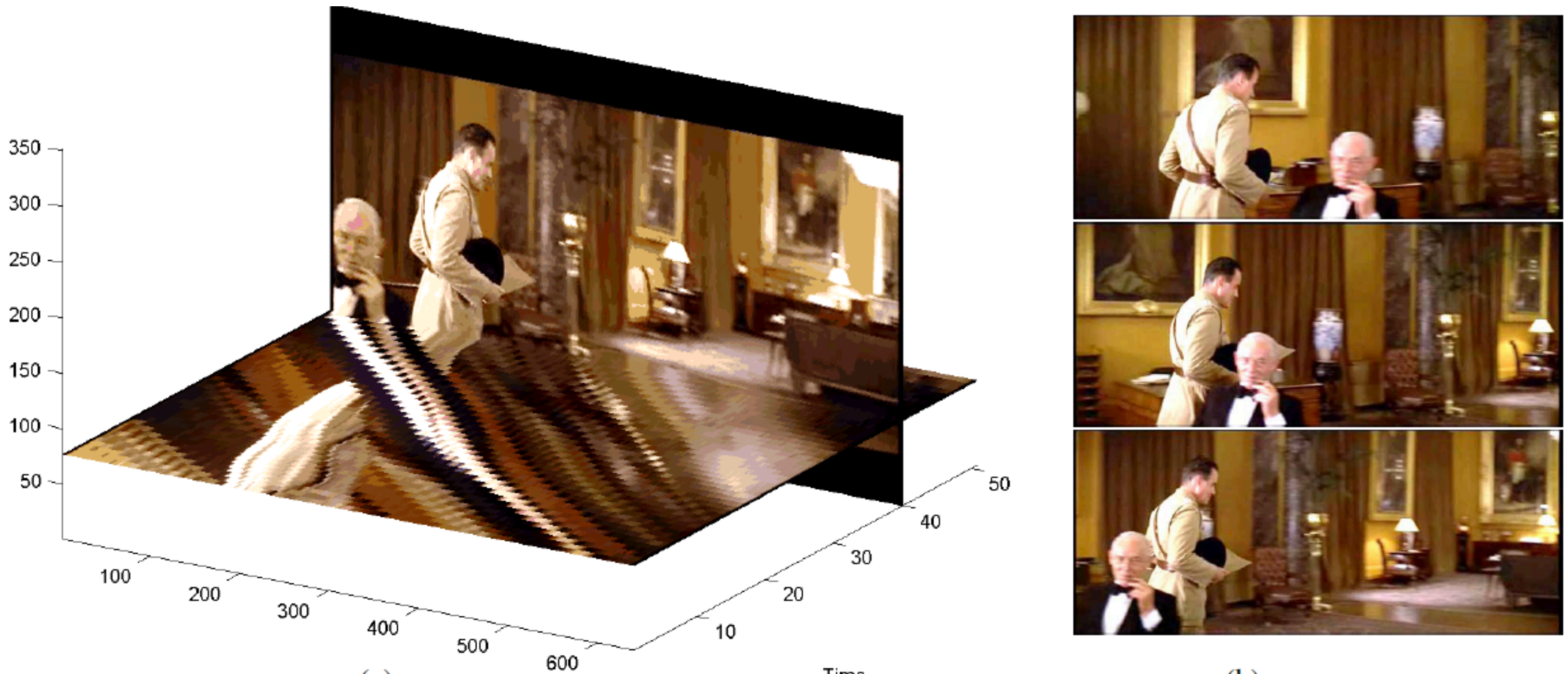

Sky
Building
Tree
Tree
Sidewalk
Car
Person
Road

Space-time video (over-) segmentation

Hollywood dataset: Laptev et al., '08

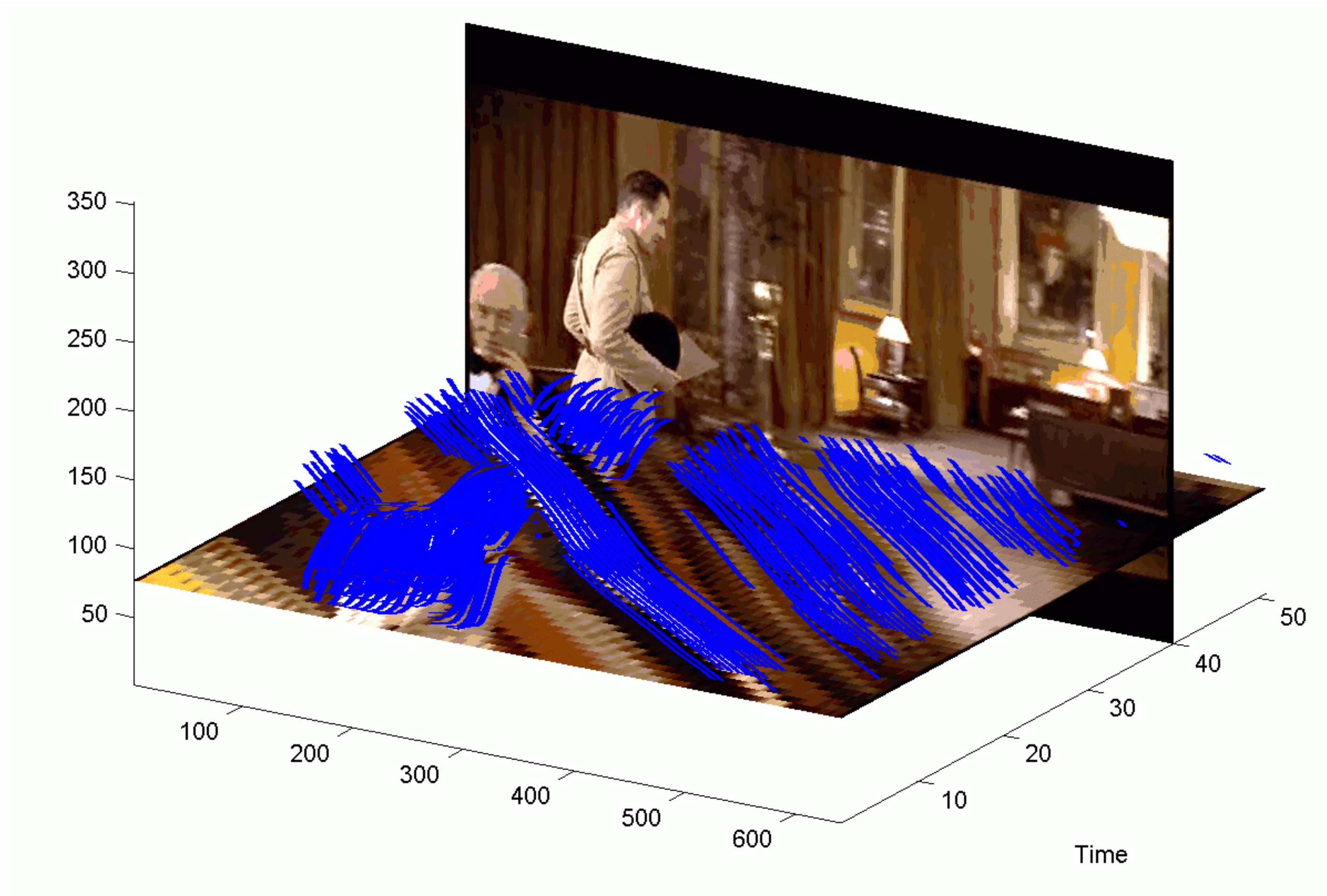
Joint work with I. Laptev, J. Lezama, J. Sivic

