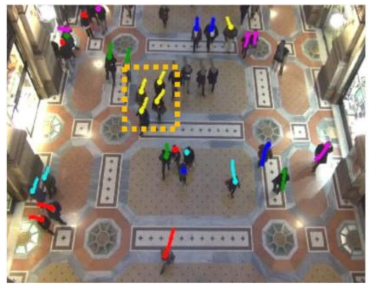
A Novel Multi-Detector Fusion Framework for Multi-Object Tracking

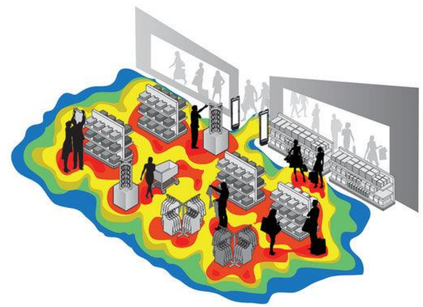
Dipl.-Math. Roberto Henschel

henschel@tnt.uni-hannover.de

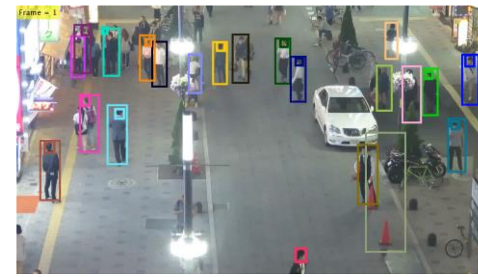
Motivation Tracking



[1]



[2]



Sport Analysis

Application:
- Obtain statistics
- Improve performance

Sociology/Economy

Application:
- Urban planning
- Optimize product placement

Surveillance

Application:
- Unusual behaviour
- Autonomous driving

[1] **Binary Quadratic Programing for Online Tracking of Hundreds of People in Extremely Crowded Scenes, Deghan et.al, CoRR 2016**

[2] http://tickto.weebly.com/uploads/4/2/0/5/42051809/5835638_orig.jpg

Standard Paradigm

Tracking-by-Detection

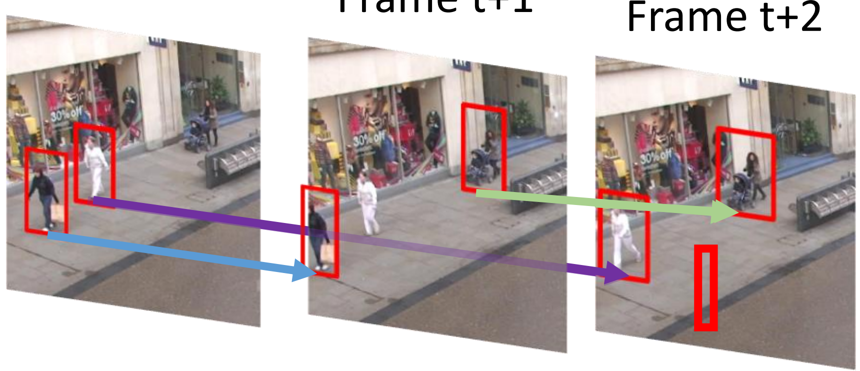
Step 1: Detect

Step 2: Track

Frame t

Frame t+1

Frame t+2



Tracking-by-Detection Paradigm

- Solve $\max_T \Pr(T|D)$ for detections D
- Information reduction by full-body boxes
  - Efficient calculation
- BUT: Usefull information might be lost

