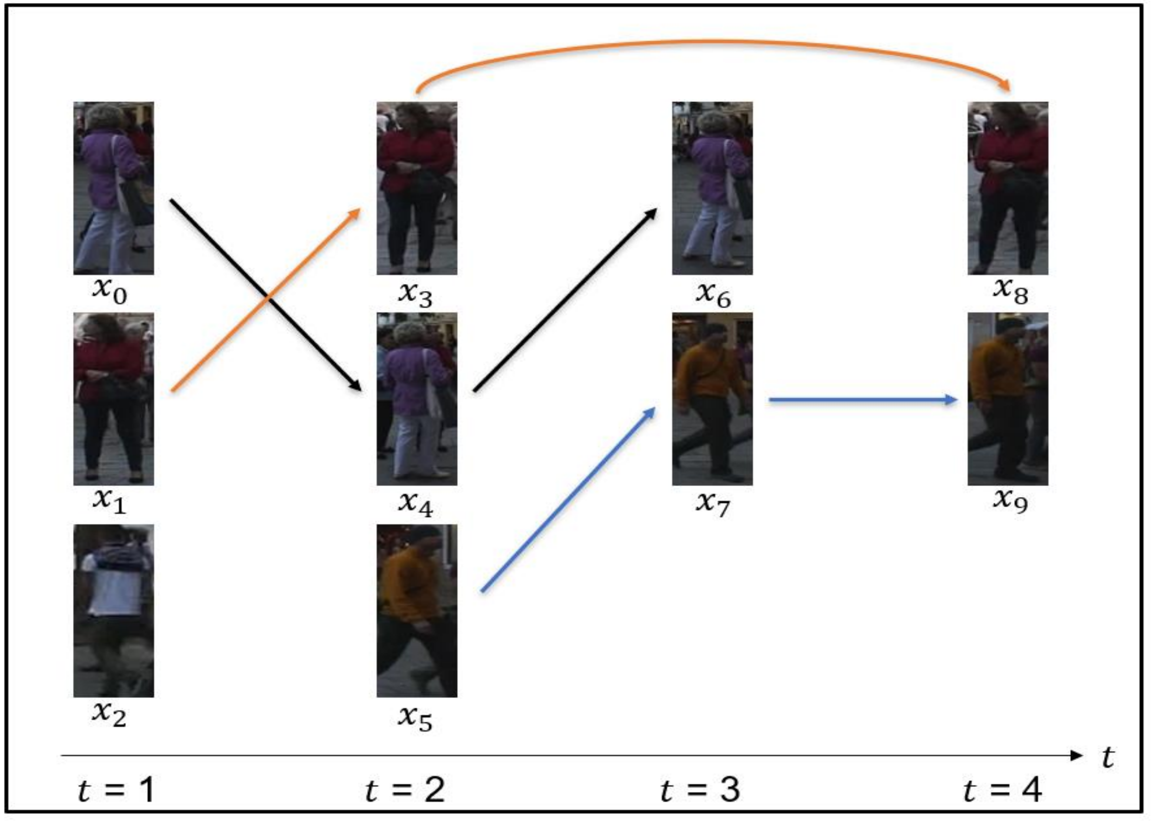
From Certain to Uncertain: Toward Optimal Solution for Offline Multiple Object Tracking

K. Zhao[1] , T. Imaseki[1], H. Mouri[1], E. Suzuki[2], T. Matsukawa[2]
1. Tokyo University of Agriculture and Technology, Japan
2. Kyushu University, Japan

