People Tracking and Re-Identification from Multiple Cameras

Ph.D. Defense

Ergys Ristani

Advisor: Carlo Tomasi

Committee Members: Ronald Parr, Pankaj Agarwal, Jiri Matas

DUKE
COMPUTER SCIENCE

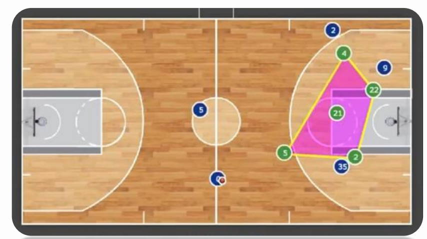
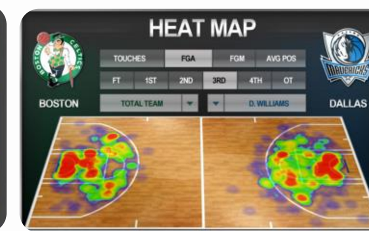
Motivation

Applications

- Security



- Sport

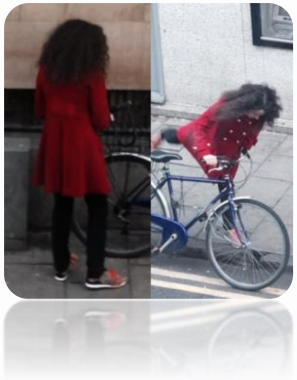
- Retail


Challenges

- Appearance can be unstable due to
  - Changes in lighting
  - Changes in pose
  - Changes in viewpoint
  - Occlusion

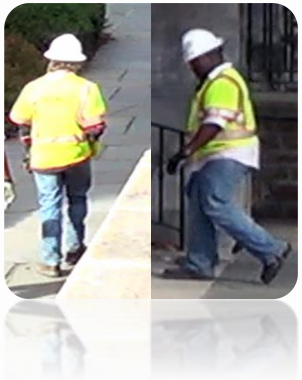
Challenges

- Appearance can be unstable due to
  - Changes in lighting
  - Changes in pose
  - Changes in viewpoint
  - Occlusion



