A Multi-View Pedestrian Tracking Framework Based on Graph Matching

Sponsored by Cisco Systems Inc., 2016-2017

Fanyi Duanmu, Xin Feng, and Yao Wang
Video Lab, NYU Tandon school of Engineering

Xiaoqing Zhu and Wai-tian Tan
Cisco System Inc.

Video LAB
NYU | TANDON SCHOOL OF ENGINEERING

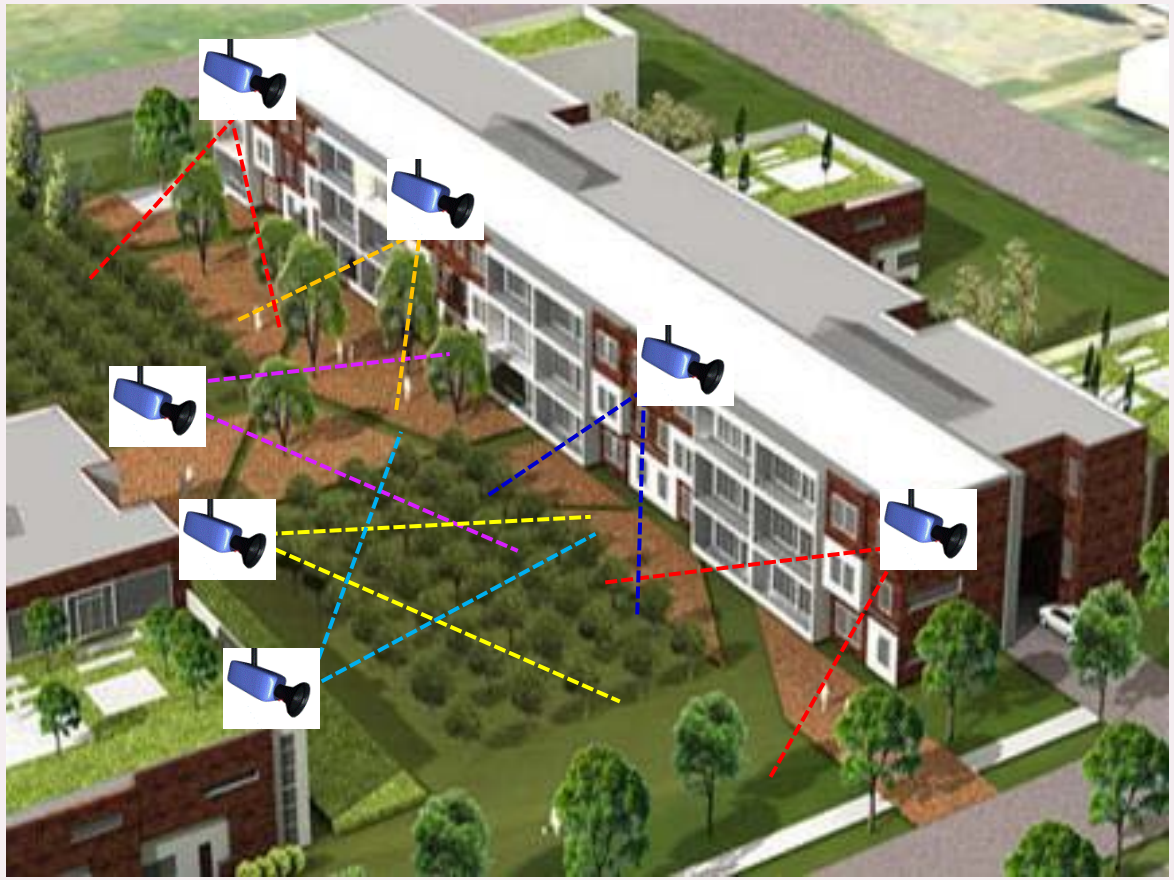
Agenda

- Motivation and Technical Challenges
- System Framework
- Pedestrian Detection
- Single-View Pedestrian Tracking
- Cross-View Pedestrian Tracking
- Tracking Results and Conclusions

- Motivation and Technical Challenges
- System Framework
- Pedestrian Detection
- Single-View Pedestrian Tracking
- Cross-View Pedestrian Tracking
- Tracking Results Conclusions

Motivation

- Video monitoring of large public or private spaces using camera network is an incoming IoE service with broad impact.
- Cross-View identification and association becomes important to track and summarize objects of interest captured from different views.