Video as a space-time volume



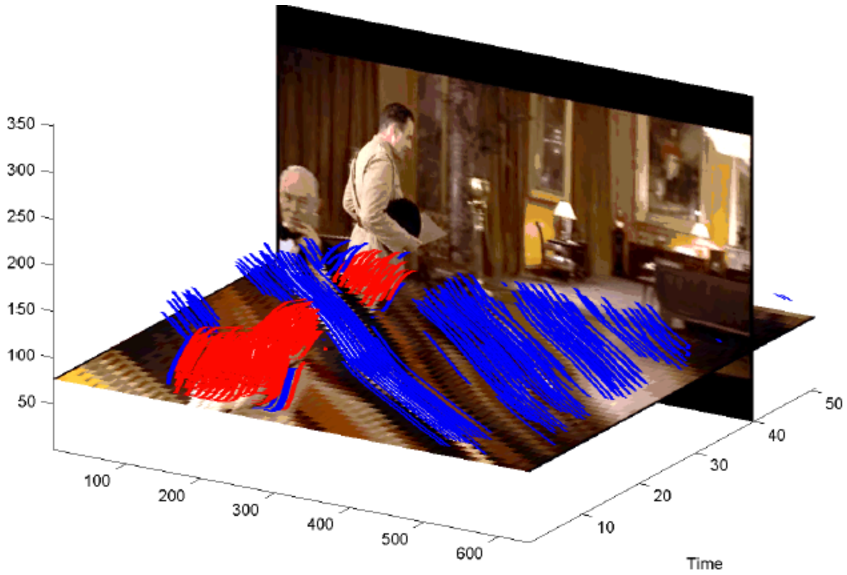
(a) (b)

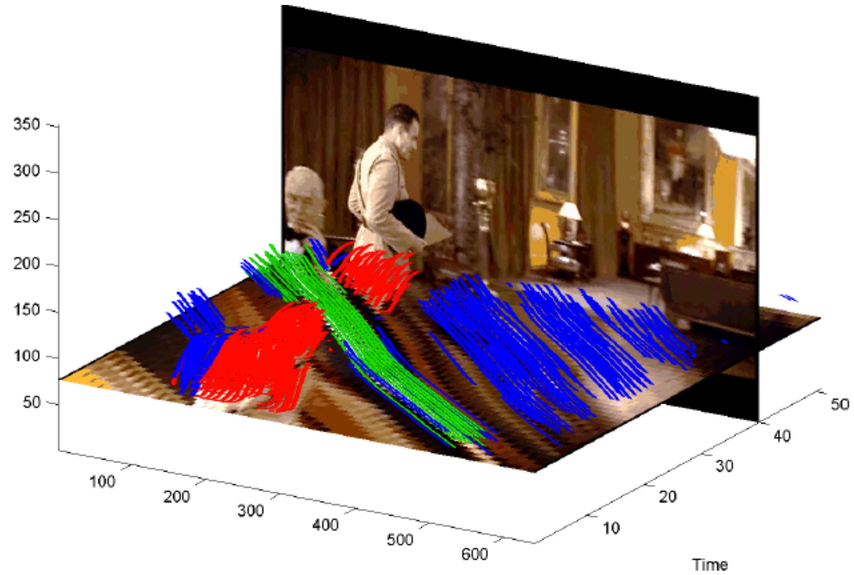
Point-tracks to capture long-range motion