Same person

Varying appearance

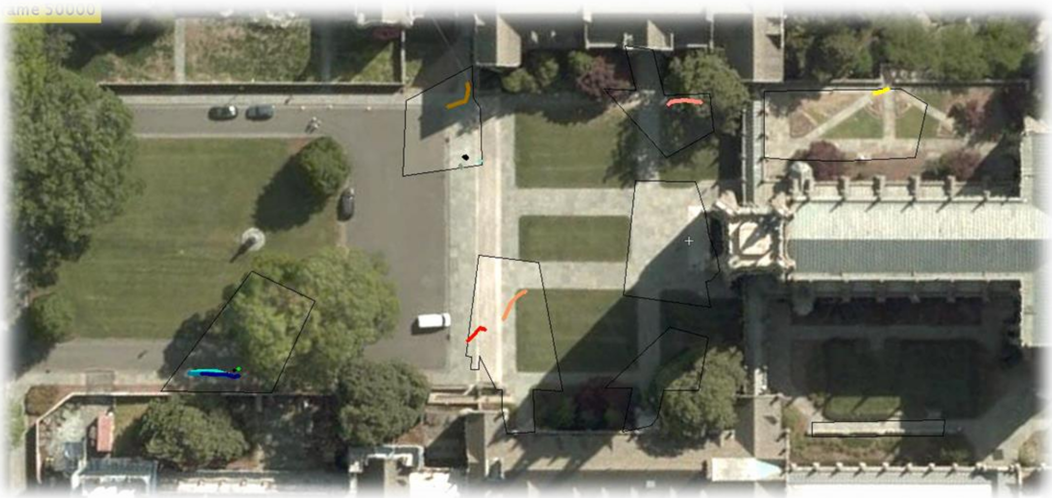


Different people

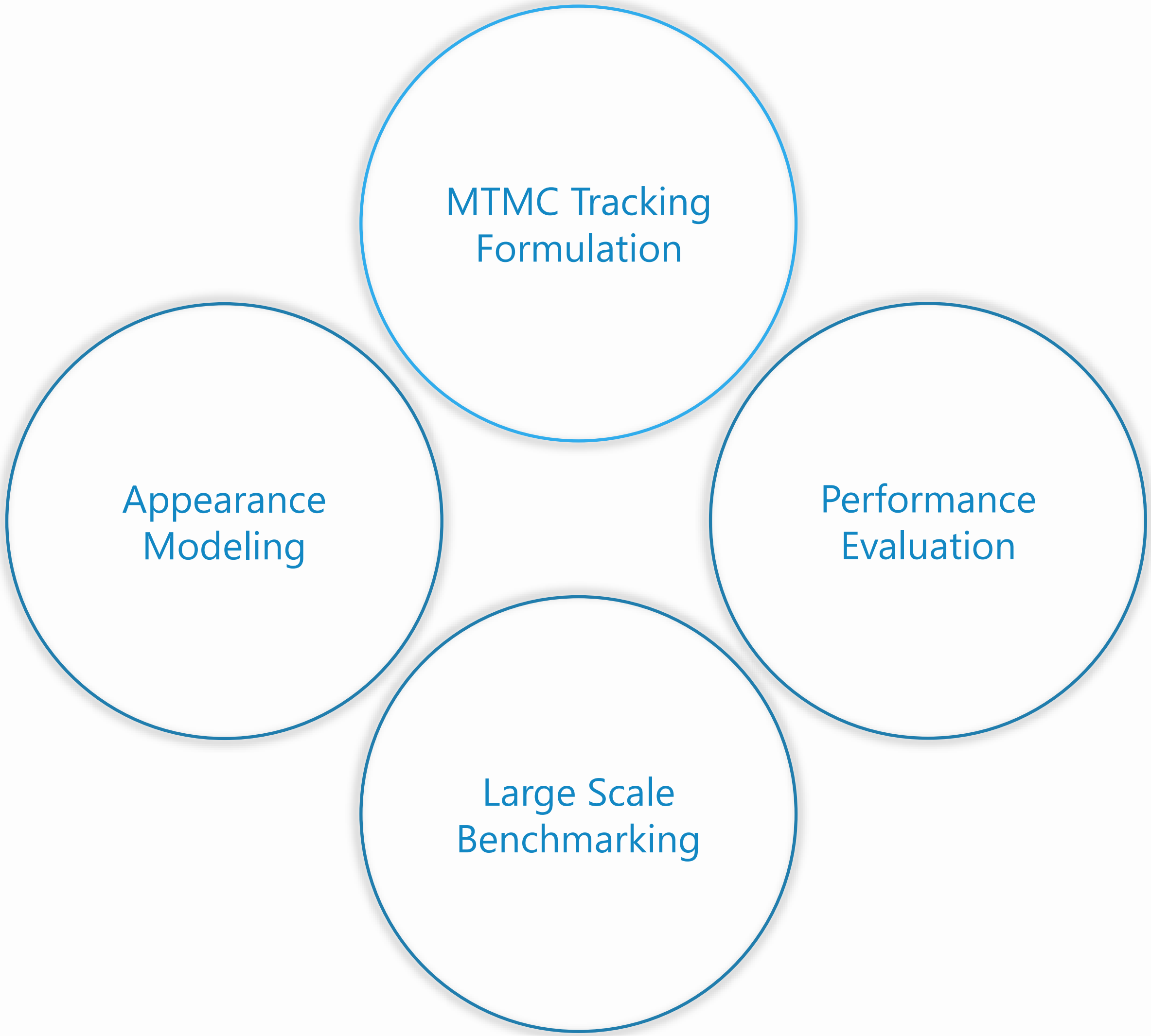
Similar appearance

Challenges

- Blind spots between cameras
- Unconstrained paths/number of people
- Large amounts of data to process