Occlusion

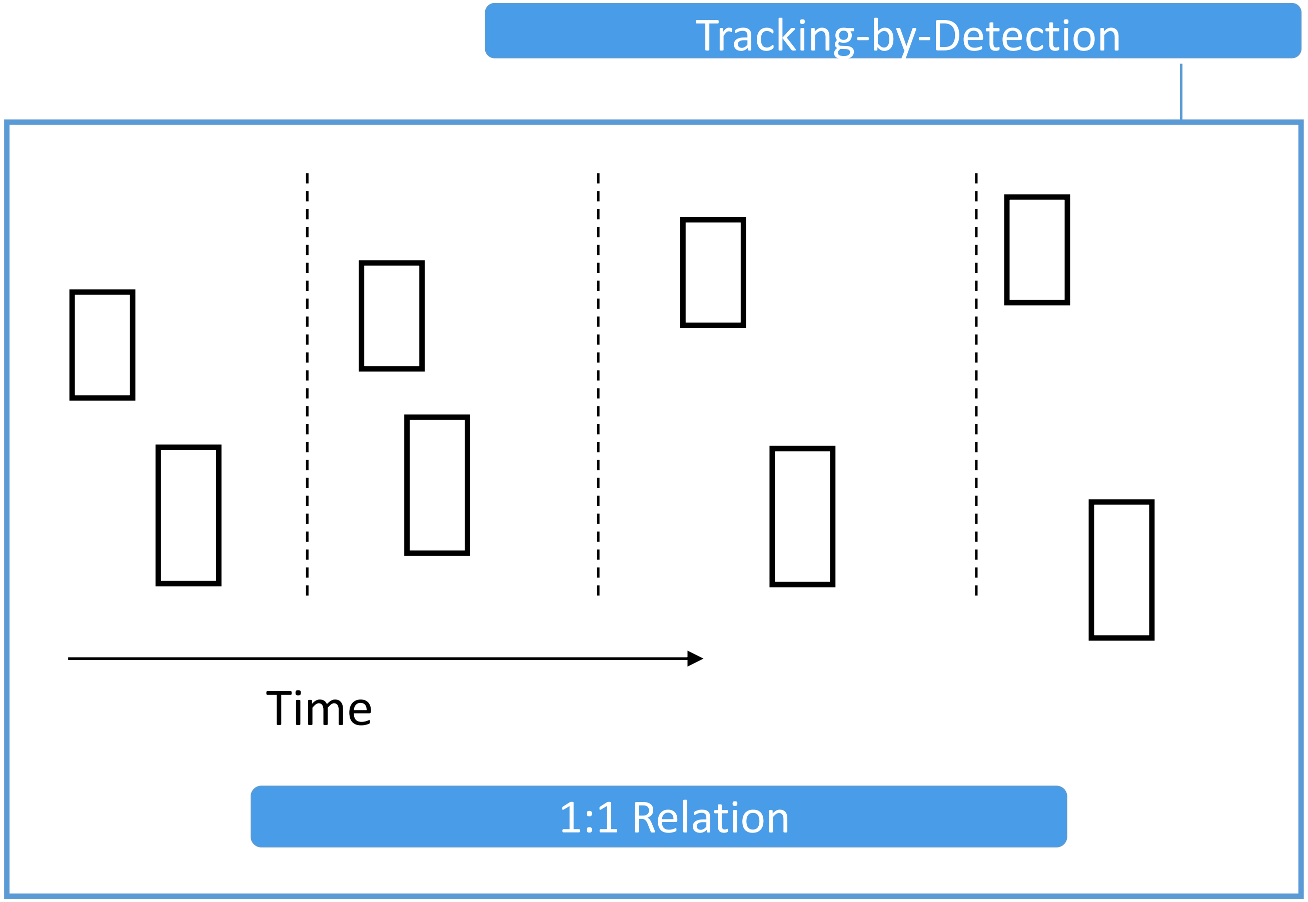
Missed detection

Rare pose

- Intrinsic Problem
- Solution via feature fusion

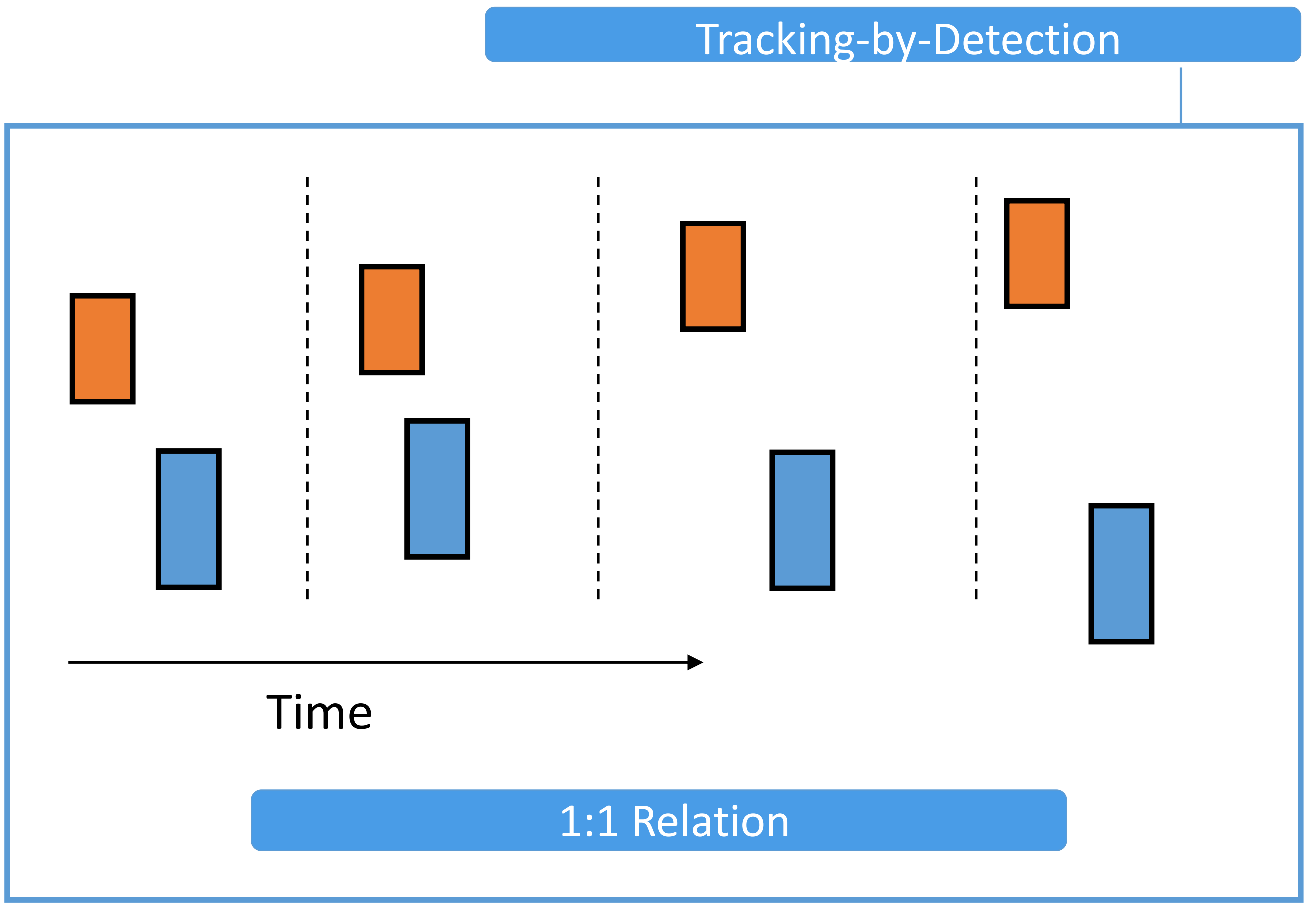
Standard Assignment Problem

TbD Ensure temporal consistency



Standard Assignment Problem

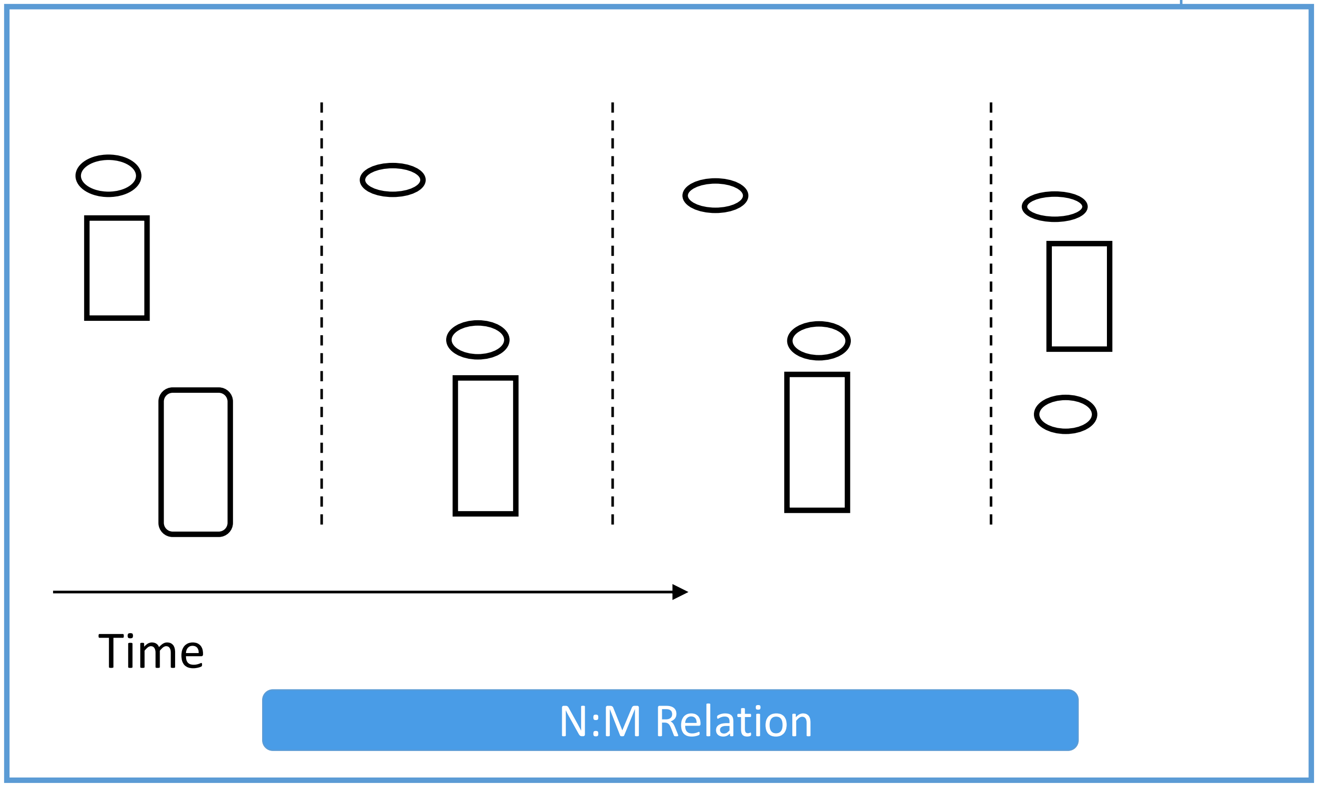
TbD Ensure temporal consistency



Fusion of Input Features

Track-by-X Ensure temporal + spatial consistency

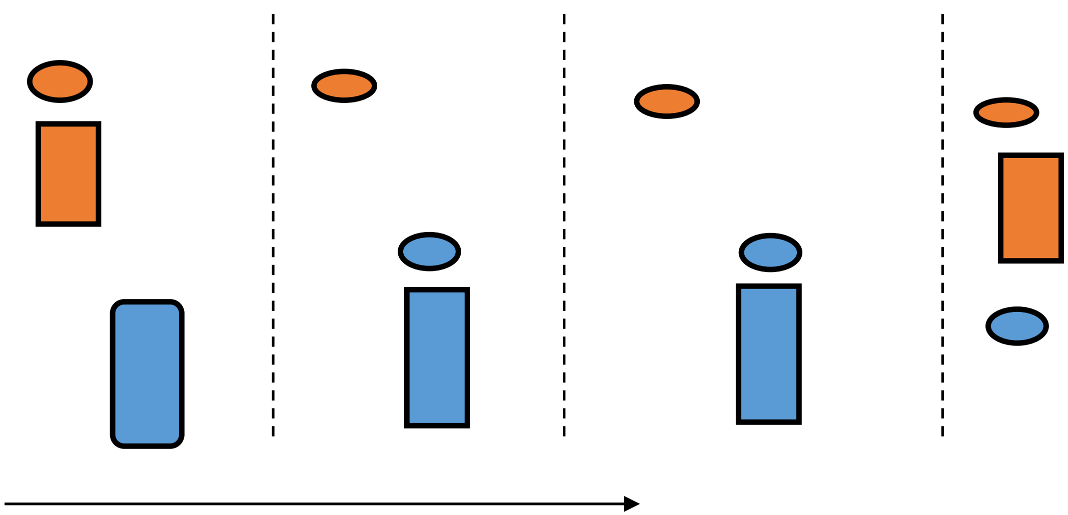
Multi-Feature Tracking



Fusion of Input Features

Track-by-X Ensure temporal + spatial consistency