Brox and Malik, ECCV '10
Wang et al., CVPR '11

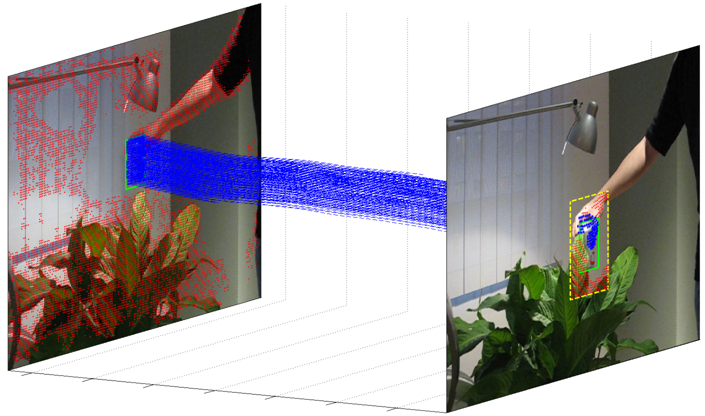
Track Clustering



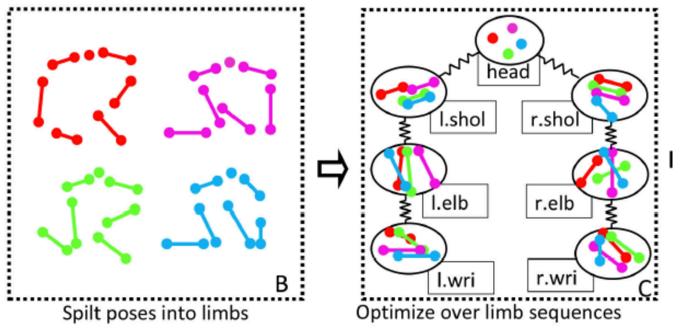


Outline

- Use the tracks to estimate the state of the object

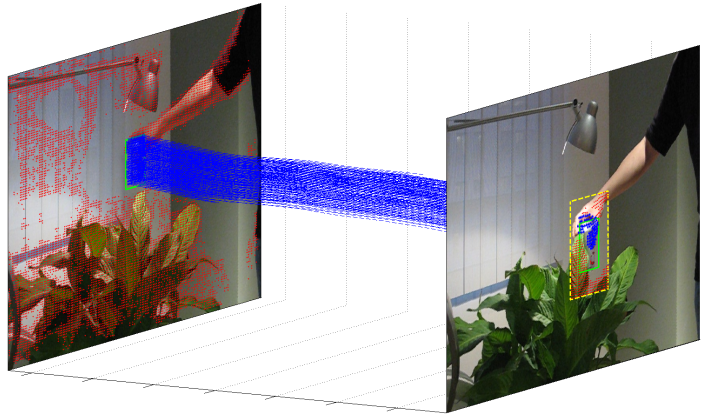


- Human pose estimation in videos

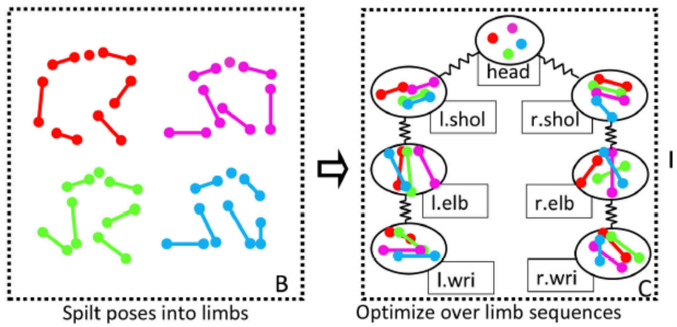


Outline

- Use the tracks to estimate the state of the object



- Human pose estimation in videos



Object Tracking

Joint work with Y. Hua, C. Schmid

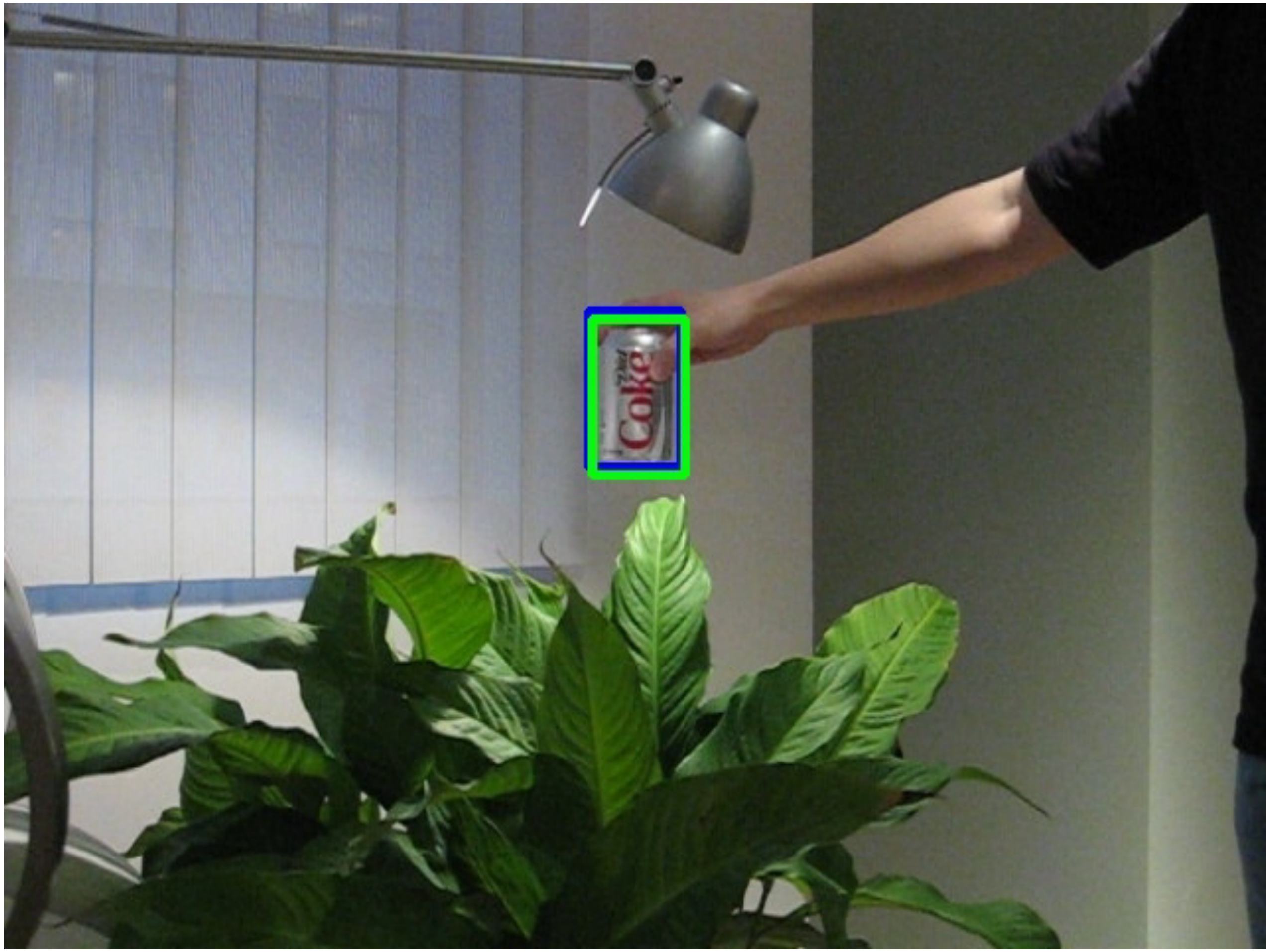
Object Tracking

Joint work with Y. Hua, C. Schmid

Object Tracking

Joint work with Y. Hua, C. Schmid

Object Tracking



TLD SPLTT Struck Ours

Object Tracking

- Tracking-by-detection approaches
  - Struck [Hare et al., ICCV 2011]
  - SPLTT [Supancic III et al., CVPR 2013]

Object Tracking

- Tracking-by-detection approaches
  - Struck [Hare et al., ICCV 2011]
  - SPLTT [Supancic III et al., CVPR 2013]

Frame 1

Object Tracking

- Tracking-by-detection approaches
  - Struck [Hare et al., ICCV 2011]
  - SPLTT [Supancic III et al., CVPR 2013]

Frame 1

Object labelled in frame 1

Object Tracking

- Tracking-by-detection approaches
  - Struck [Hare et al., ICCV 2011]
  - SPLTT [Supancic III et al., CVPR 2013]

Frame 1

Object labelled in frame 1

Learn a model with this annotation

Object Tracking

- Tracking-by-detection approaches
  - Struck [Hare et al., ICCV 2011]
  - SPLTT [Supancic III et al., CVPR 2013]

Frame 2

Evaluate the model on new frames

Object Tracking

- Tracking-by-detection approaches
  - Struck [Hare et al., ICCV 2011]
  - SPLTT [Supancic III et al., CVPR 2013]

Frame 2

Evaluate the model on new frames

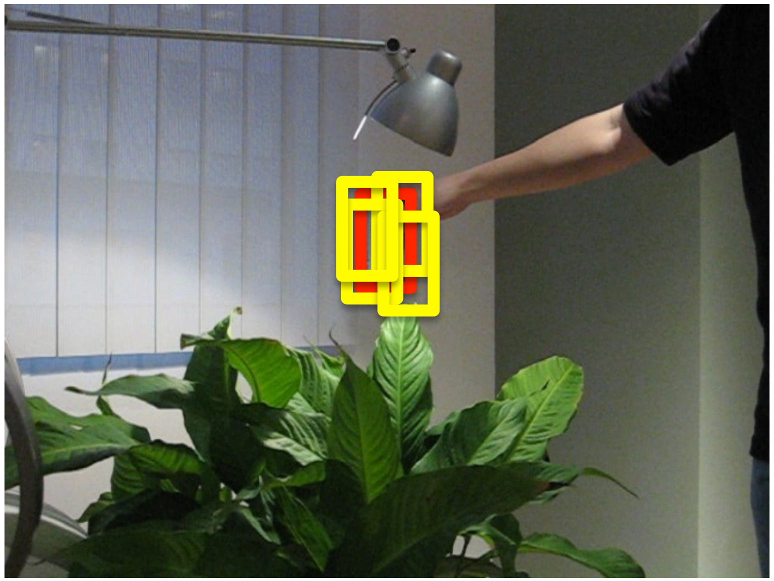
Update the model

When to update?

- Struck [Hare et al., ICCV 2011]
  - With every new detection

When to update?

- Struck [Hare et al., ICCV 2011]
  - With every new detection



When to update?

- SPLTT [Supancic III et al., CVPR 2013]
  - A selection of detections

When to update?

- Continuous update
  - Leads to drifting

Object occluded or leaves the frame

When to update?

- Continuous update
  - Leads to drifting

Object occluded or leaves the frame

Object changes in appearance

When to update?