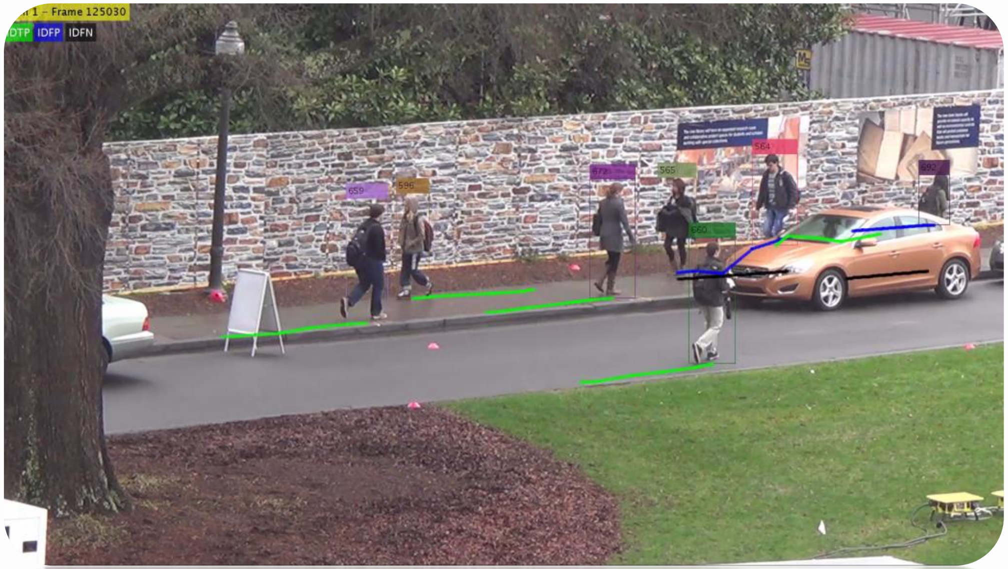


My Research

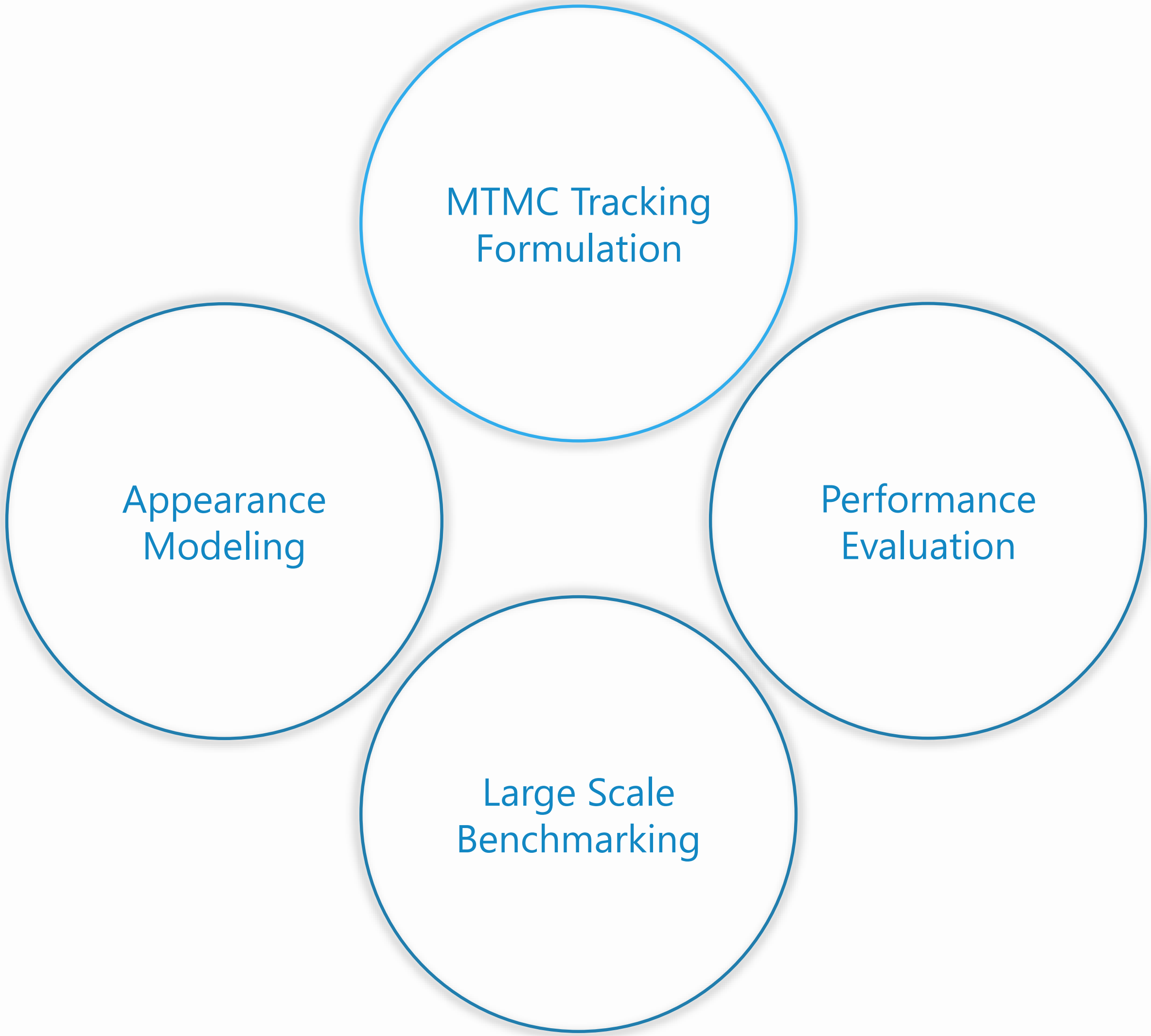

MTMC Tracking Formulation
Appearance Modeling
Performance Evaluation
Large Scale Benchmarking

