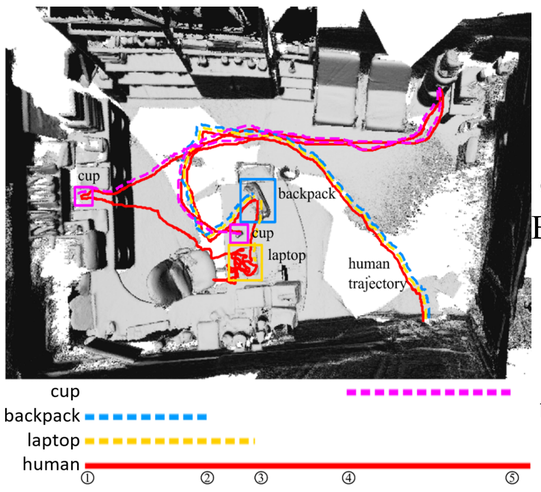
Tracking Occluded Objects and Recovering Incomplete Trajectories by Reasoning about Containment Relations and Human Actions

Wei Liang*♦ Yixin Zhu♦ Song-Chun Zhu♦
*Beijing Institute of Technology, China ♦University of California, Los Angeles

Motivation

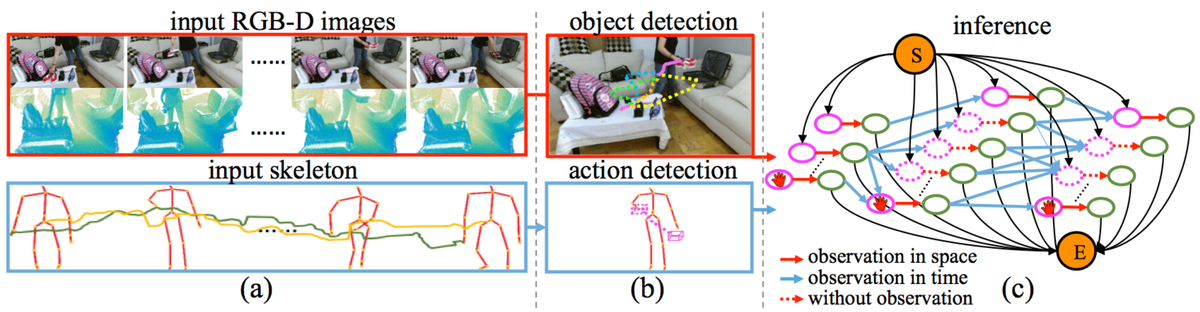
This paper studies a challenging problem of tracking severely occluded objects in long video sequences. The proposed method reasons about the containment relations and human actions, thus infers and recovers occluded objects identities while contained or blocked by others.



An example scenario for tracking occluded objects in an indoor scene. The dashed lines represent the inferred trajectories and different colors indicate different objects in the scene.

By explicitly reasoning about containment relations, the proposed algorithm is capable of recovering full trajectories of objects even they are contained or occluded by other objects in the video.

Framework



(a) A sequence of RGB-D images and human skeleton captured by a Kinect sensor.
(b) Object detection and human action detection algorithms were applied to extract the object location and human actions per frame.
(c) Inference on a network flow representation. The solid red lines denote the observations in space. The dashed red lines denote that the present state of the object is hidden and there is no observation. The blue lines denote the observations in time.

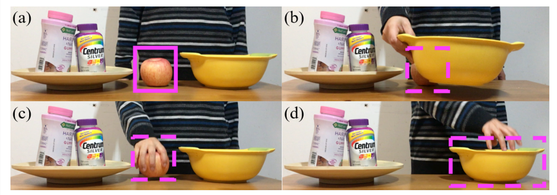
Formulation

$$T^* = \arg\max_T P(T|I) = \arg\max_T P(T|\mathcal{X}, H)$$
$$\propto \arg\max_T P(\mathcal{X}|T)P(T|H)$$
$$= \arg\max_T \prod_k P(\mathcal{X}|T^k)P(T^k|H)$$

$$P(\mathcal{X}|T^k) = \prod_{t=1}^{\tau} P(\mathcal{X}_t|x_t^k)$$
$$P(T^k|H) = P(\{x_1^k, x_2^k, \cdots, x_\tau^k\}|H)$$
$$= P_S(x_0^k)\prod_{t=1}^{\tau} P(x_t^k|x_{t-1}^k, H_{t-1})\, P_E(x_\tau^k)$$