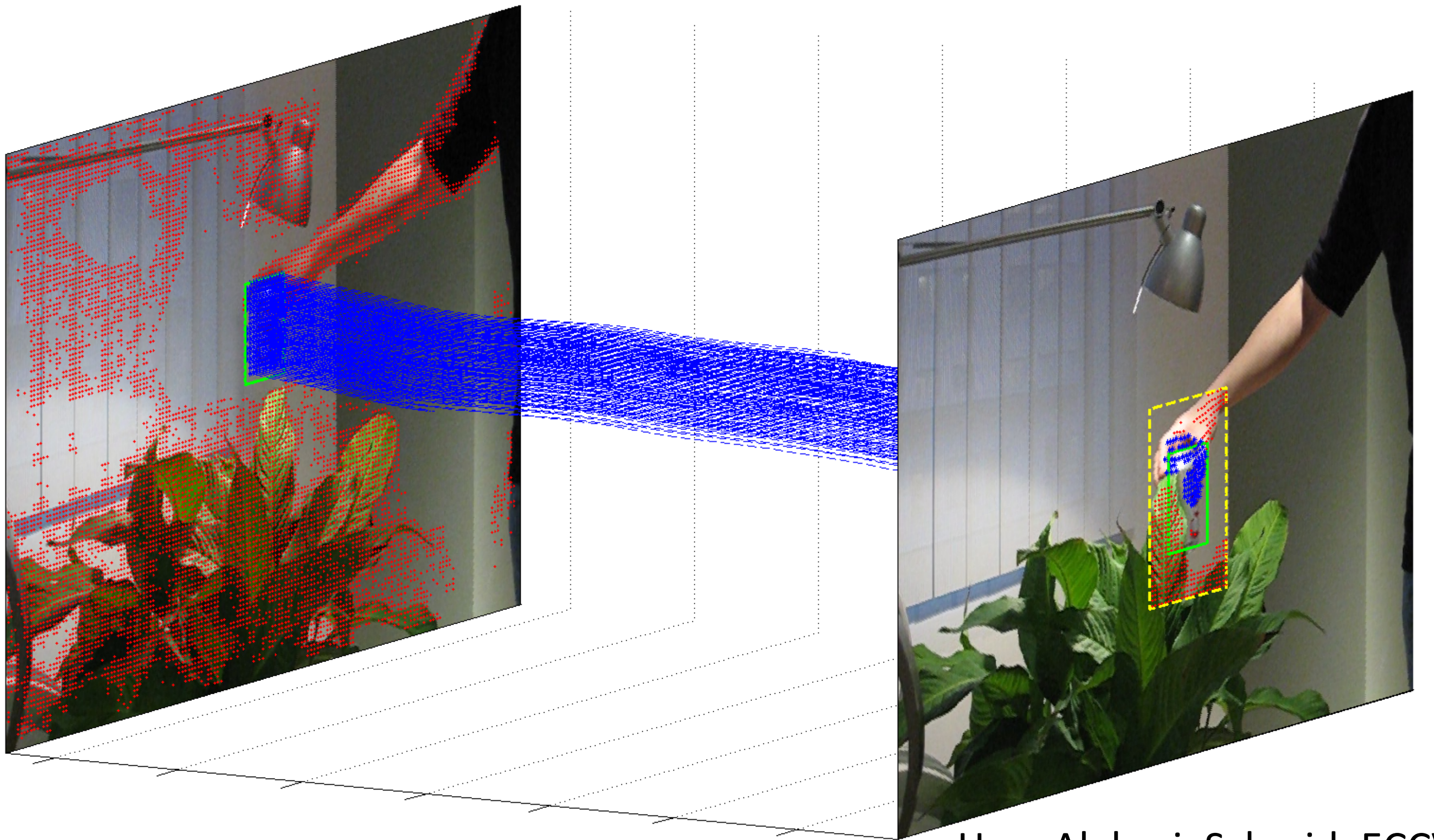
- Continuous update
  – Leads to drifting

Object occluded or leaves the frame

Object changes in appearance

Determine the object state

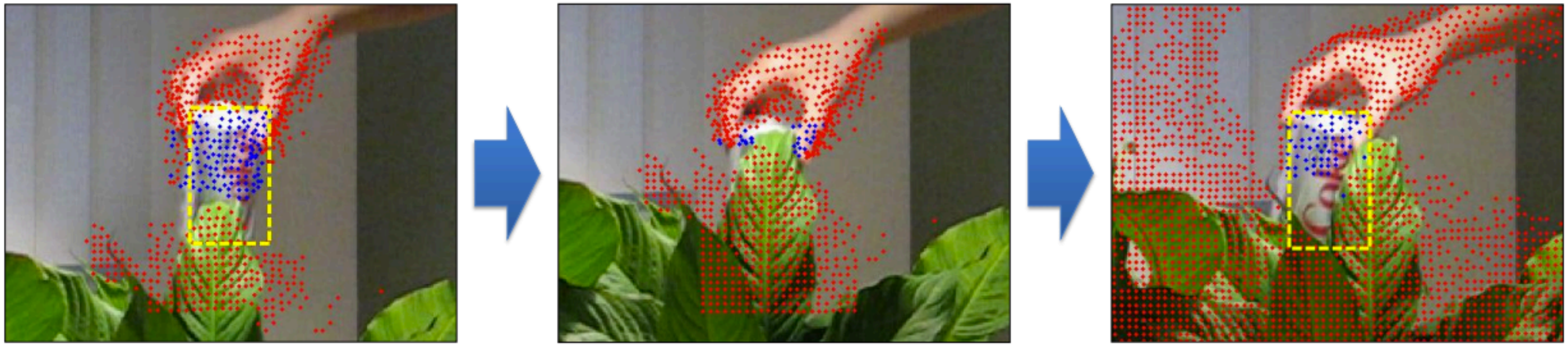
- e.g., occlusion



Hua, Alahari, Schmid, ECCV 2014

Determine the object state

- e.g., occlusion



Frame 251: **Partial occlusion**
Continue to track and no model updating

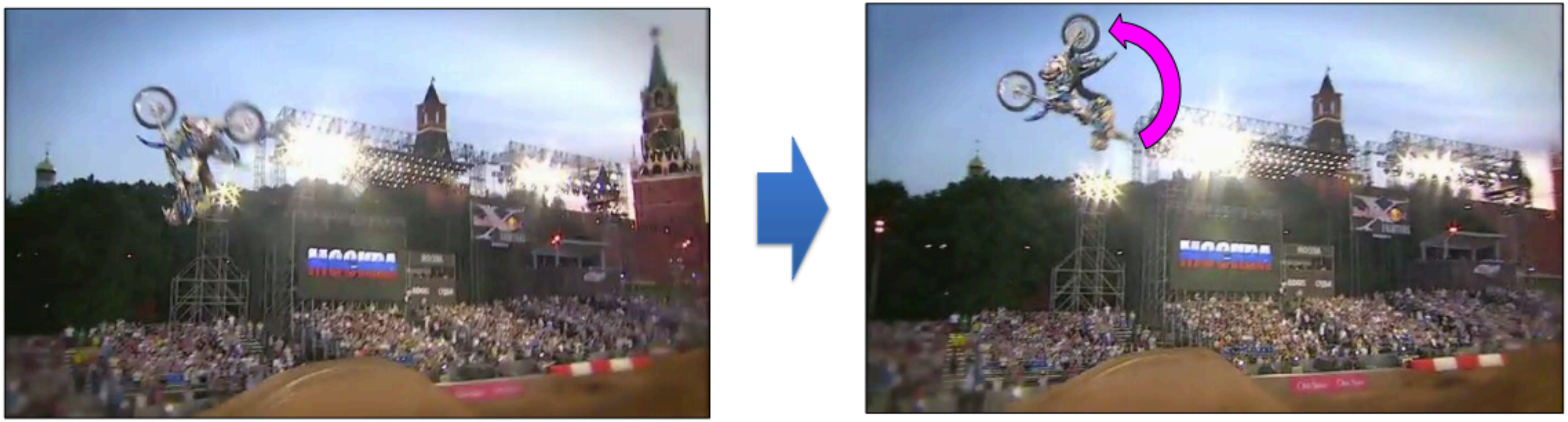
Frame 254: **Full occlusion**
Stop tracking and no model updating

Frame 269: **Object reappears**
Recover from occlusion with global detector scanning