Technical Challenges

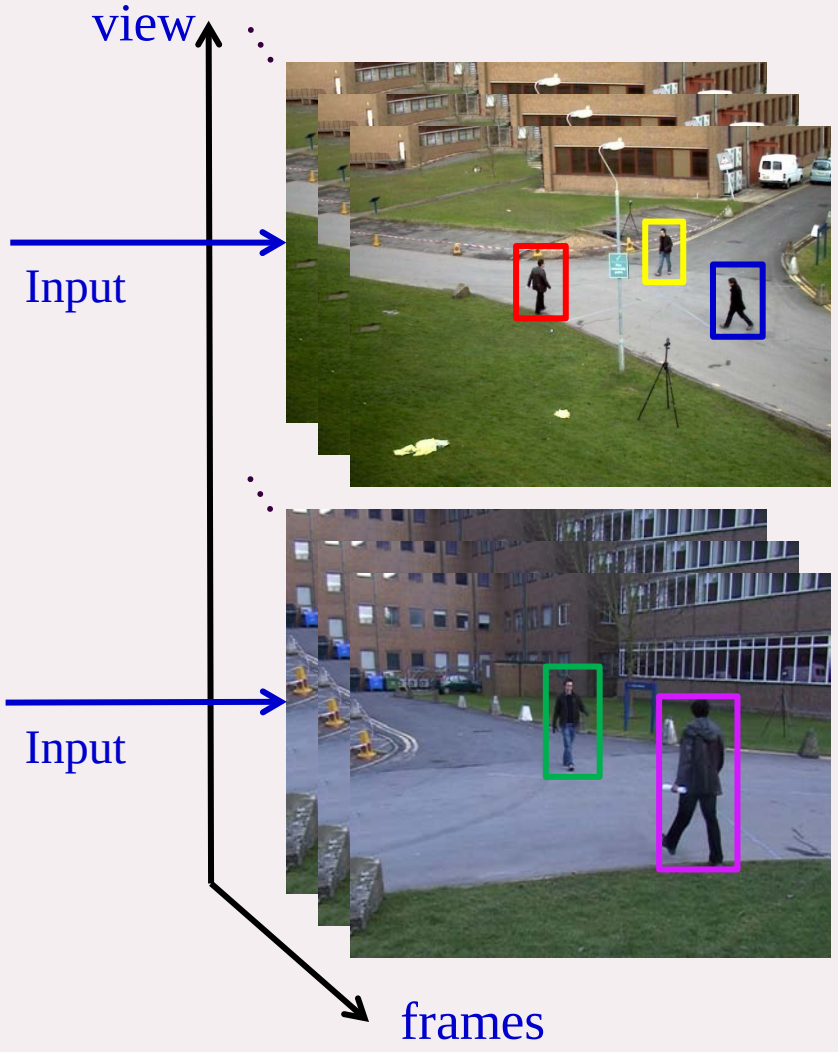
Challenges:

- Object occlusion
- Trajectory intersection
- Detection failure
- Illumination change
- Appearance difference across view
- etc…

- Motivation and Technical Challenges
- System Framework
- Pedestrian Detection
- Single-View Pedestrian Tracking
- Cross-View Pedestrian Tracking
- Tracking Results and Conclusions