2021/1/14 ICPR 2020

Offline multiple object tracking

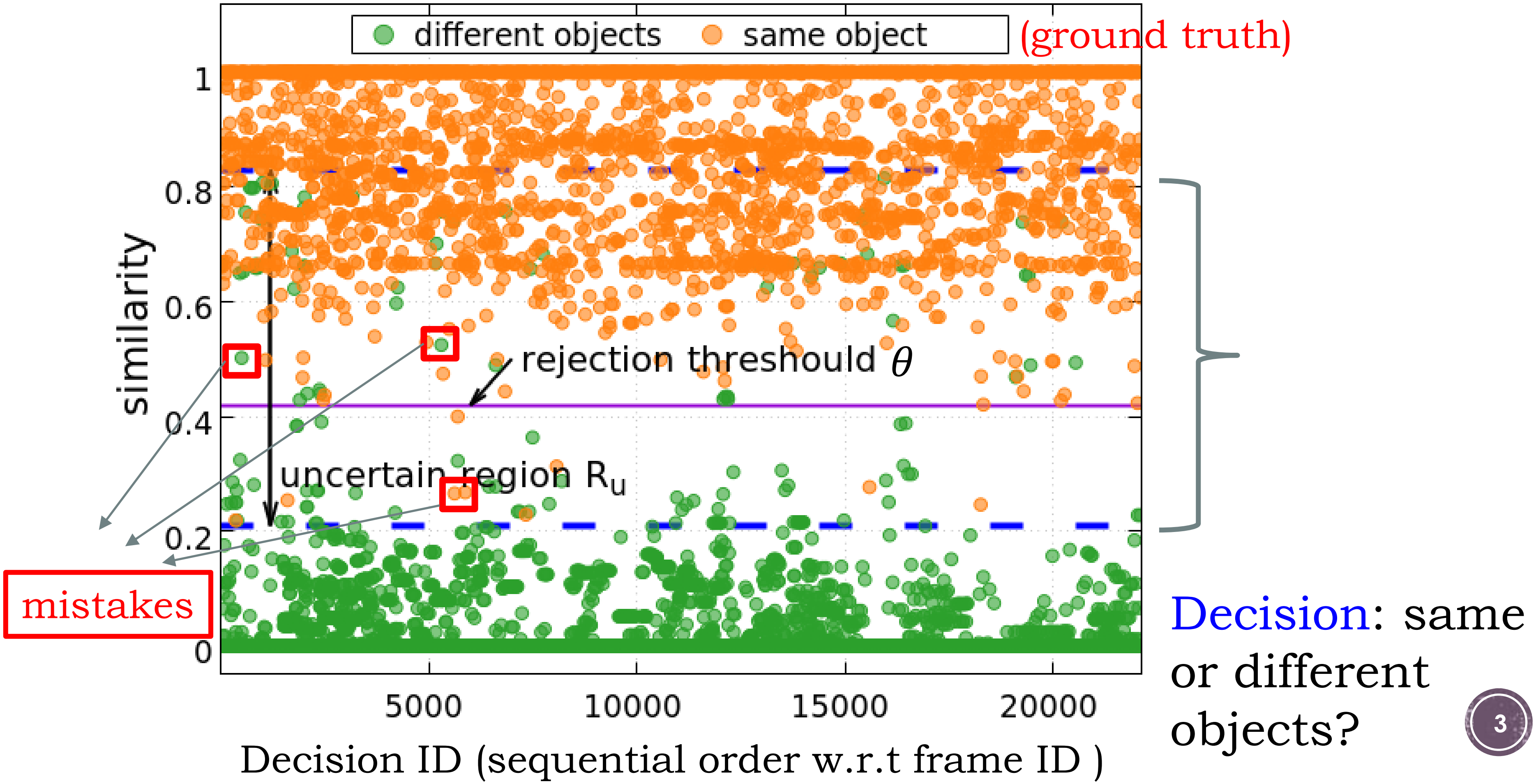
assigning identities for detected objects across a series of frames



same or different objects? -> affinity measure

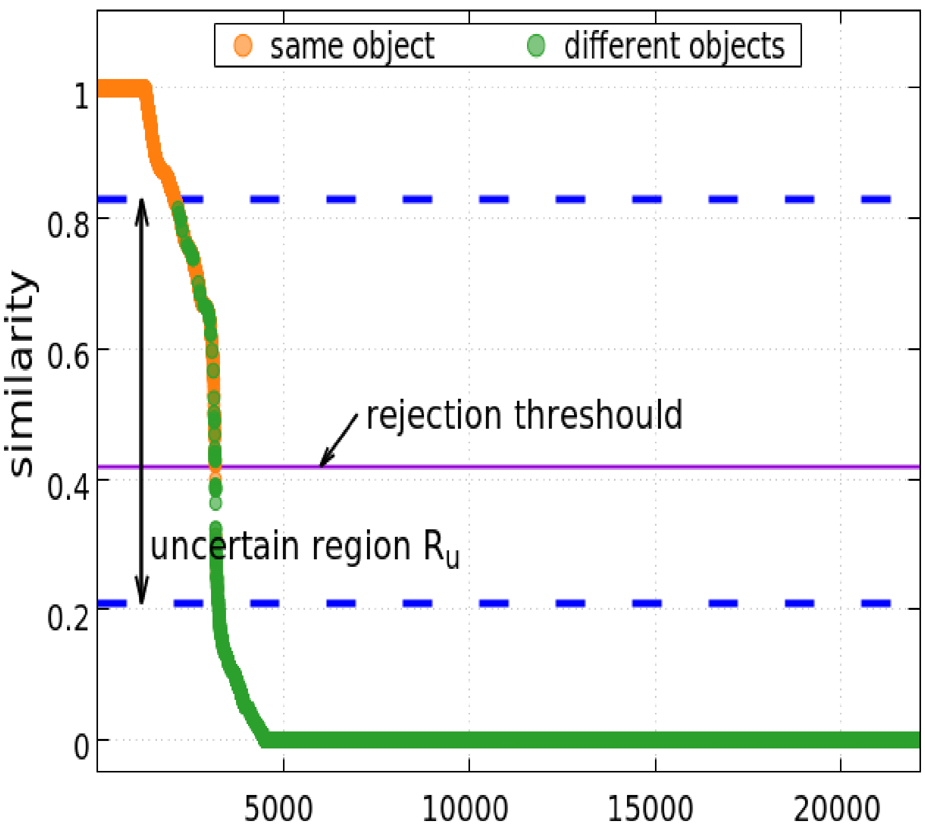
Uncertain region & early mistake issues

- imperfect affinity measure -> uncertain region -> threshold θ
- sequential tracking with pre-decided θ -> early mistakes