Hua, Alahari, Schmid, ECCV 2014

Determine the object state

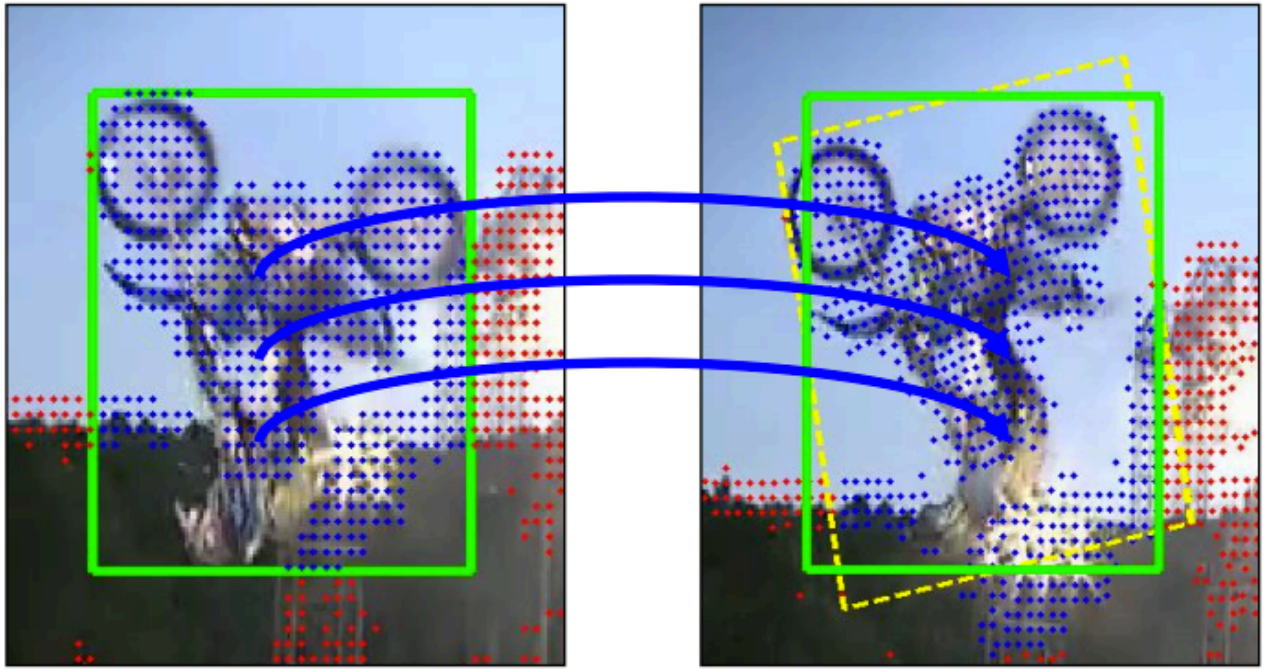
- e.g., geometric transformation



Hua, Alahari, Schmid, ECCV 2014

Determine the object state

- e.g., geometric transformation

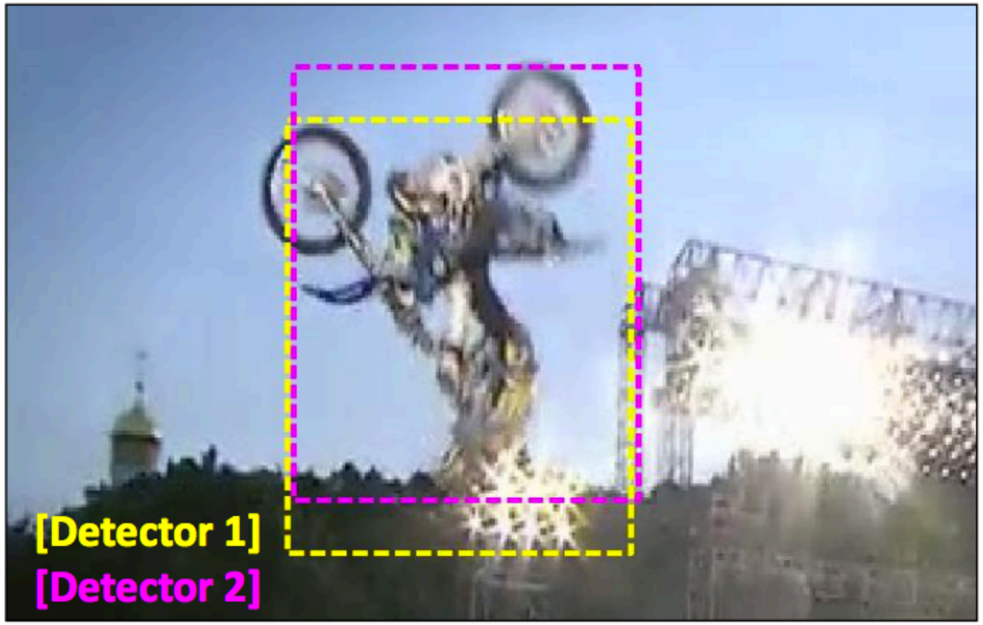


- We estimate a similarity transform

Hua, Alahari, Schmid, ECCV 2014

Determine the object state

- If a significant change occurs, we
  - train a new detector; and
  - maintain a set of exemplar detectors



Hua, Alahari, Schmid, ECCV 2014

Object Tracking: Results

- Evaluated on a benchmark & TLD dataset

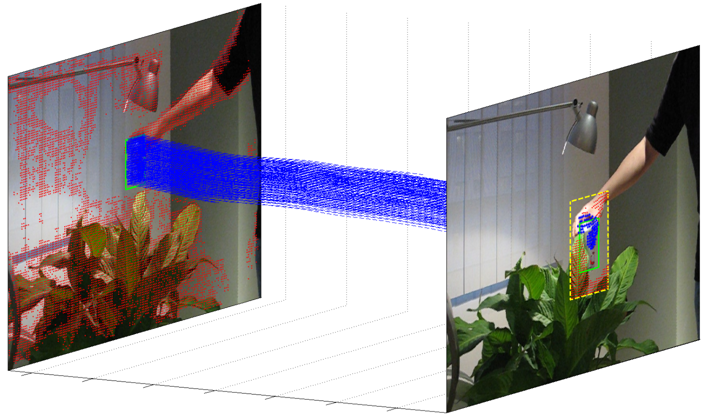
Dataset	Sequence	Struck	TLD	SPLTT	Ours (plain)	Ours (Occ + VP)
Benchmark Dataset	Coke	**0.948**	0.694	0.804	0.801	0.880
	MotorRolling	0.146	0.110	0.128	0.134	**0.512**
	Football1	0.378	0.351	0.554	**1.000**	**1.000**
	Trellis	0.821	0.455	0.701	0.838	**0.919**
	Walking	0.585	0.379	0.541	0.476	**0.922**
	Freeman4	0.177	0.134	0.145	**0.205**	0.004
TLD Dataset	Pedestrian2	0.175	0.500	0.950	0.107	**0.979**
	Carchase	0.036	**0.340**	0.290	0.098	0.312

Benchmark dataset: Wu et al., 2013
SPLTT: Supancic and Ramanan, 2013
Struck: Hare et al., 2011
TLD: Kalal et al., 2012

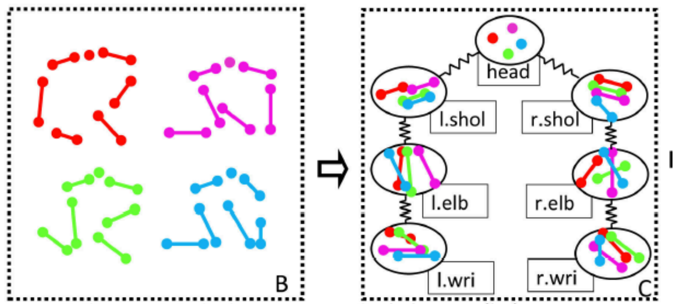
Object Tracking: Summary

Outline

- Use the tracks to estimate the state of the object



- Human pose estimation in videos



Spilt poses into limbs

Optimize over limb sequences

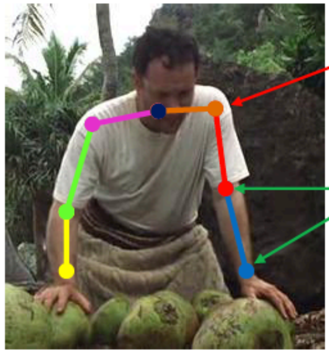
Human Pose Estimation

Poses in the Wild dataset

Joint work with A. Cherian, J. Mairal, C. Schmid

Human Pose Estimation (in an image)