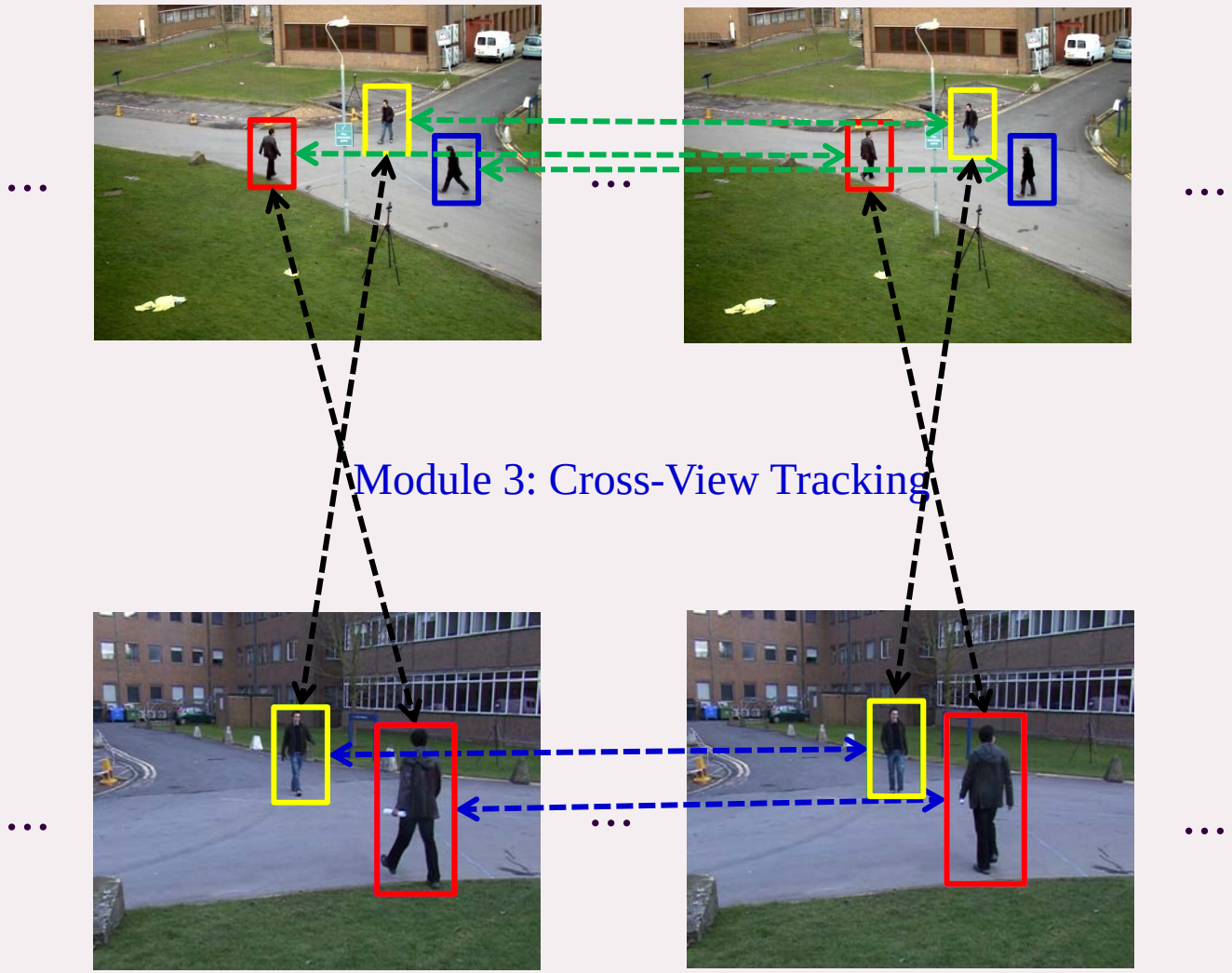
System Framework

Module 1: Pedestrian Detection



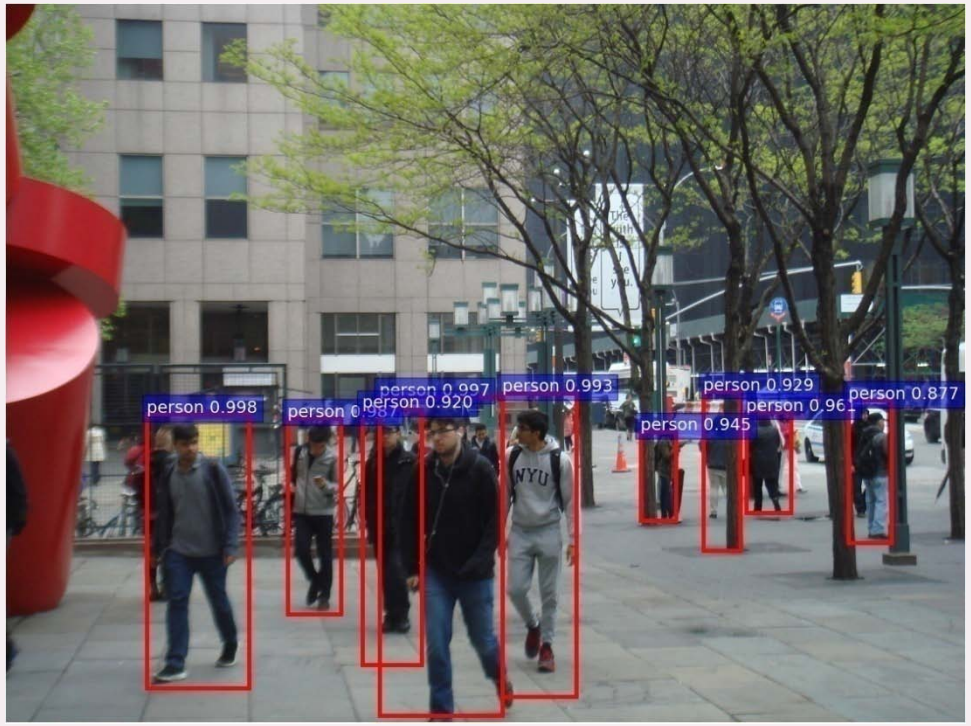
Module 2: Single-View Tracking

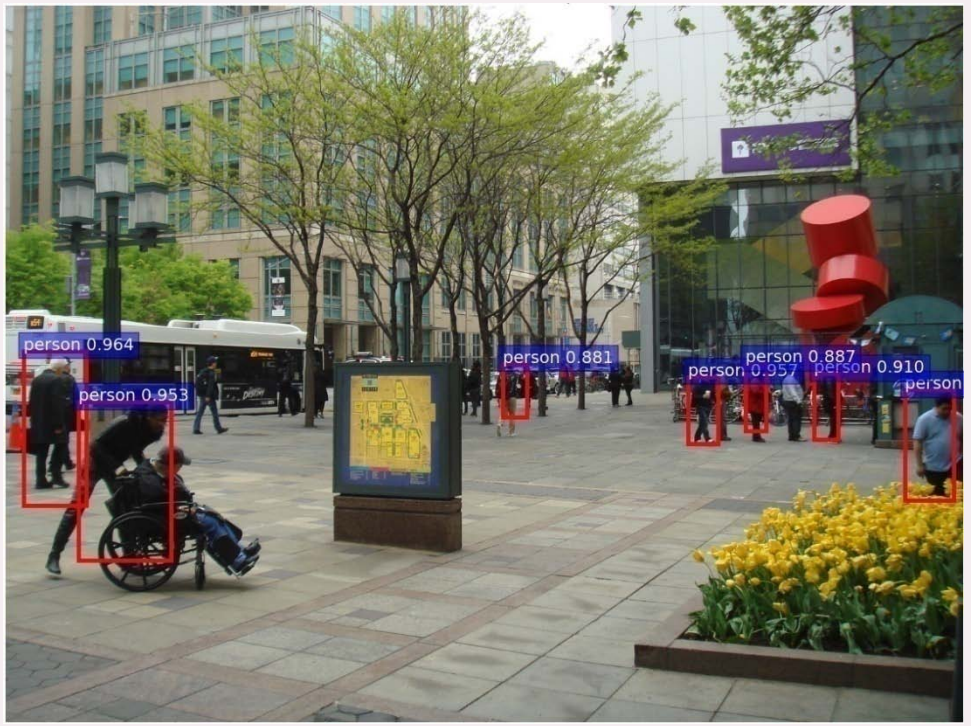
Module 3: Cross-View Tracking



- Motivation and Technical Challenges
- System Framework
- Pedestrian Detection
- Single-View Pedestrian Tracking
- Cross-View Pedestrian Tracking
- Tracking Results and Conclusions

Pedestrian Detection





- Use RCNN[1] detector to identify pedestrians.
- Return the bounding box coordinates, confidence, feature vectors, etc.
- Accurate and Robust to illumination change.

[1] Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", Advances in Neural Information Processing Systems (NIPS), 2015.

- Motivation and Technical Challenges
- System Framework
- Pedestrian Detection
- Single-View Pedestrian Tracking
- Cross-View Pedestrian Tracking
- Tracking Results and Conclusions

Single-View Pedestrian Tracking

Scenarios:

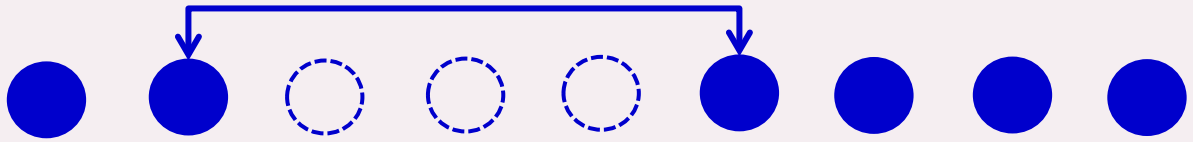
1. Short-term mis-detection

Solution: Historical Object Lookup

1-2 frames missed

2. Mid-term mis-detection

Solution: Single-View Graph Matching



<10 frames missed

3. Long-term mis-detection

Solution: New Label Assignment

>10 frames missed

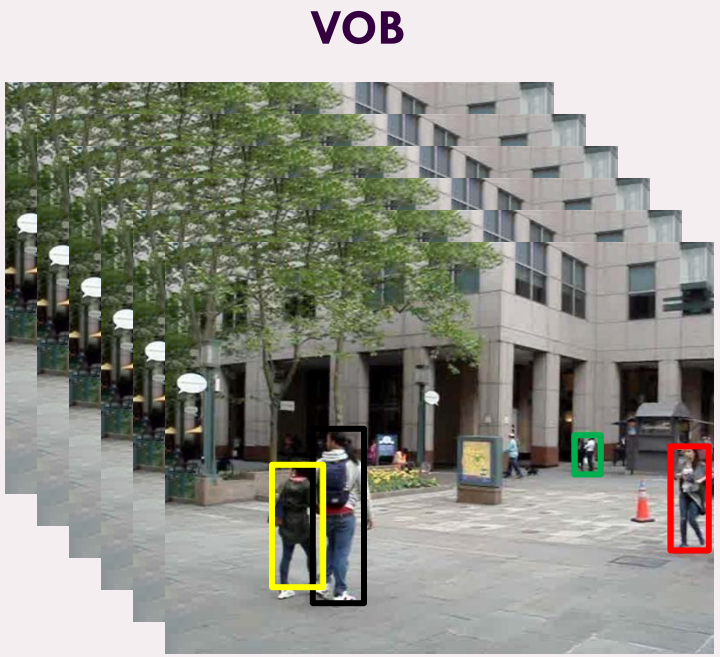
Single-View Pedestrian Tracking

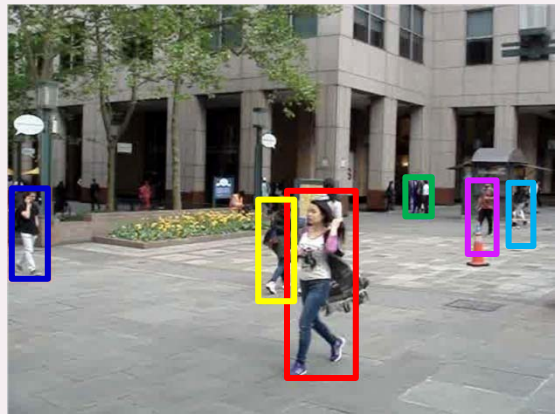
Short-term Tracking

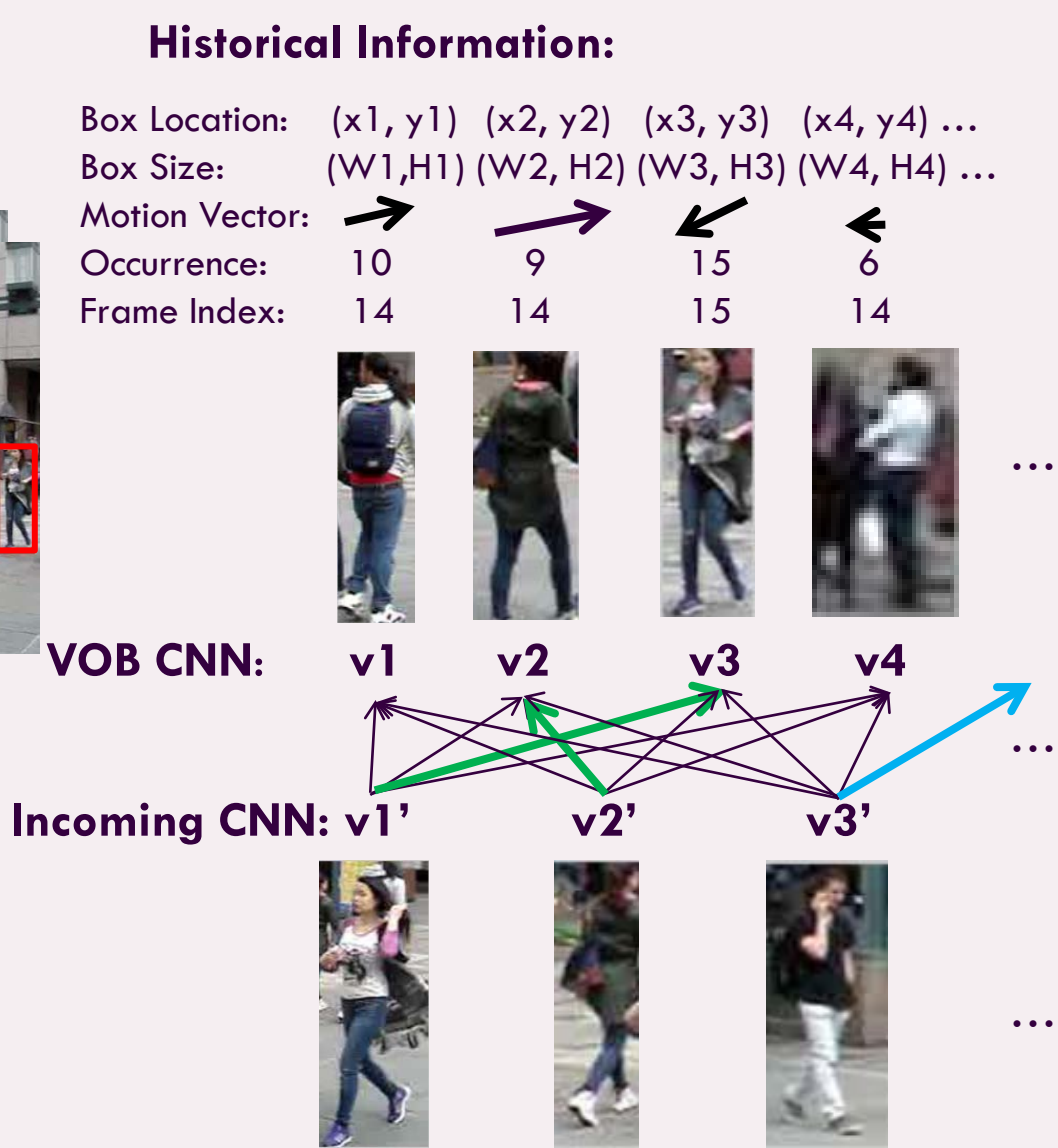
- Based on Virtual Object Buffer (VOB) lookup.