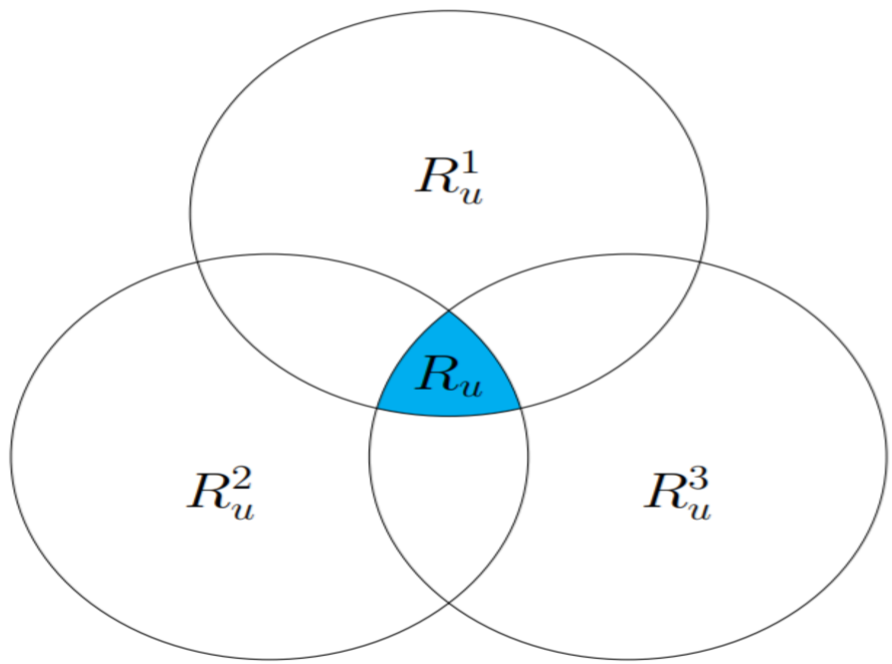
Ideas to reduce early mistakes & uncertain region

tracking from certain to uncertain



Decision ID (decreasing certainty)

ensemble multiple tracking experts

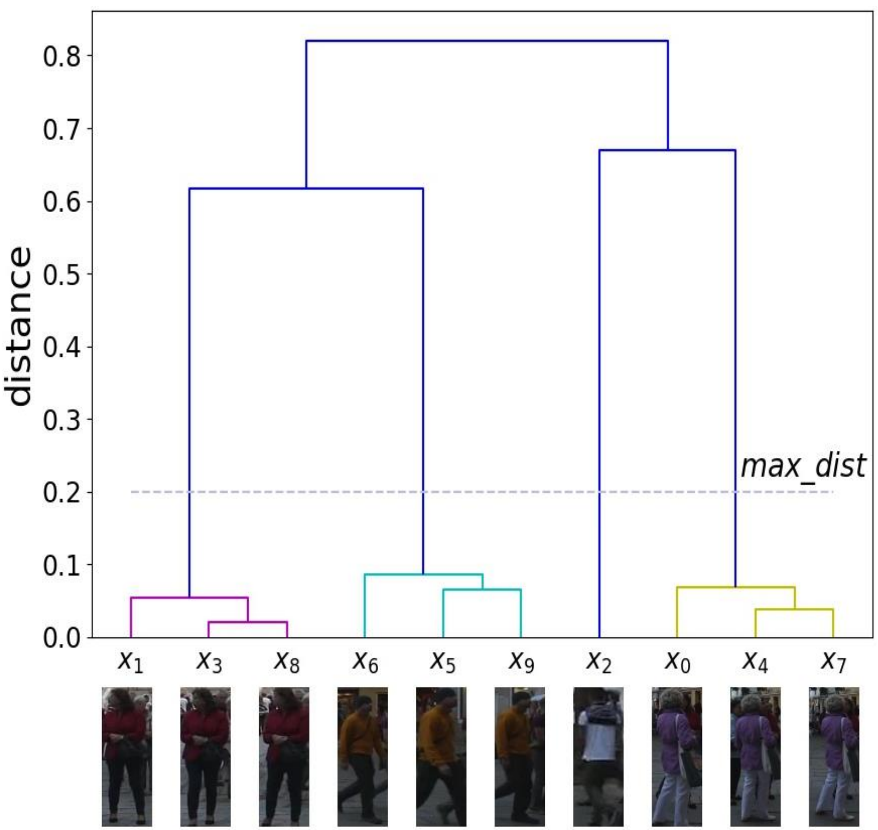


$$R_u = R_u^1 \cap R_u^2 \cap R_u^3$$

R_u^k: uncertain region of kth expert

Proposal: implementing the idea of tracking from certain to uncertain

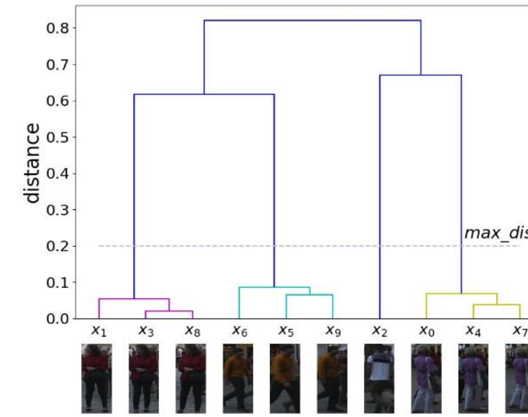
Agglomerative Hierarchical Clustering (AHC, [Day and Edelsbrunner, 84]): grouping similar observations into clusters