Multi-Feature Tracking



Time

N:M Relation

Which Feature to use ?

- Head detections[1]
  - Less prone to occlusion.
  - Less prone to pose variation.
  - Sparse

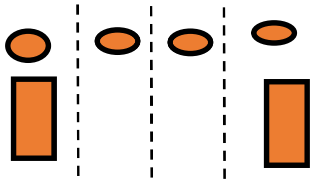
[1] R. Stewart, M. Andriluka, and A. Y. Ng. End-to-end peopledetection in crowded scenes. In CVPR, 2016.

People Tracking via Feature Fusion

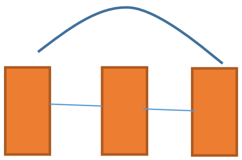
Approach **Fusion of multiple input features**

Concept

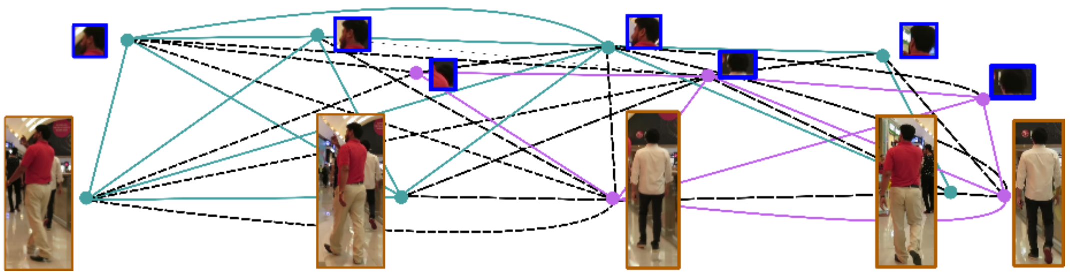
- Any feature can be used to create a trajectory