Association Criteria

- IOU between new object and motion-compensated VOB candidate is large.
- CNN feature Euclidean-distance is minimal.
- If not satisfied, create new entry into the VOB.







Single-View Pedestrian Tracking

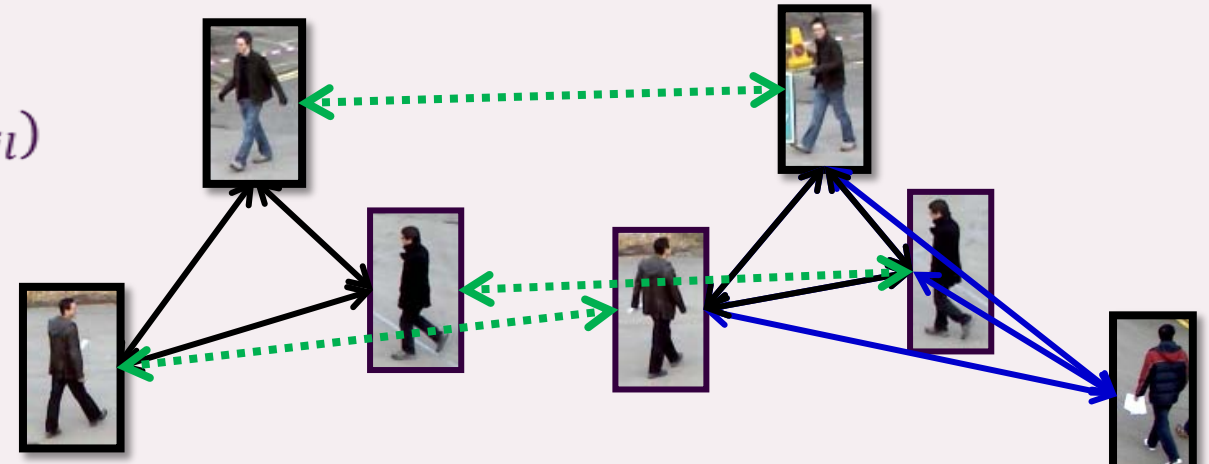
Mid-term Tracking

- Based on Graph Matching[2](GM) Technique.
- Aimed to decide the optimal correspondence between object nodes, to ultimately maximize the global sum of node/edge compatibilities.

$$J(X) = \sum_{i,j} x_{i,j} KP(n_i, n_j) + \sum_{e_{ik} \in E_1, e_{jl} \in E_2} x_{ij}\, x_{kl} KQ(e_{ik}, e_{jl})$$

Node Affinity — Edge Affinity

- Pedestrian → Graph Node. Node attributes: CNN feature L-2 distance.
- Spatial relationships → Graph Edge. Edge Attributes: displacement vector L-2 distance.



Frame: 10 Frame: 14

[2] F. Zhou and F. De la Torre, "Factorized Graph Matching", IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 38(9):1774-1789, 2016.