- Formulated as a graph optimization problem



ϕ_u : unary potential

$\psi_{u,v}$: pairwise potential

For an image *I*, pose model $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, and

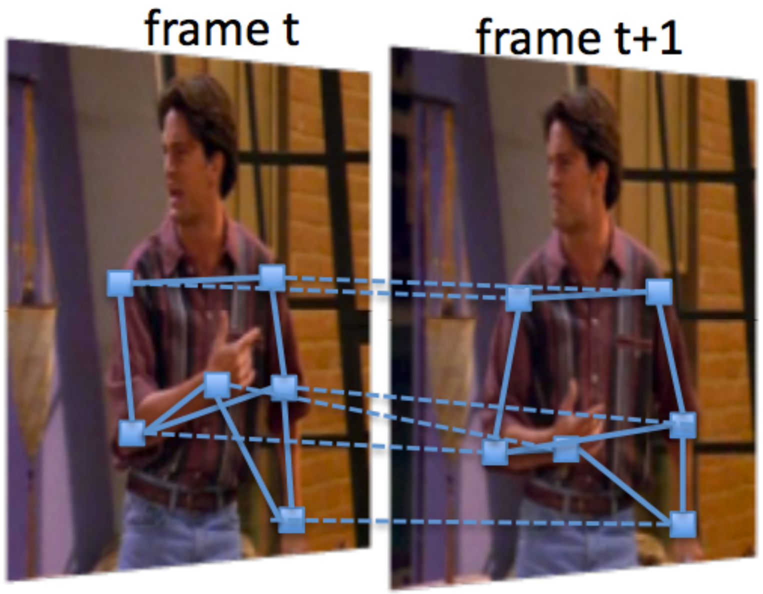
$$p = \{p^u = (x^u, y^u) \in \mathbb{R}^2 : \forall u \in \mathcal{V}\}$$

$$\min C(I, p) := \sum_{u \in \mathcal{V}} \phi_u(I, p^u) + \sum_{(u,v) \in \mathcal{E}} \psi_{u,v}(p^u - p^v)$$

Yang and Ramanan, CVPR 2011

Human Pose Estimation

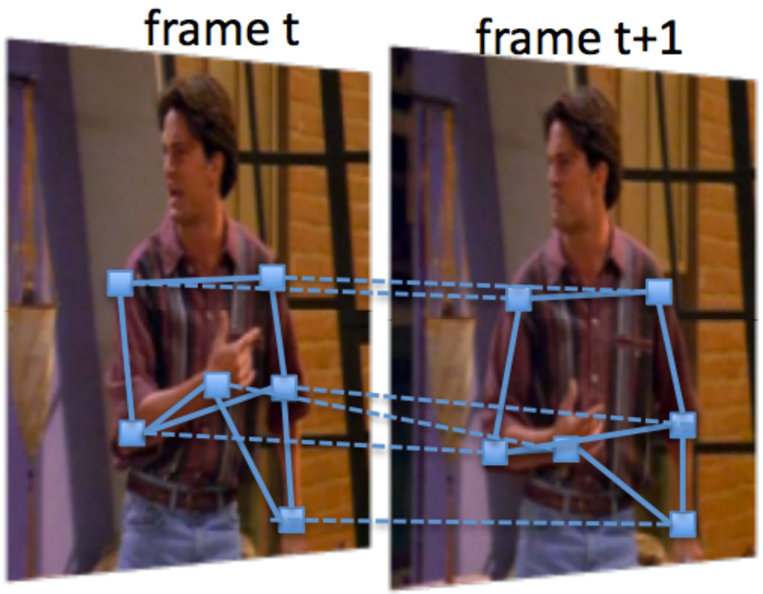
- Extension to videos: introduce temporal links



e.g., Sapp et al., '11, Tokola et al., '13

Human Pose Estimation

- Extension to videos: introduce temporal links
- Inference is now intractable – requires approximate methods



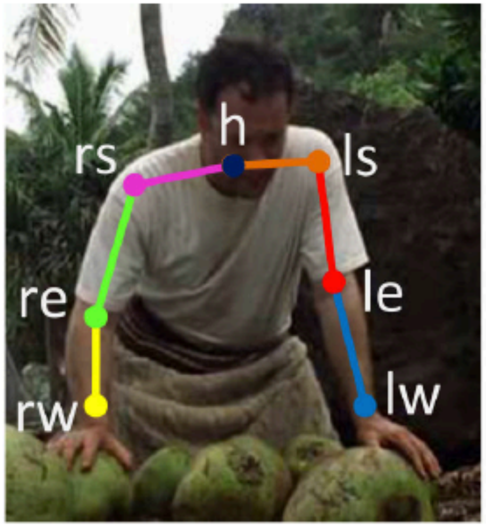
e.g., Sapp et al., '11, Tokola et al., '13

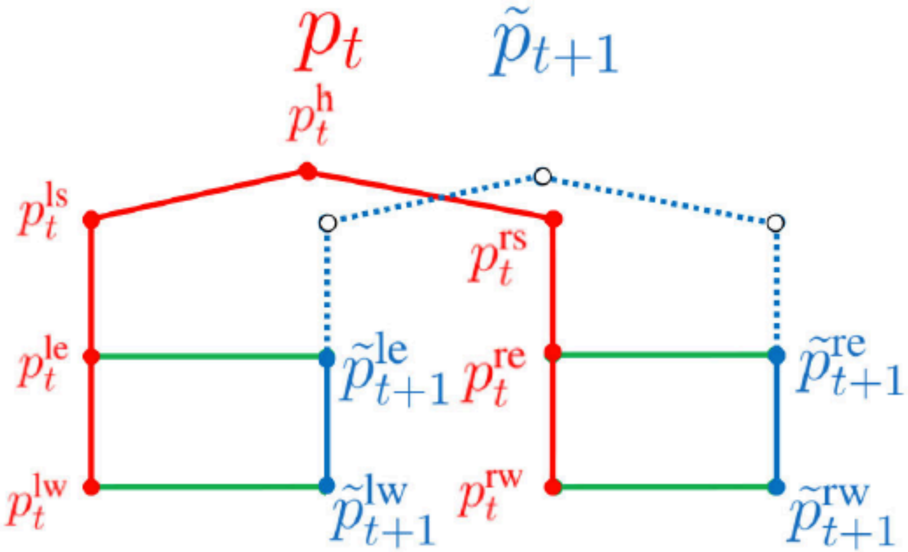
Human Pose Estimation

- Extension to videos: introduce temporal links
- Inference is now intractable – requires approximate methods
- e.g.,
  - Sapp et al. '11: Convex combination of trees
  - Park & Ramanan '11: Candidate set of poses
  - Tokola et al. '13: Restrict the set of part tracks

Our Approximations

- Stabilize the lower-limb pose estimates
- Decompose poses and perform limb-tracking