- Ensure consistency **within all** time steps
- Ensure consistency within each frame

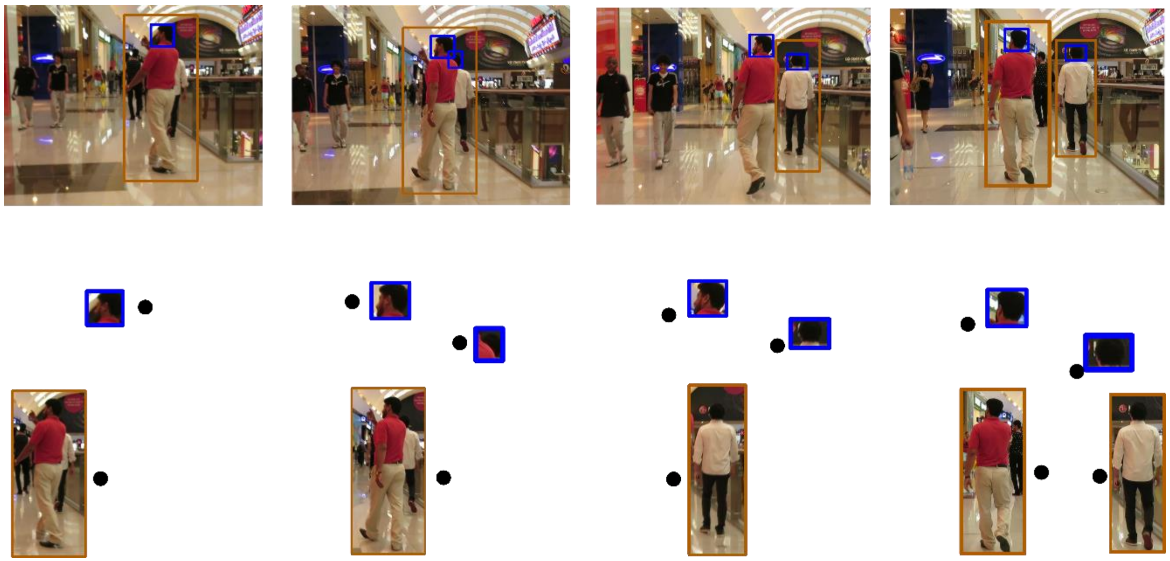


Solution Best fusion via Graph Labeling



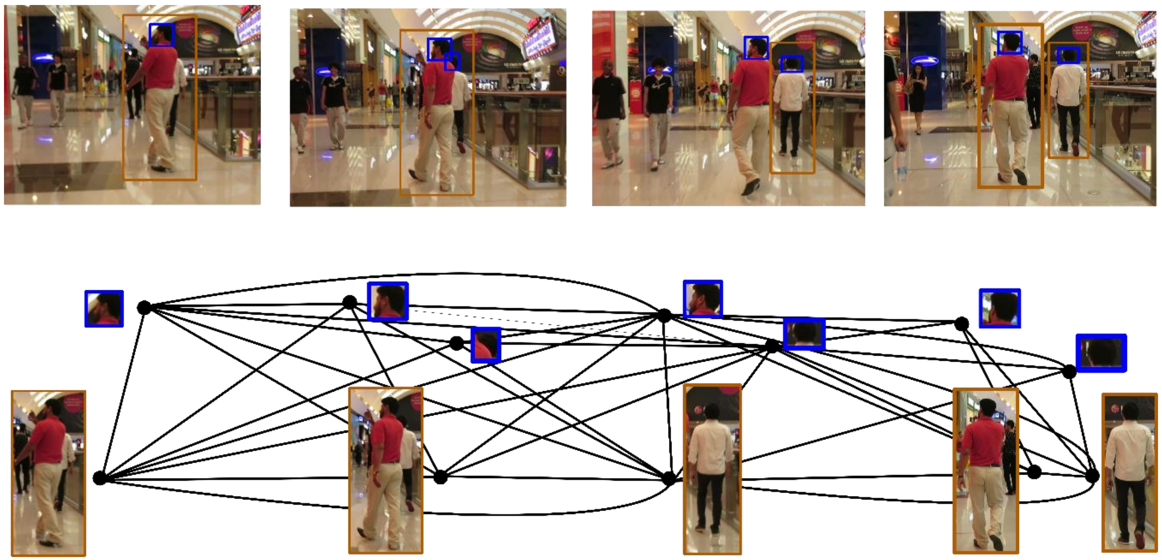
Feature Fusion

- Multi-Feature Tracking via Graph Labeling



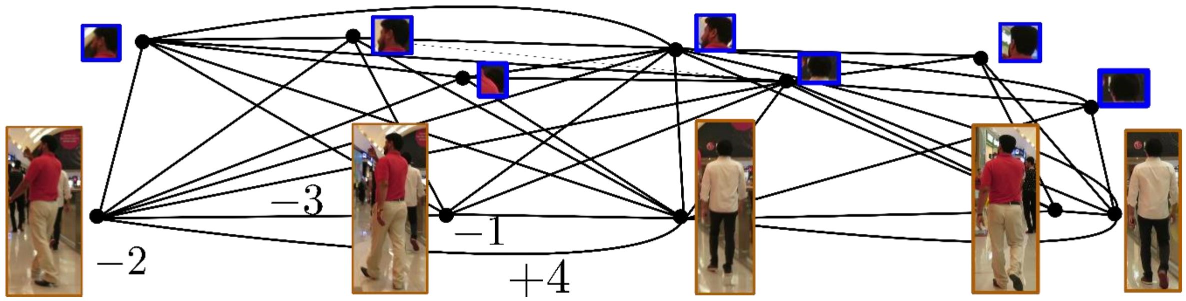
Feature Fusion

- Multi-Feature Tracking via Graph Labeling



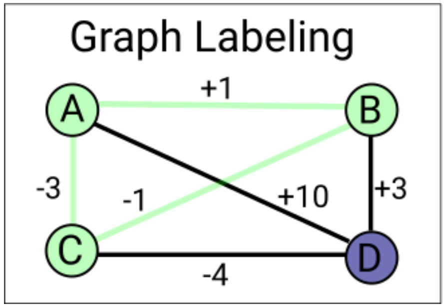
Feature Fusion

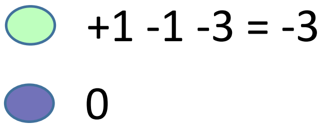
- Multi-Feature Tracking via Graph Labeling
  - Each node has a weight, reflecting probability of the feature.
  - Each edge has a weight, reflecting how likely the edge connects two features that belong to the same person.