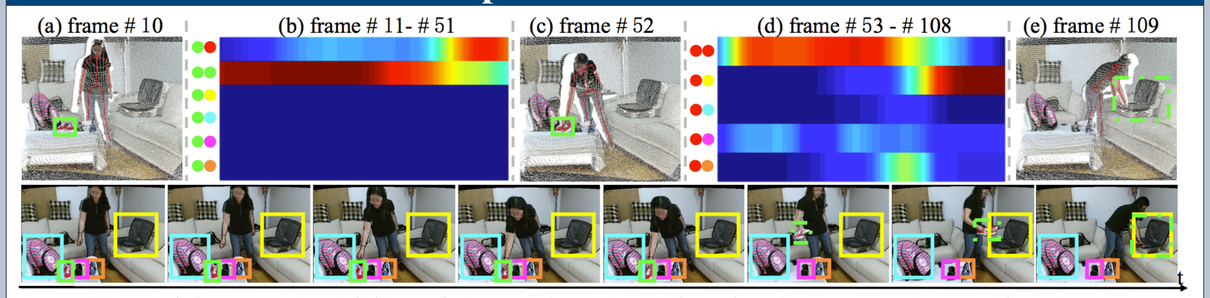
T: Trajectories of all objects I: Input video $\mathcal{X}$: The object detection in space H: Human action in time

$P(x_t^k|x_{t-1}^k, H_{t-1})$: The transition probability of two consecutive frames based on the observation of human action



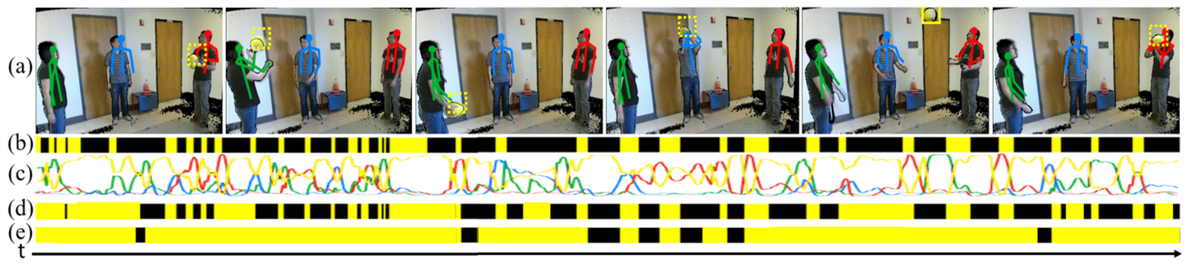
Spatial hypotheses by containment relations:
i) Blocked: An apple (a) can be detected at first, later becomes occluded by a bowl.
ii) Contained: The apple is contained by a person (c) and a bowl (d), respectively.

Experiment Results



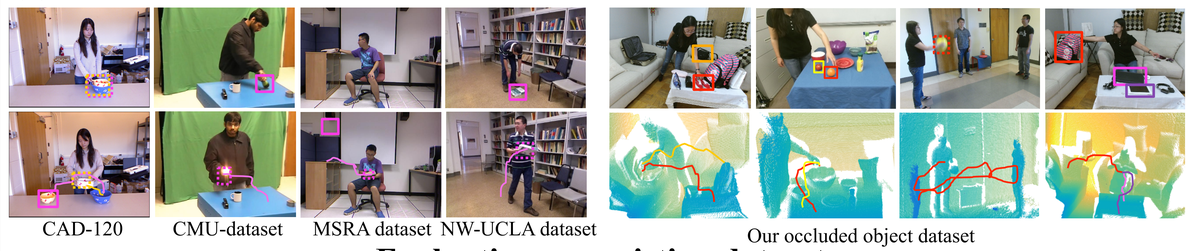
Transition probability of the object location in the green bounding box

The solid boxes depict that the object is tracked by object detectors. The dashed boxes depict that the object is recovered by inference.
(a), (c) and (e) show detected bounding boxes and human skeletons on point cloud.
(b) and (d) are the transition probabilities between two possible locations.
(b)The bottom four bars with low probability keep the same since we constrain the impossible object moving that are not caused by human actions.



An example of tracking a ball

In each bar, the yellow represents the correct results, and the black represents the wrong results. Different colors denote different objects: actor 1 (green), actor 2 (blue), actor 3 (red) and ball (yellow).
(a) Examples of tracking results. The dashed boxes depict the object is occluded.
(b) Temporal-suppression results.
(c) The scores of consistency between object movement and human action.
(d) Spatial-suppression results.
(e) Full model results.



Evaluations on existing datasets