CDT: Cooperative Detection and Tracking for Tracing Multiple Objects in Video Sequences

Han-Ul Kim, and Chang-Su Kim
School of Electrical Engineering, Korea University

Motivation

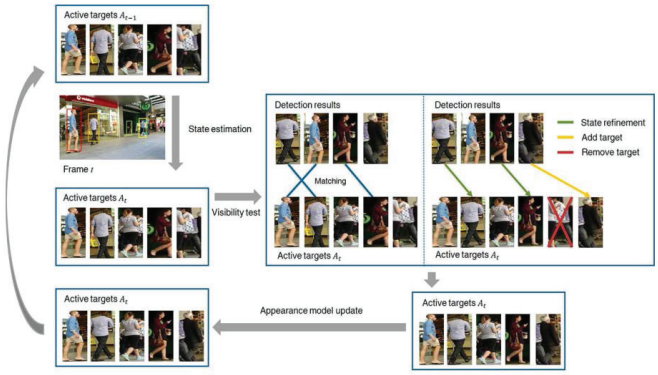
- Motivation
  - In tracking-by-detection, the tracking performance is closely related to the detection quality
  - A model-free tracking algorithm can be used to find undetected objects
- Our approach
  - We improve the detection quality by combining an object detector with a model-free tracker

Object Detection

- Training
  - We employ the pre-trained Fast R-CNN detector[1]
  - Then, we fine-tune the detector on MOT challenge dataset[2]
- Test
  - We use only the detection results with high confidence to reduce false positives
  - In the proposed algorithm, the impacts of undetected objects are less severe than those of false positives.

Forward Tracking

- Overview

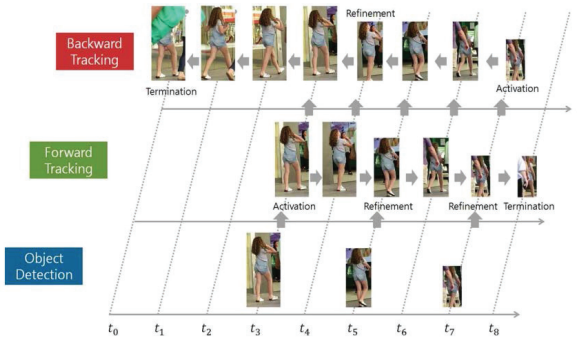


- State estimation
  - We perform model-free tracking to find target states
- Detection guidance
  - Visibility test
    - Find the disappearing target by computing the detection score
    - Determine occluded targets by computing IoU overlap ratio
  - Matching between detection results and active targets
    - Improve target states using matched detection results
    - Add new active target using unmatched detection results
- Appearance Model Update
  - We model the appearance of an active target using SSVM

$$\mathbf{w}_i = \arg\min_{\mathbf{w}} \frac{1}{2}\|\mathbf{w}\|^2 + \beta \sum_{\mathbf{x}} \max\{0, (1 - \Delta(\mathbf{x}_{i,t}, \mathbf{x})) - \mathbf{w}^T(\phi(\mathbf{x}_{i,t}) - \phi(\mathbf{x}))\}$$

Backward Tracking

- State refinement

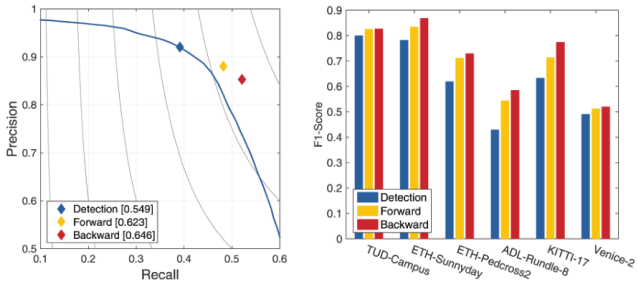


Experimental Results

- MOT challenge 2015 dataset[2]
  - We partition training sequences into two subsets

Training	TUD-Stadtmitte, ETH-Bahnhof, ADL-Rundle-6, PETS09-S2L1, KITTI-13
Validation	TUD-Campus, ETH-Sunnyday, ETH-Pedcross2, ADL-Rundle-8, KITTI-17, Venice-2

- Analysis
  - Detection performance on validation sequences



  - Tracking performance on validation sequences

Setting	MOTA↑	MOTP↑	MT↑	ML↓	FP↓	FN↓	IDS↓
A: Forward w/o	26.8	67.1	12.8	28.2	3,160	13,457	293
B: Forward	40.8	72.5	22.2	27.8	1,516	11,964	180
C: Backward w/o	42.1	72.1	32.1	26.5	2,119	11,072	184
D: Backward	42.4	72.2	32.5	26.5	2,077	11,050	178

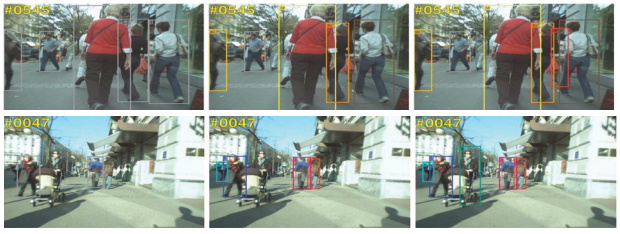
- Comparison Results
  - We use the same detector for comparison

Algorithm	MOTA↑	MOTP↑	MT↑	ML↓	FP↓	FN↓	IDS↓
MHT	39.9	74.7	18.8	38.9	616	13,180	70
JPDA	27.2	72.2	14.1	42.3	2,970	13,625	220
Proposed	42.4	72.2	32.5	26.5	2,077	11,050	178

- Post-processing

Algorithm	MOTA↑	MOTP↑	MT↑	ML↓	FP↓	FN↓	IDS↓
Proposed	42.4	72.2	32.5	26.5	2,077	11,050	178
Proposed+MHT	44.6	73.2	36.3	27.0	2,045	10,639	112

- Tracking results



[1] R. Girshick, "Fast R-CNN," In *ICCV*, 2015.
[2] https://motchallenge.net

MCL Media Communications Lab