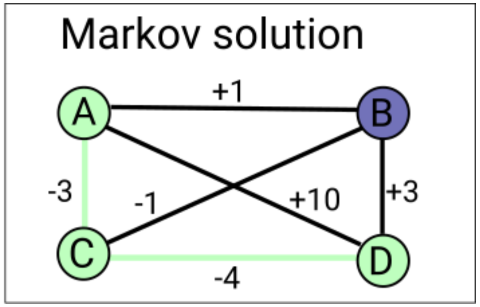
Tracking via Graph Labeling

- Goal: Find minimizing labeling



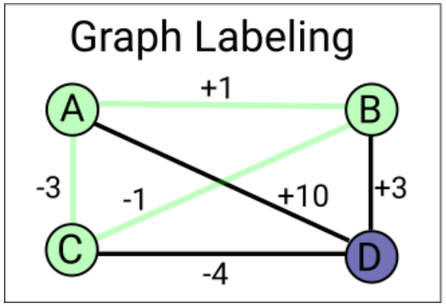


- Former trackers: 1st order Markov chain
  - current position depends only on last position



A and D not compatible!

Long-term Consistency



Result

We achieve consistency to **all** equally labeled nodes!

Model

$$\arg\min_x f(x) := \sum_{l=1}^{K} \left(\sum_{v \in V} c_v x_v^l + \sum_{\{u,v\} \in E} q_{u,v} x_u^l x_v^l \right)$$

s.th. $\forall v \, : \sum_{l=1}^{K} x_v^l \leq 1, \; x \in \{0,1\}^{|F| \cdot K}$

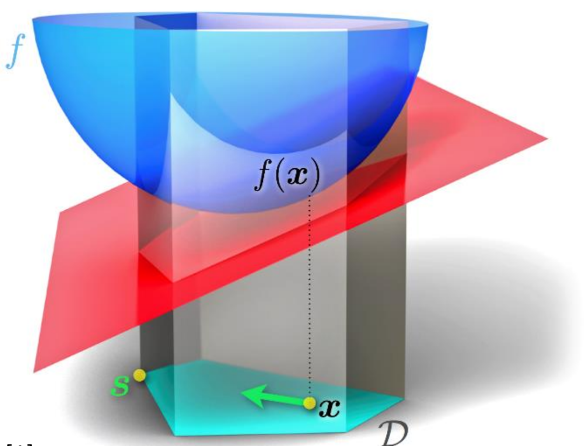
$|F| = \text{Number of features} \quad K = \text{max. number of Labels}$

x_v^l : Assignment of label l to node v

$c = (c_v)$: Node affinity $\; Q = (q_{u,v})$: Edge affinity

Solving a BQP

- The graph labling problem is a BQP (Binary Quadratic Problem)
- NP-hard to solve!
- We approximate it using the Frank-Wolfe algorithm on the relaxation (delivers local optimum).



[1]

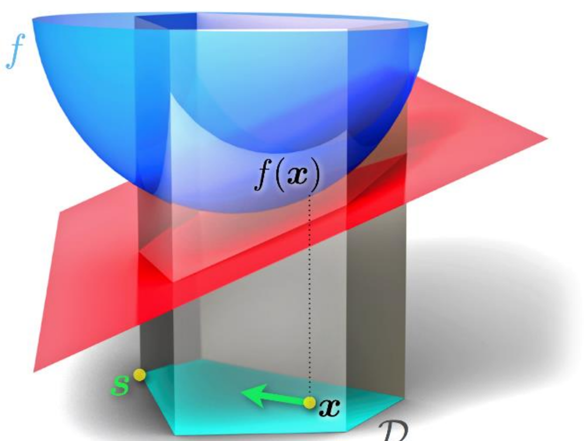
$$\arg\min_x f(x) := \sum_{l=1}^{K} \left(\sum_{v \in V} c_v x_v^l + \sum_{\{u,v\} \in E} q_{u,v} x_u^l x_v^l \right)$$

s.th. $\forall v \, : \sum_{l=1}^{K} x_v^l \leq 1, \, x \in \{0,1\}^{|F| \cdot K}$

[1] https://upload.wikimedia.org/wikipedia/commons/e/e5/Frank-Wolfe_Algorithm.png

Solving a BQP

- The graph labling problem is a BQP (Binary Quadratic Problem)
- NP-hard to solve!
- We approximate it using the Frank-Wolfe algorithm on the relaxation (delivers local optimum).



[1]

$$\arg\min_x f(x) := \sum_{l=1}^{K} \left(\sum_{v \in V} c_v x_v^l + \sum_{\{u,v\} \in E} q_{u,v} x_u^l x_v^l \right)$$

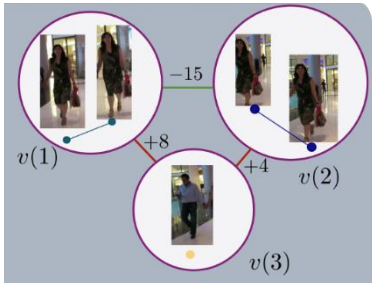
s.th. $\forall v : \sum_{l=1}^{K} x_v^l \leq 1,\ x \in \cancel{\{0,1\}^{|F| \cdot K}}$

$[0,1]^{|F|K}$

[1] https://upload.wikimedia.org/wikipedia/commons/e/e5/Frank-Wolfe_Algorithm.png