$(x_3) + (x_8)$	0.02
$(x_4) + (x_7)$	0.04
$(x_1) + (x_3,x_8)$	0.06
$(x_5) + (x_9)$	0.07
$(x_0) + (x_4,x_7)$	0.07
$(x_6) + (x_5,x_9)$	0.09
$(x_1,x_3,x_8) + (x_5,x_6,x_9)$	0.62
$(x_2) + (x_0,x_4,x_7)$	0.67
$(x_1,x_3,x_5,x_6,x_8,x_9) + (x_0,x_2,x_4,x_7)$	0.82

Pros: 1) merging tracks with strictly increasing distance, 2) considering all tracks in each iteration

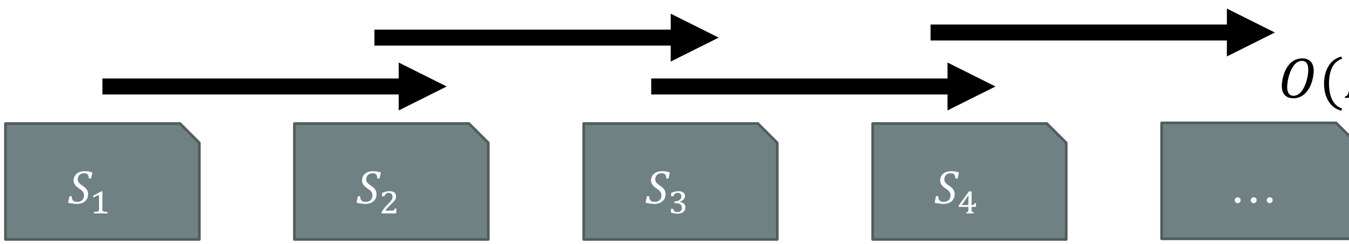
Adapting AHC for object tracking



Cons of AHC:

1) memory complexity: $O(N^2)$, where N: #detections

→ dividing sequence S into $S_1, S_2, \ldots$

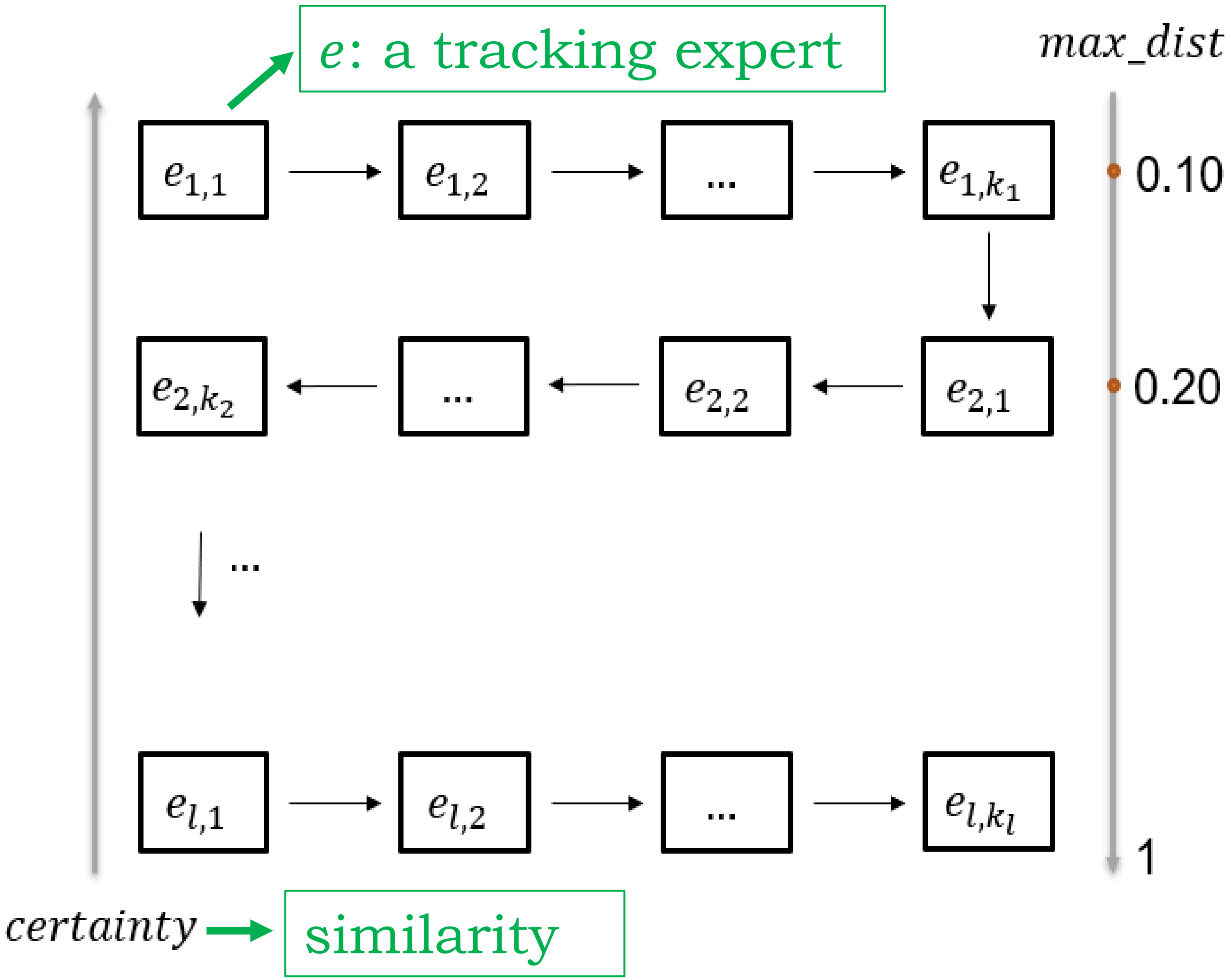


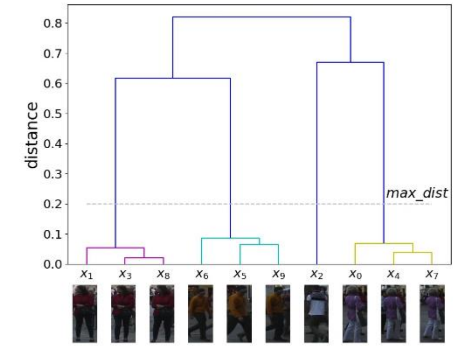
$O(N^2)$ -> $O(N_i^2)$

2) spatiotemporal constraint: detections in the same image should not belong to the same track

→ building cannot-link constraints

AHC with ensemble of tracking experts (AHC_ETE)





incrementally build the dendrogram with each expert contributing its most certain mergings in turn

need to define es; distance measure, max_*dist* for each e

Defined distance measures

an integration of appearance, motion and temporal distances

$$dist(T_u, T_v) = dist_{major}(T_u, T_v) * F_1(\cdot) * F_2(\cdot) * \cdots$$

major distance measure (appearance)

filter for other distance measures (motion, temporal)

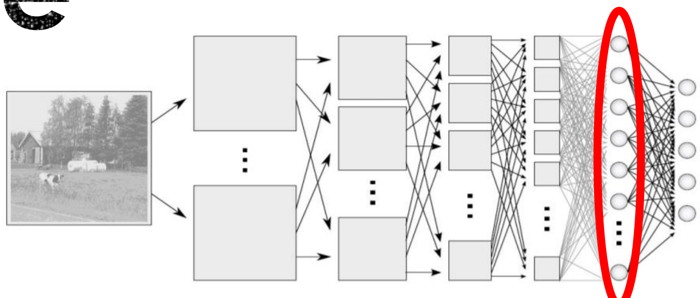
where

$$F(v, condition) = \begin{cases} 1, & \text{if } v \text{ satisfies } condition \\ inf, & \text{else} \end{cases}$$

for imposing cannot-link constraints

Appearance distance

for two detections x_i, x_j:



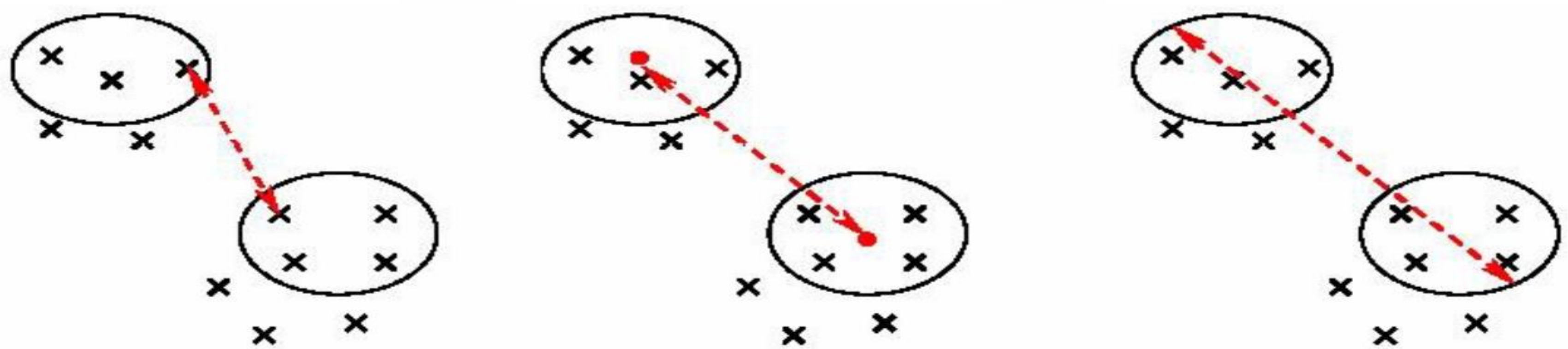
https://www.cs.colostate.edu/~cs440/fall2015/more_progress/DCNN.pdf

$$dist_{appe}(x_i, x_j) = 1 - \frac{a_i^T a_j}{\|a_i\| \, \|a_j\|}$$

where a_i extracted CNN feature vector (128-dim, output of the penultimate layer) of x_i