- Motivation and Technical Challenges
- System Framework
- Pedestrian Detection
- Single-View Pedestrian Tracking
- Cross-View Pedestrian Tracking
- Tracking Results and Conclusions

Cross-View Pedestrian Tracking

Cross-View Association:

- Step 1: Homography Projection
- Step 2: Graph Matching

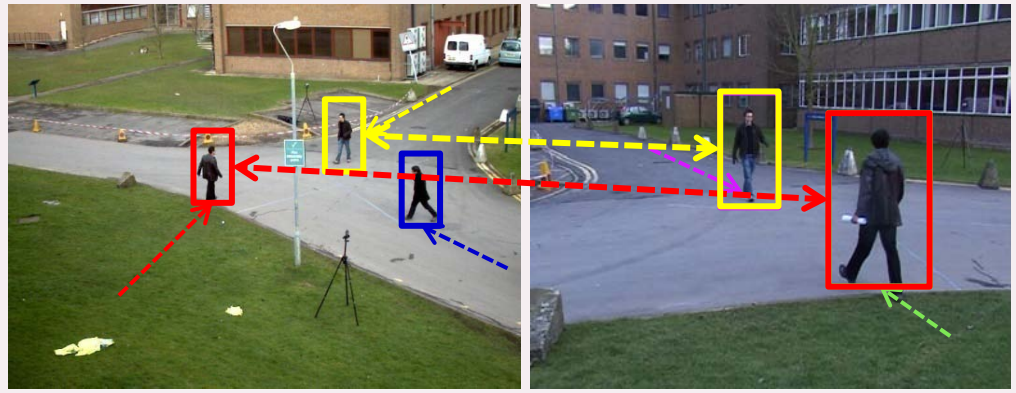
Node attributes:

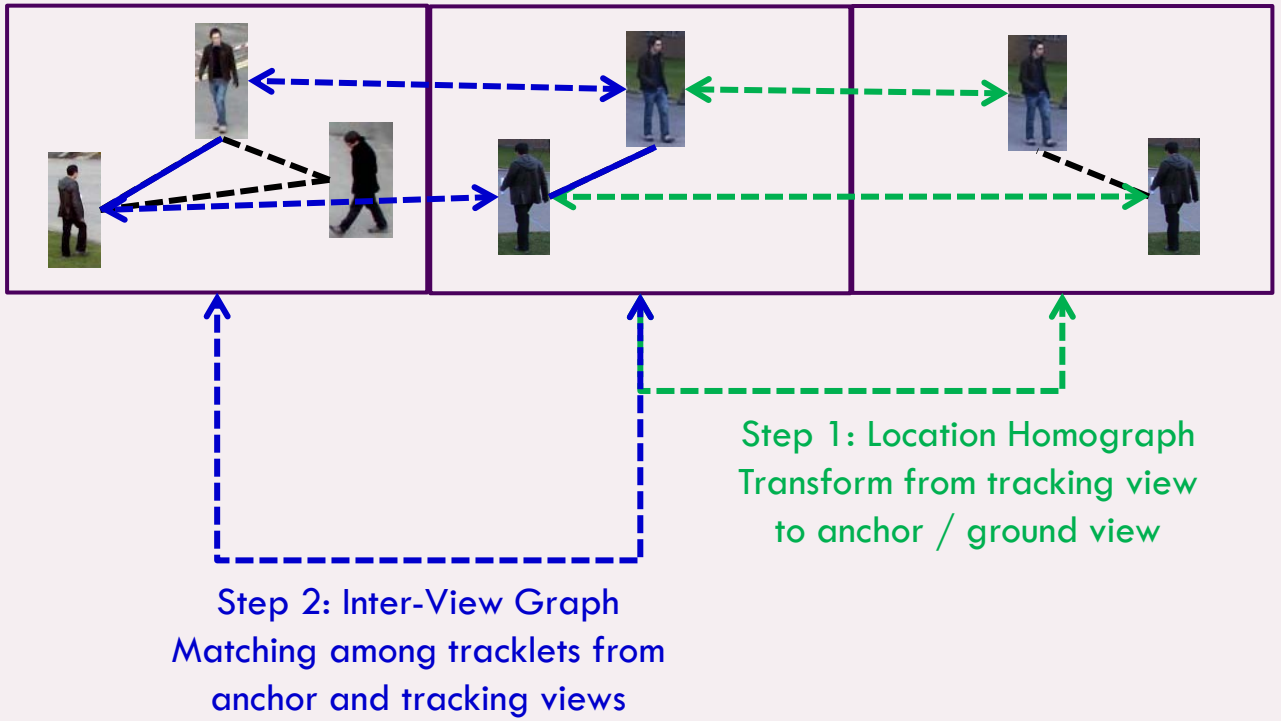
- Tracklet average CNN similarity.
- Tracklet bounding box location similarity over co-detected frames.