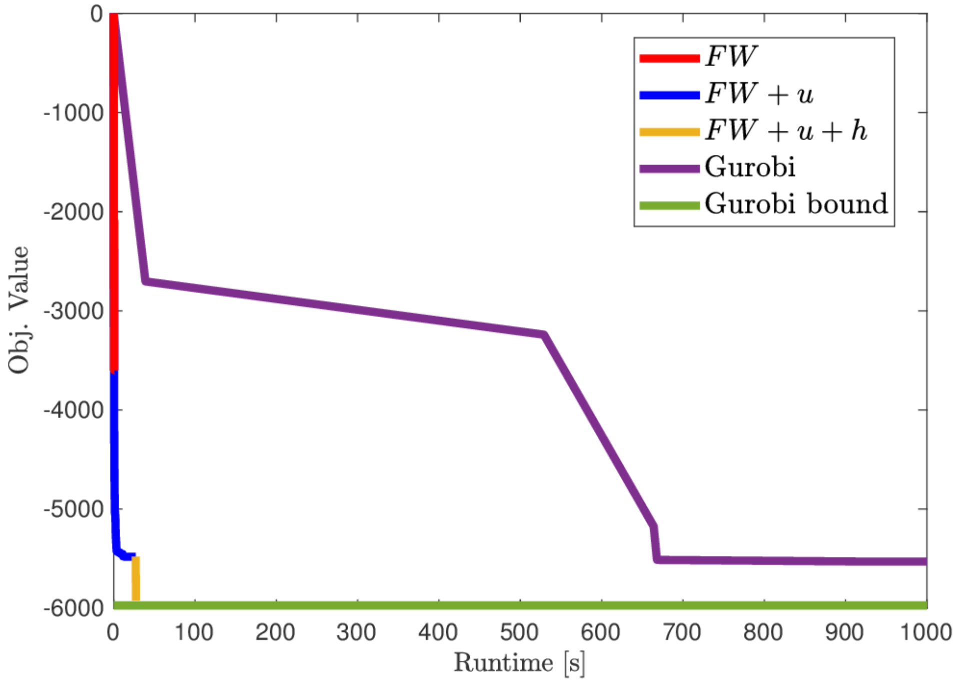
Improving the BQP Solver

- We improved the standard Frank-Wolfe via
  - Speed up
  - Regularization
  - Hierarchical improvement
- Experiment: Our BQP solver vs Gurobi
  - 40 frames
  - Only body detections
  - K=70 labels allowed
  - 400 detections
  - >28.000 variables



Improving the BQP Solver

Minimization quality



Full evaluation

	Time[sec]	Obj. Value
FW	0.7	-3060
FW+u	27	-5481
FW+u+h	27.5	-5925
Gurobi	1000	-5531
Bound	1000	-5973

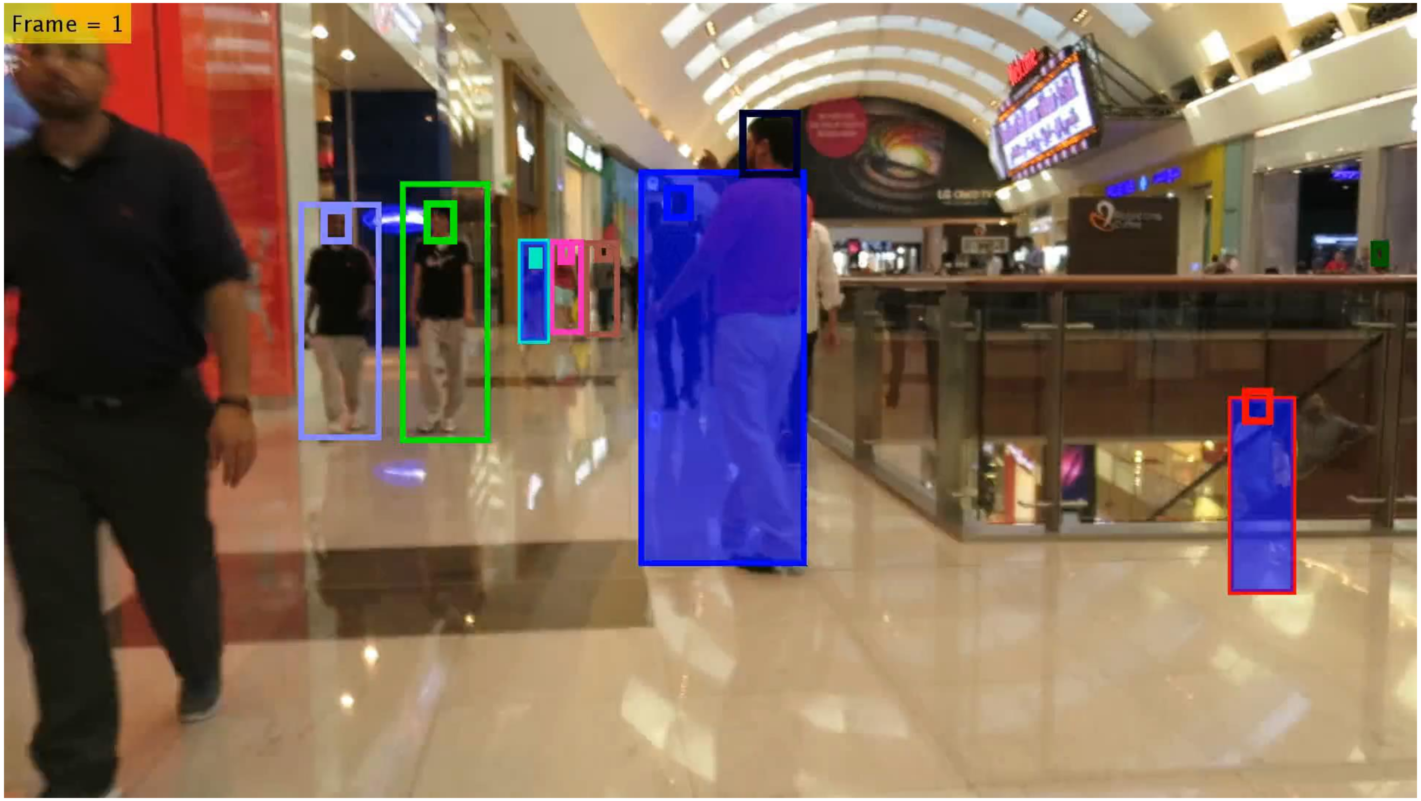
Experiment Feature Fusion

- Evaluation on MOT16 train set
  - 5300 frames
  - 110.000 body boxes
  - 517 trajectories
- Metrics
  - MT = number of mostly tracked trajectories ($\geq 80\%$ of the track is correct)
  - FP = numer of false positive detections
  - MOTA = multiple object tracking accuracy, incorporates: missing detections, identity switches, false positives.