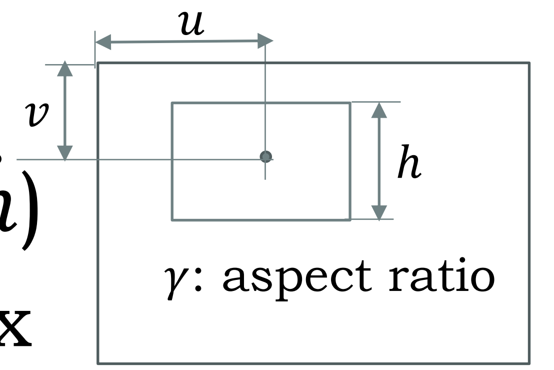
for two tracks T_u, T_v

- Simple linkage
- Average linkage
- Complete linkage



https://medium.com/datadriveninvestor/hierarchical-clustering-514b9d1aa2c1

Motion (Kalman filter) distance



state of object: $(u, v, \gamma, h, \dot{u}, \dot{v}, \dot{\gamma}, \dot{h})$

centers, aspect ratio, height of a bbox

[Wojke et al., 17]

$$dist_{kf}(T, x) = \sqrt{(y - \hat{y})^T \Sigma^{-1} (y - \hat{y})}$$

T: a track; x: a single detection;

y: detection transferred to the measurement space;

$\hat{y}$: prediction of Kalman Filter.

Temporal distance

$$dist_{temp}(T_u, T_v) = \begin{cases} |\Gamma_u \cap \Gamma_v| - |\Gamma_u \cup \Gamma_v| & \text{if } \Gamma_u \cap \Gamma_v \neq \emptyset \\ min(\Gamma_v) - max(\Gamma_u) & \text{elseif } max(\Gamma_u) < min(\Gamma_v) \\ min(\Gamma_u) - max(\Gamma_v) & \text{elseif } max(\Gamma_v) < min(\Gamma_u) \\ 0 & \text{else} \end{cases}$$

Γ_k: set of frame IDs for detections in track T_k

frame IDs overlap → negative value;

one track appears later than another → closest frame gap;

no overlap & not earlier, later tracks → 0

Defined tracking experts

1. Preprocessing: build T_{fp} for detections with score ≤ 0.3 or suppressed by NMS with threshold 0.1; impose cannot-links for T_{fp}, i.e., for any track T_k, $dist(T_k, T_{fp}) = inf$

2. Connecting detections to tracks: track with complete linkage (expert e_1), then single linkages (e_2 and e_3) → remove cannot-links on T_{fp} and track with weak constraints (e_4 and e_5)

3. Post-processing: remove T_k if $|T_k| < 3$

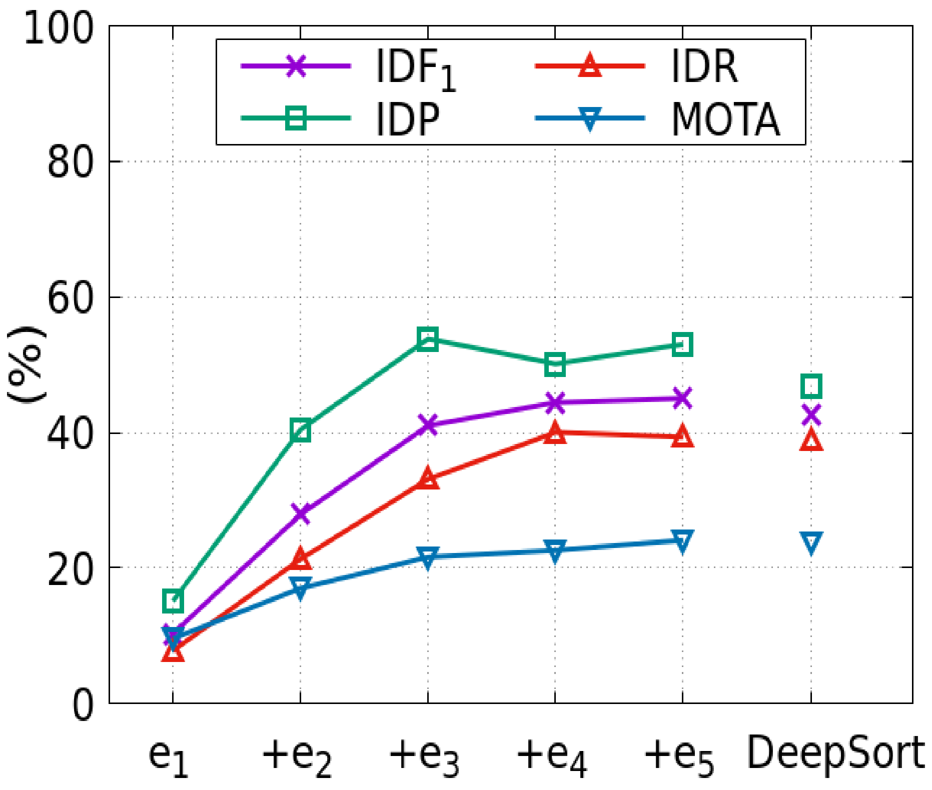
E	$dist_{appe}$	$F_1(temp)$	$F_2(kf)$	$F_3(appe)$	max_dist
e_1	complete	≥ 0	complete:< 9.5	-	0.10
e_2	single	≥ 0	-	-	0.05
e_3	single	≥ 0	complete:< 9.5	-	0.10
e_4	single	≥ 0	complete:< 9.5	-	0.10
e_5	single	≥ 0	average:< 9.5	complete:≤ 0.30	0.20