My Research

MTMC Tracking
Formulation

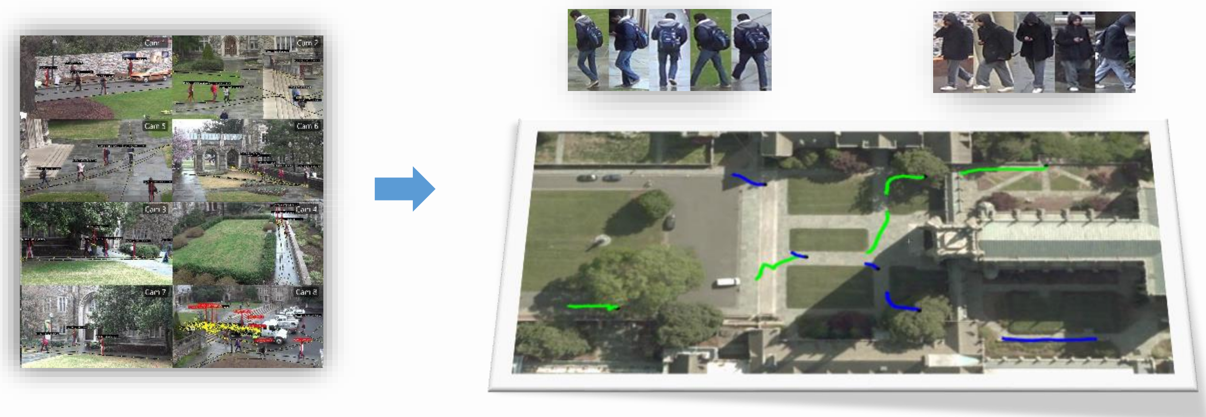


My Research


MTMC Tracking Formulation
Appearance Modeling
Performance Evaluation
Large Scale Benchmarking

Problem Definitions

- Multi-Target Multi-Camera Tracking (MTMCT):
  - Given *n* camera streams, determine who is where at all times



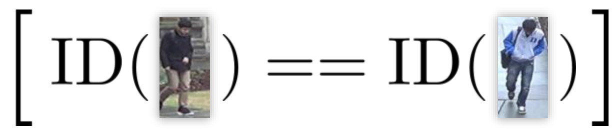
- Person Re-Identification (ReID):
  - Given a query image, retrieve the most similar identities from a database

Query Gallery

Problem Definitions

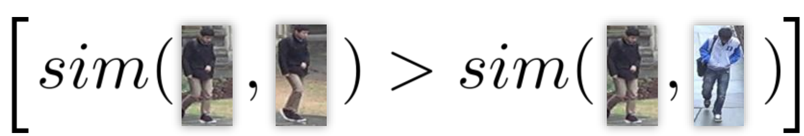
- Multi-Target Multi-Camera Tracking (MTMCT):
  - Given *n* camera streams, determine who is where at all times

Classification



- Person Re-Identification (ReID):
  - Given a query snapshot, retrieve the most **sim**ilar identities from a database