	MOTA $\uparrow$	MT $\uparrow$	FP $\downarrow$
Body	33.0	76	11949
Body+Head	38.2	86	4972

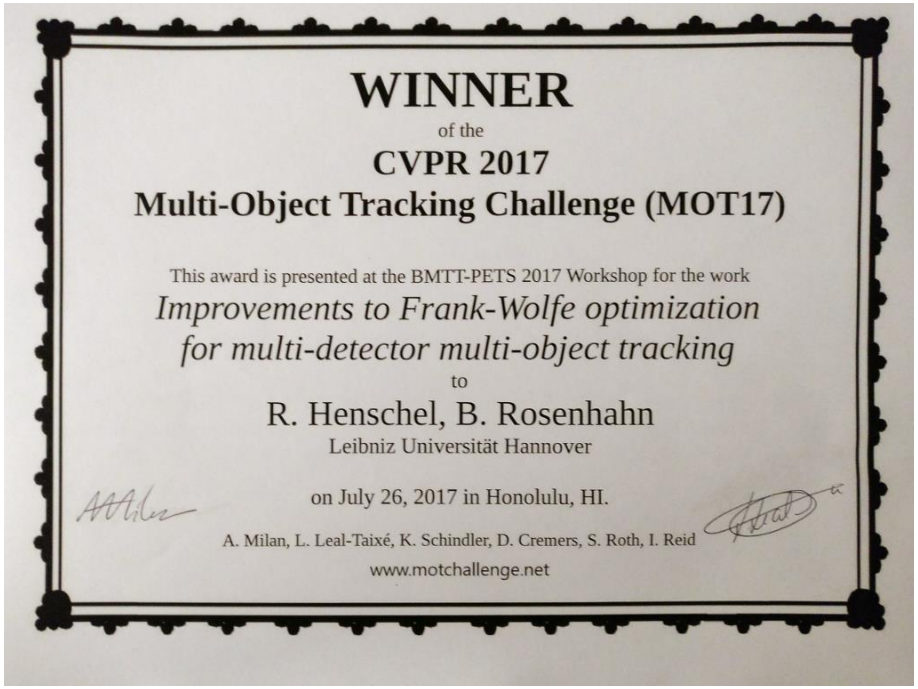
Experiment Feature Fusion

Visual result



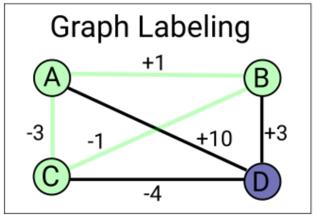
Tracking Challenge

- 7 Sequences, 3 different detectors
- Compared to other methods we perform state-of-the-art



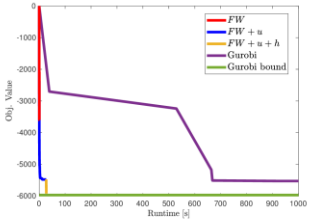
Summary

- Tracking via **Graph Labeling**
  - Long-term temporal consistency
  - State-of-the-art results



- Feature Fusion via labeling
  - Fusion improves results considerably

- Solver for labeling problem
  - More efficient than Gurobi
  - Can be applied to other labeling problems



Experiment Test Set Evaluation

- Evaluation on MOT16 test set
  - 5900 frames
  - 180.000 body boxes
  - 760 trajectories

- Evaluation on MOT17 test set
  - 17000 frames
  - 560.000 body boxes
  - 2300 trajectories

Experiement Test Evaluation

- MOT16
  - We rank 2nd

	MOTA	MT	ML	FP	FN
[1]	48.8	18.2	40.1	6654	86245
Ours	47.8	19.1	38.2	8886	85487
[2]	47.6	17.0	40.4	5844	89093

- Metrics
  - MT = number of mostly tracked trajectories ($\geq 80\%$ of the track is correct)
  - ML = number of mostly lost trajectories ($\leq 20\%$ of the track is correct)
  - FP = numer of false positive detections
  - FN = numer of false negative detections
  - MOTA = multiple object tracking accuracy, incorporates: missing detections, identity switches, false positives.

Experiement Test Evaluation

- MOT17
  - We rank 1st

	MOTA	MT	ML	FP	FN
Ours	51.3	21.4	35.2	24101	247921
[1]	50.7	20.8	36.9	22875	252889
[2]	50.0	21.6	36.3	32279	247297

- Metrics
  - MT = number of mostly tracked trajectories ($\geq 80\%$ of the track is correct)
  - ML = number of mostly lost trajectories ($\leq 20\%$ of the track is correct)
  - FP = numer of false positive detections
  - FN = numer of false negative detections
  - MOTA = multiple object tracking accuracy, incorporates: missing detections, identity switches, false positives.