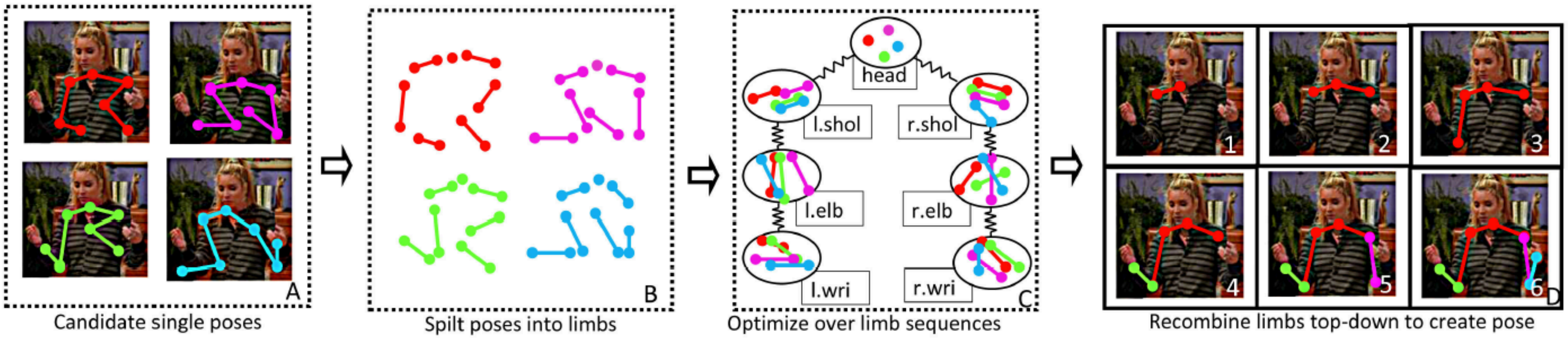




Cherian, Mairal, Alahari, Schmid, CVPR 2014

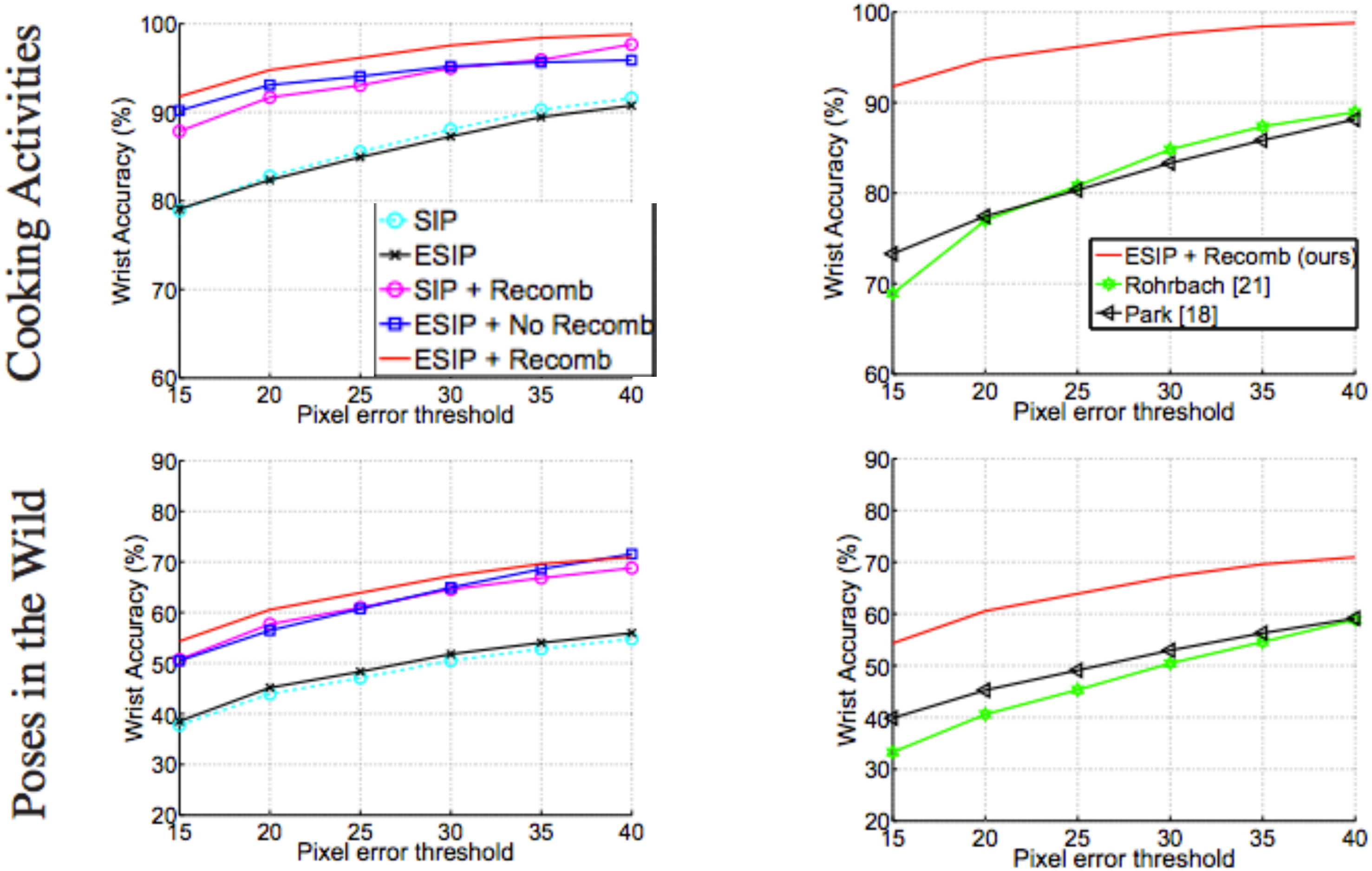
Our Approximations

- Stabilize the lower-limb pose estimates
- Decompose poses and perform limb-tracking