Design of experiments

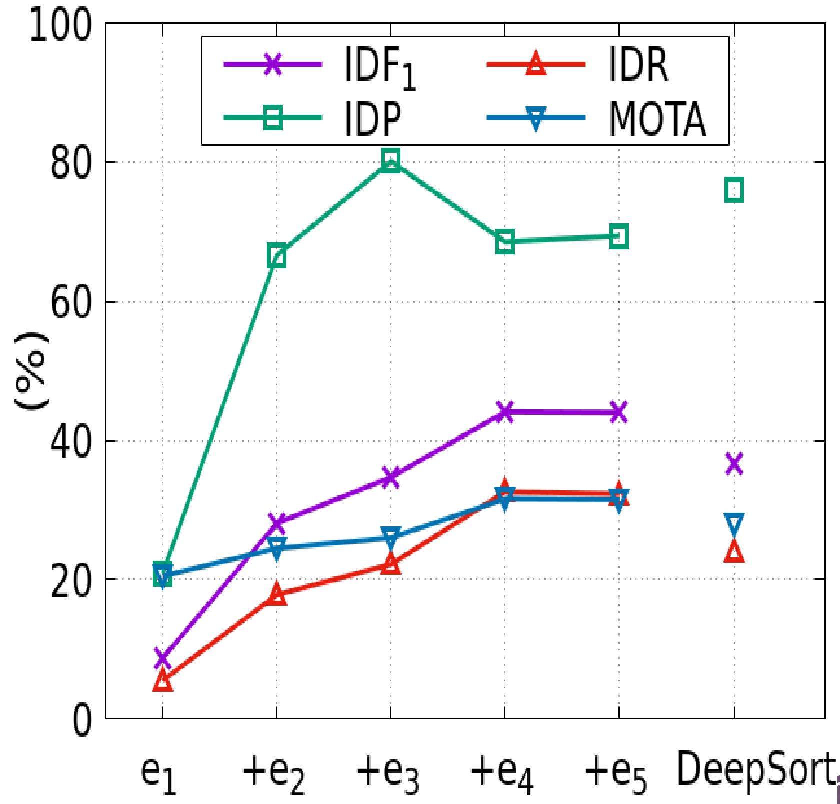
- Dataset: MOT15, MOT16 [Milan et al., 16] training sequences
- Evaluation metrics: multiple object tracking accuracy (MOTA [Bernardin and Stiefelhagen, 08]), identification precision (IDP), recall (IDR), corresponding F1 score (IDF1 [Ristani et al., 16])
- Benchmark method: Deep Sort [Wojke et al., 17] (same features, appearance and motion distances)

Effects of merging order

- our method generally outperforms Deep Sort [Wojke et al., 17]
- IDF1s, IDPs, IDRs and MOTAs generally increase as more experts integrated