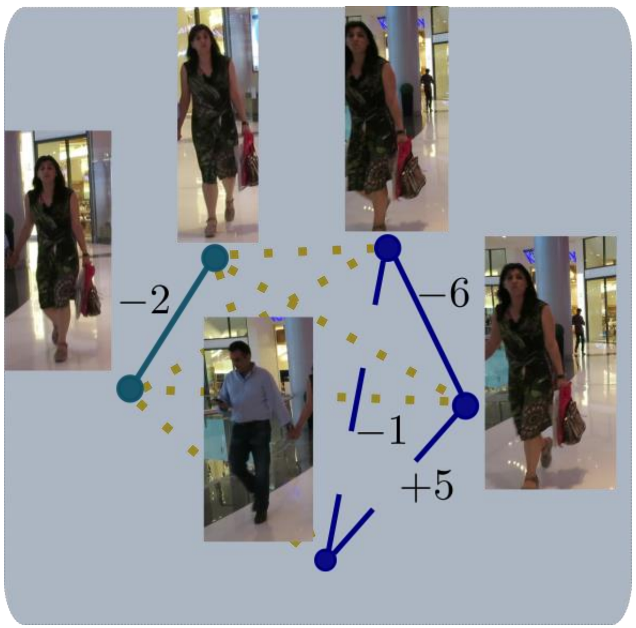
Regularization

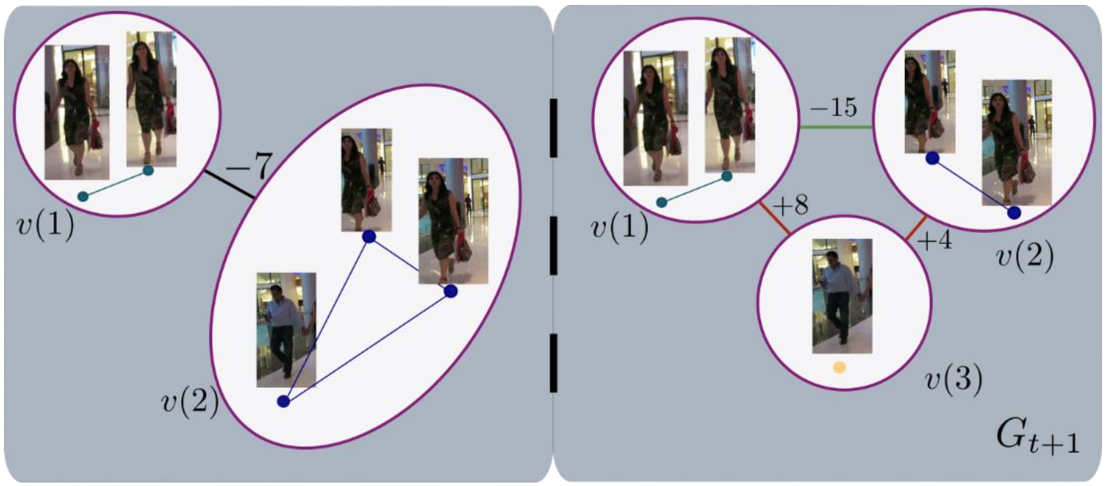
- Improvement of Frank-Wolfe optimization via

$$f_u(x) = f(x) + u\sum_i (x_i^2 - x_i)$$

- Note that $f_u(x) = f(x)$, if x is binary
- The function $(x_i^2 - x_i)$ is minimal at 0.5
- The function $-(x_i^2 - x_i)$ is minimal at 0 and 1, within [0,1].
- For u<0, results are closer to discrete solutions
- For u>0 big enough, the function becomes convex

Hierarchical Improvement

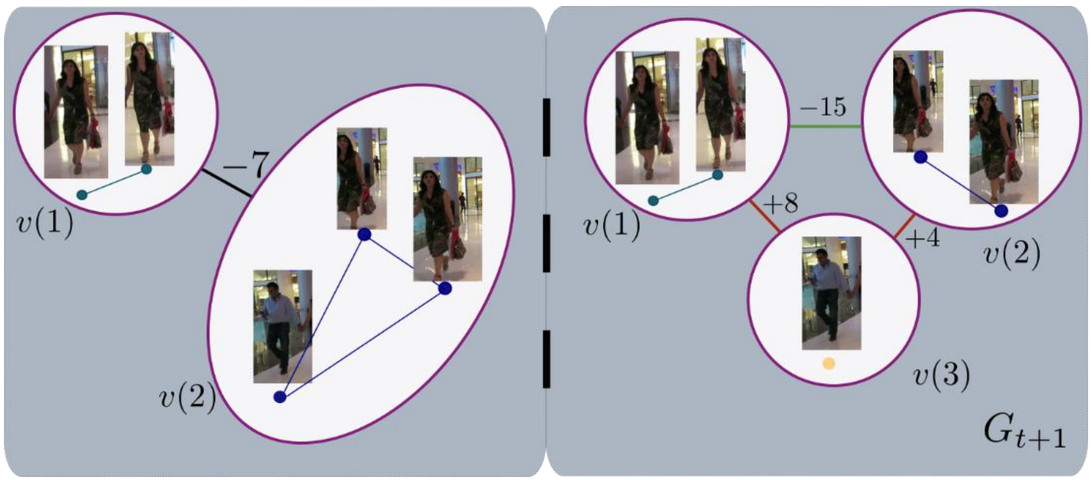


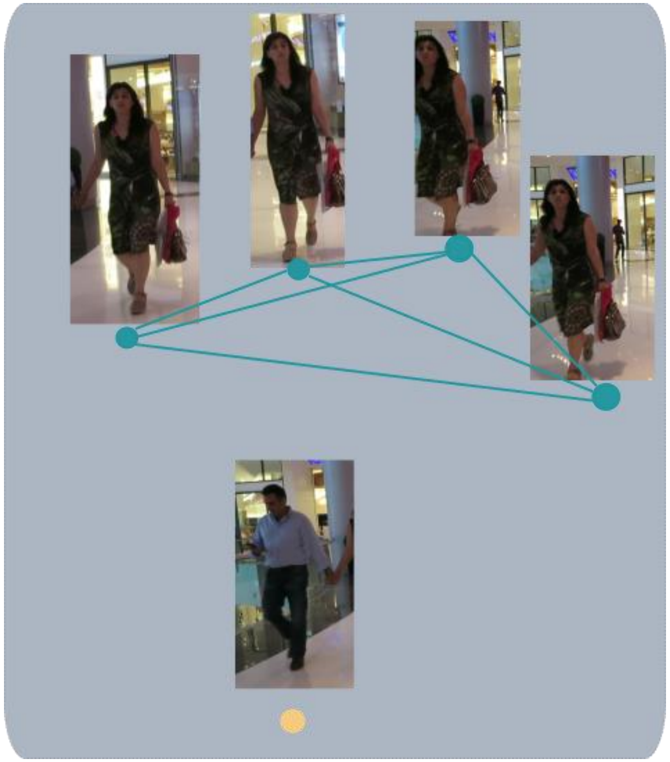


1. Label result

2. Correction + Recompute

Hierarchical Improvement

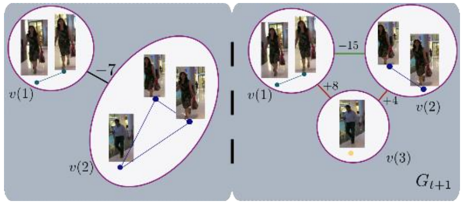




2. Correction + Recompute

3. Expand

Hierarchical Improvement



Correction step

- Fixes obvious error due to
  - Rounding
  - Local optimality

Recompute step

- Enourmos reduction of problem size
  - Can be solved optimally via Gurobi
- Continues optimizing the labeling problem
  - New label problem w.r.t. last solution
  - Recompute weights accordingly