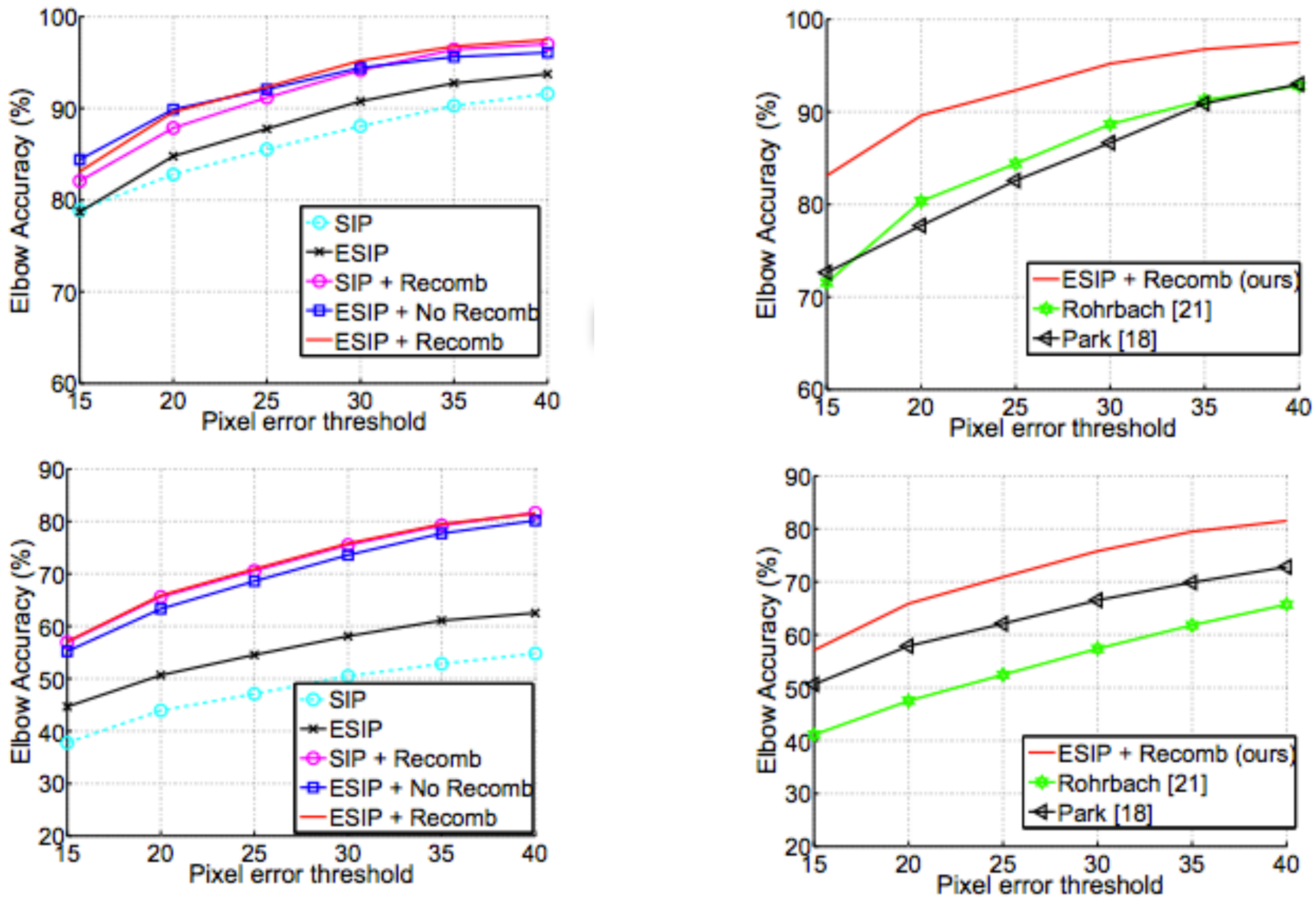


Cherian, Mairal, Alahari, Schmid, CVPR 2014

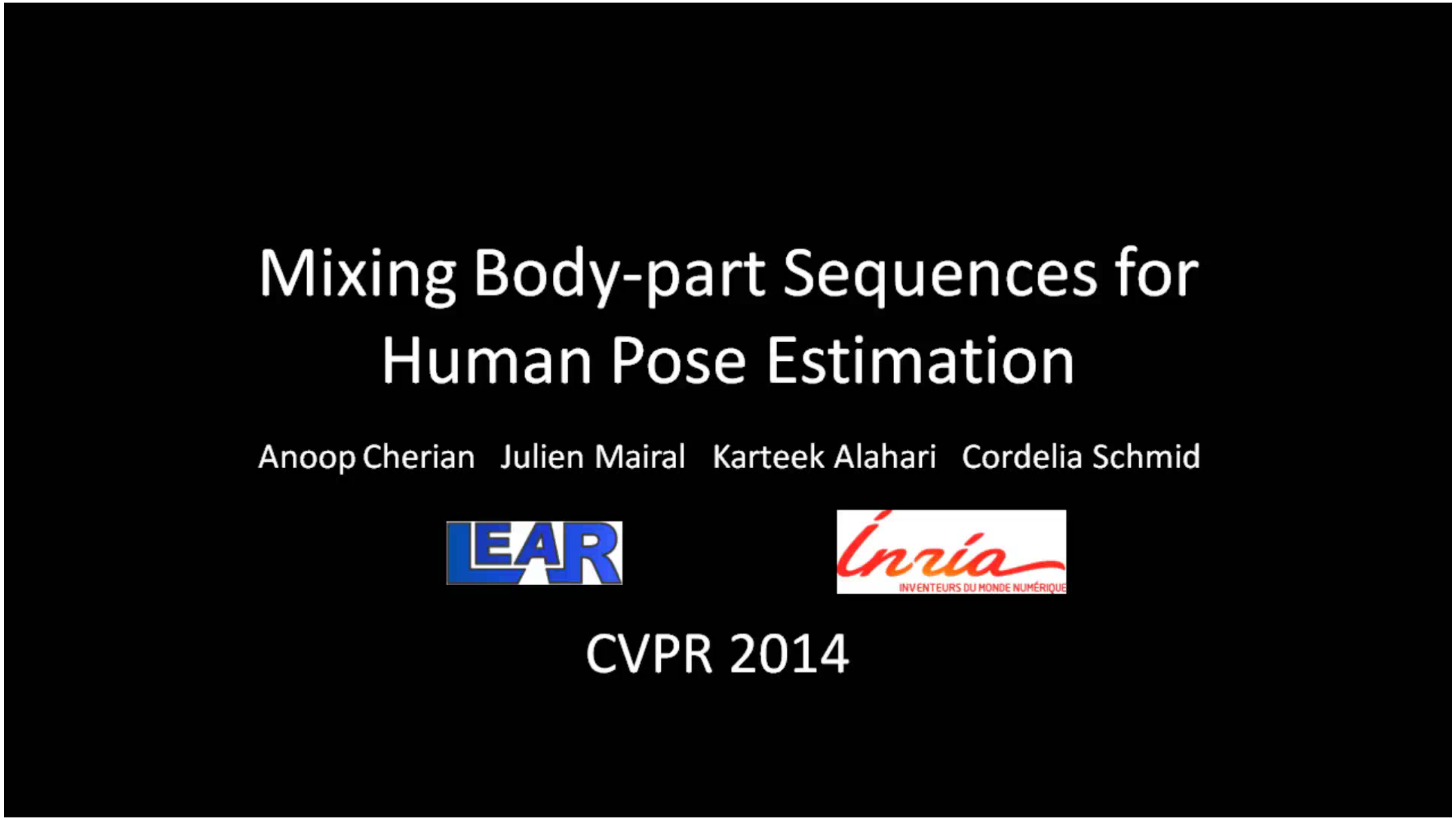
Human Pose Estimation



Human Pose Estimation

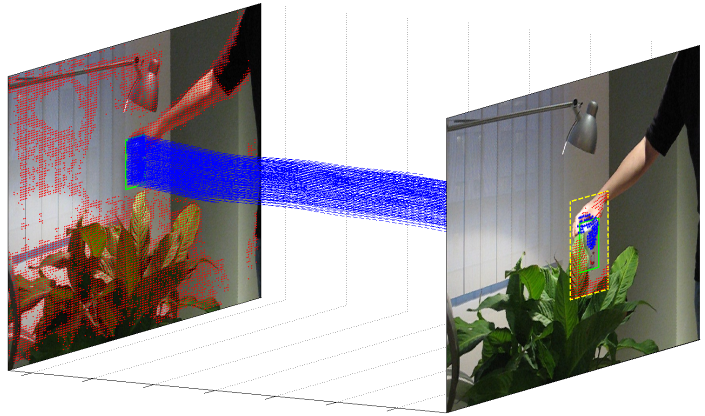


Human Pose Estimation

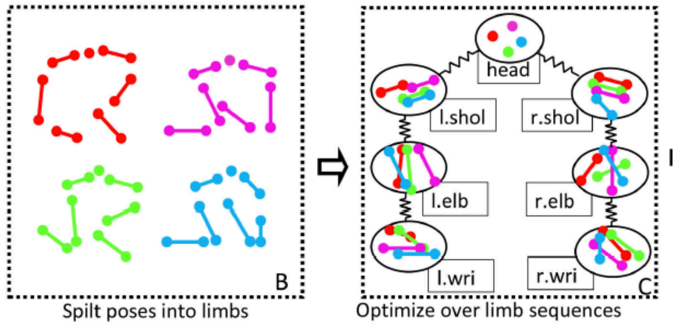


Summary

- Use the tracks to estimate the state of the object



- Human pose estimation in videos



Thank you!