MOT15

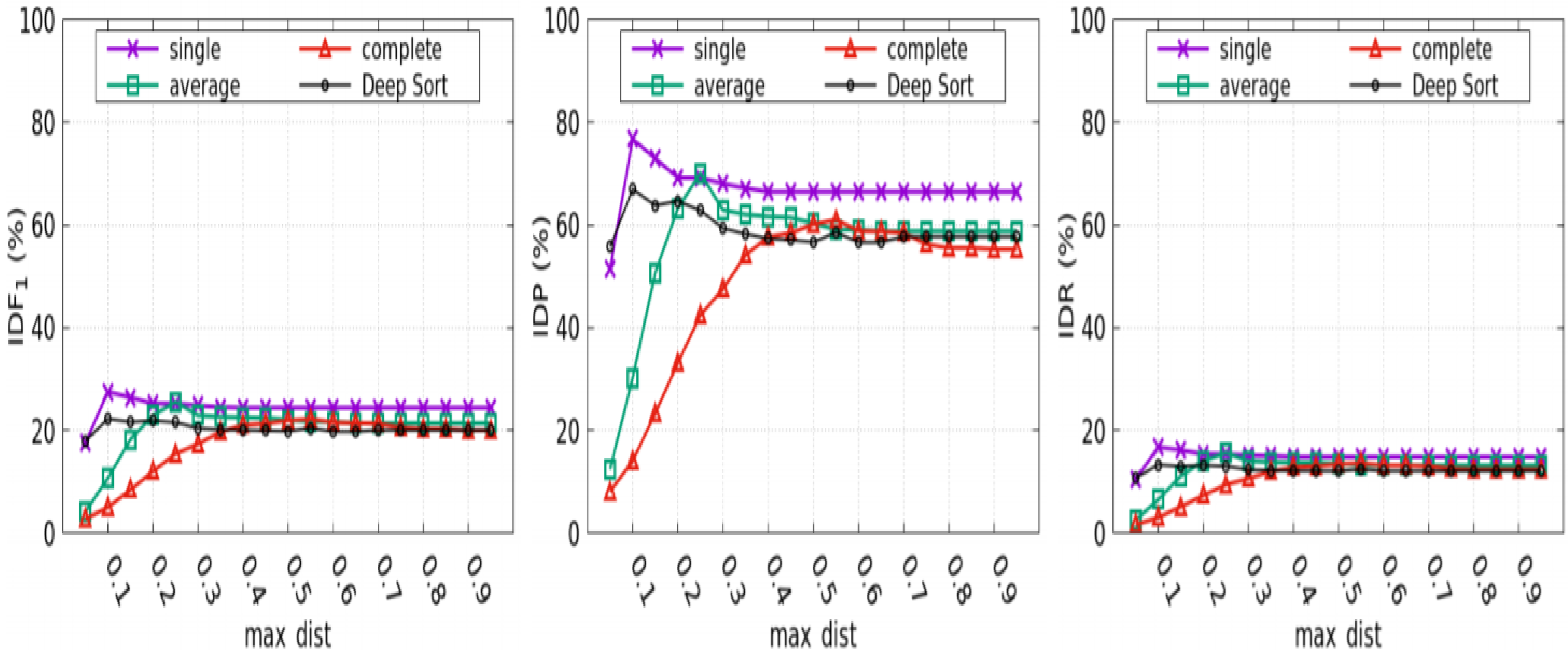


MOT16

Effects of different linkages

best IDF1s, IDPs and IDRs differ;

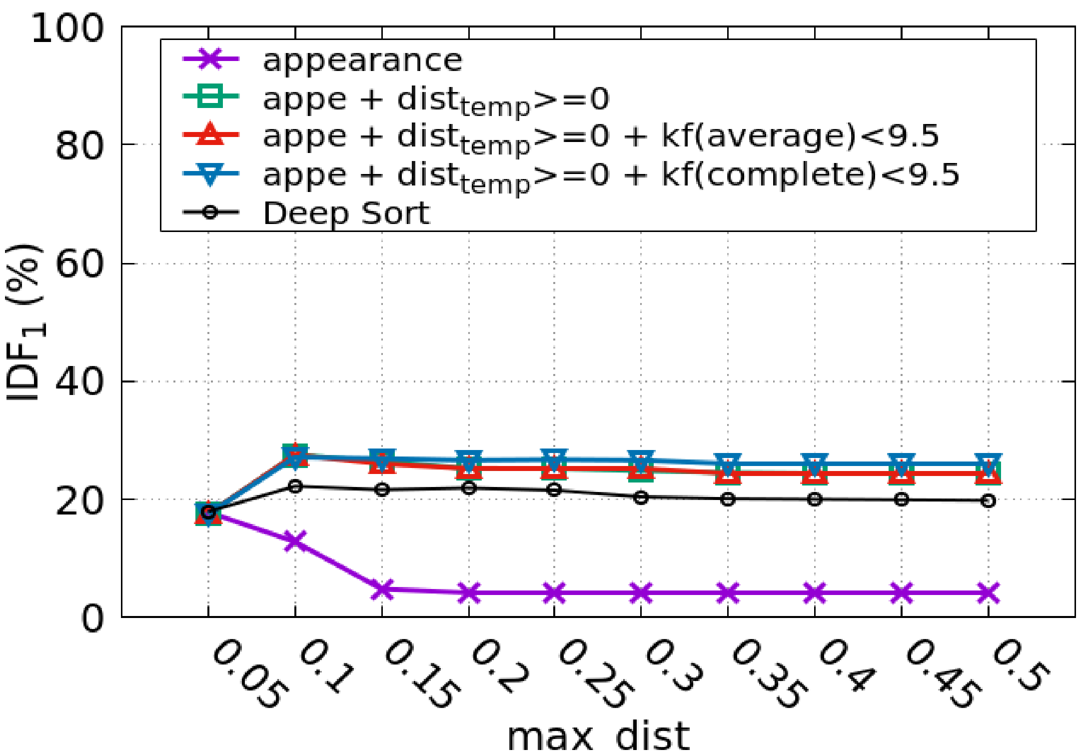
certain region of single < average < complete linkage



standard AHC [Day and Edelsbrunner, 84] based tracking;
Test data: MOT16-02

Effects of imposing cannot-link constraints

only appearance -> deteriorates significantly when max_*dist* increases for *single* linkage;
with temporal, Kalman Filter constraints -> IDF1 generally increases



standard AHC [Day and Edelsbrunner, 84] based tracking;
Test data: MOT16-02

Conclusions

Tackling two typical issues for object tracking: 1) uncertain region, 2) early mistakes

Proposed AHC_ETE: tracking from certain to uncertain, ensemble multiple tracking experts

(a general framework for various distance measures and tracking experts)

Code: cyoukaikai / ahc_ete

GitHub

Limitations and future work

- accepted all the progress made by the earlier tracking experts as the starting point of the later ones

  -> proposed algorithm sensitive to the ordering of experts

- further experiments comparing with the state-of-the-art methods needed