Edge attribute:

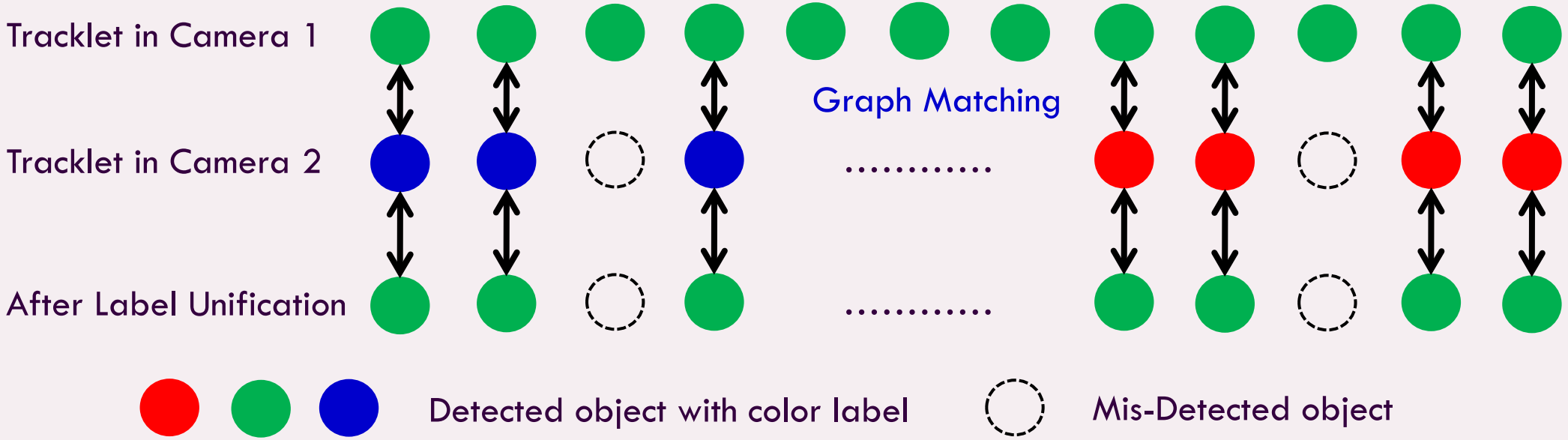
- Tracklet motion vector similarity over co-detected frames.

- Step 3: Label Unification





Cross-View Pedestrian Tracking



Label Unification:

The shorter tracklet in one view will carry over the label from the associated object with longer tracklet in the other view, as long as it does not overlap temporally with any other previous associated tracklet(s) in the same view.

- Motivation and Technical Challenges
- System Framework
- Pedestrian Detection
- Single-View Pedestrian Tracking
- Cross-View Pedestrian Tracking
- Tracking Results and Conclusions

Comparison with Prior works on PETS-2009 dataset

We simulate the multi-view tracking among *View* 1, *View* 5 and *View* 7 for 200 frames, using *View* 1 as the reference view. Objects from *View* 5 and *View* 7 are projected onto *View* 1 and associated and then the labels are unified across views.

Cross-View Tracking Experimental Results				
Method	Z. Wu, et.al[3] 3-view	Z. W, et.al[3] 3-view	C. Li, et. al[4] 2-view	Proposed 3-view
MOTP[5]	53.4%	60.0%	79.9%	83.6%
MOTA[5]	74.1%	74.0%	93.2%	94.8%

[3] Z. Wu, N.I. Hristov, T.H. Kunz, M. Betke, “Tracking-reconstruction or reconstruction tracking Comparison of two multiple hypothesis tracking approaches to interpret 3D object motion from several camera views”, Workshop on Motion and Video Computing (WMVC), pp. 1–8, 2009.
[4] C. Li, S. Ping, H. Sheng, J. Chen, and Z. Xiong, "Multi-view Multi-object Tracking Based on Global Graph Matching Structure." Pacific Rim Conference on Multimedia, 2016.
[5] K. Bernardin, and S. Rainer, "Evaluating multiple object tracking performance: the CLEAR MOT metrics.", EURASIP Journal on Image and Video Processing, Issue: 1, pp: 1-10, 2008.


QUESTIONS?