Ranking



My Research

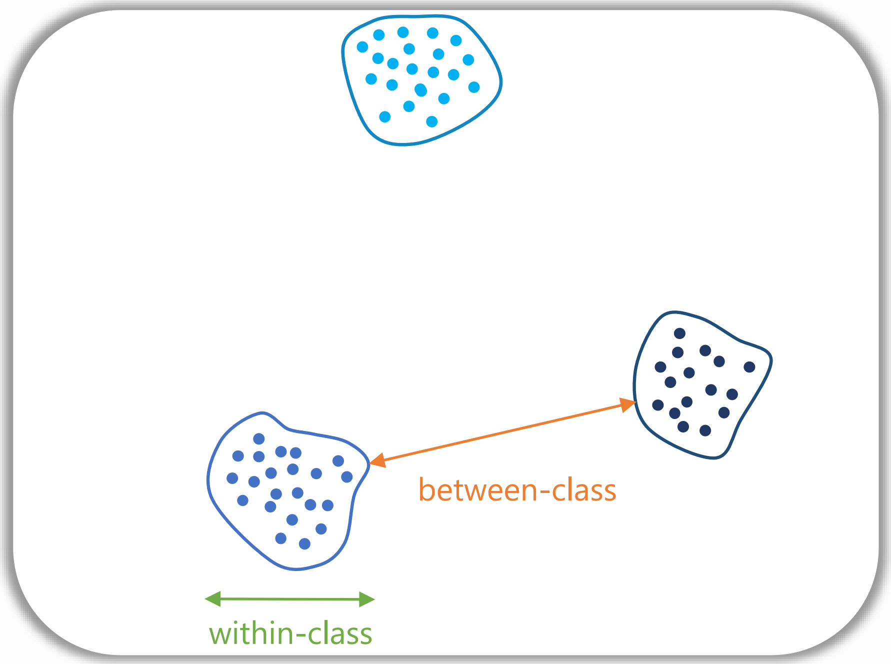
MTMC Tracking Formulation

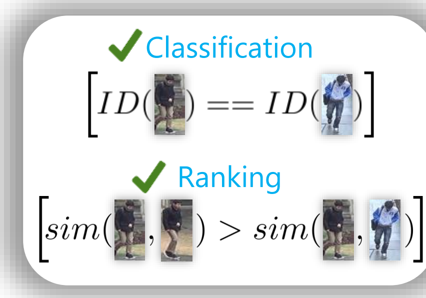
Appearance Modeling

Performance Evaluation

Large Scale Benchmark

Ideal Feature Space for MTMCT and ReID





$$\overbrace{\max_{a \sim p} d(a,p)}^{\text{largest within-class distance}} \quad < \quad \overbrace{\min_{a \nsim n} d(a,n)}^{\text{smallest between-class distance}}$$

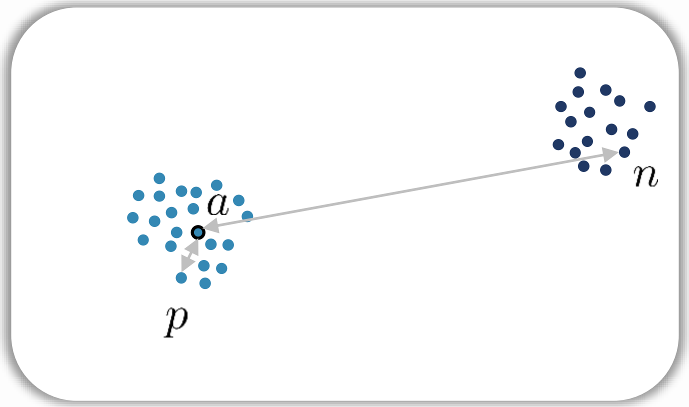
Well-separated classes guarantee correct ranking and classification

Triplet Loss

Triplet Loss

$$L_U = [m + d(a,p) - d(a,n)]_+$$

Schroff et al. CVPR 2015



Combinatorial triplet selection

Less discriminative

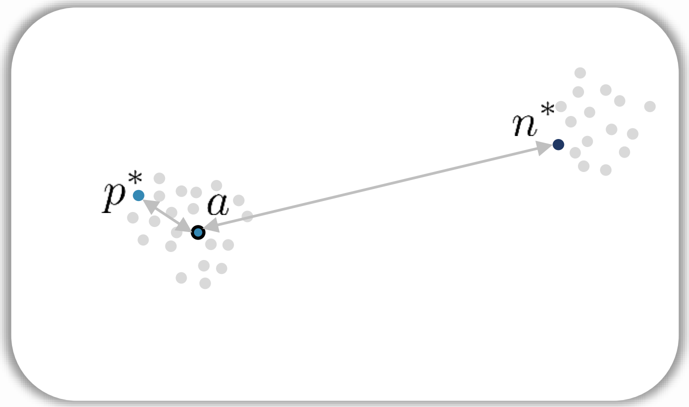
Robust to outliers

Hard Triplet Loss

Hard Triplet Loss $L_H = \left[m + \max_{p \in P(a)} d(a,p) - \min_{n \in N(a)} d(a,n) \right]_+$

Hermans et al. arXiv 2017

Mishchuk et al. NIPS 2017



No combinatorial triplet selection

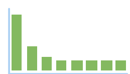
Discriminative

Sensitive to outliers

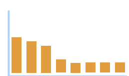
Weighted Triplet Loss

$$L_3 = \left[m + \sum_{p \in P(a)} w_p^a d(a,p) + \sum_{n \in N(a)} w_n^a d(a,n) \right]_+$$

Adaptive Weights



$$w_p^a = \frac{e^{d(a,p)}}{\sum_{i \in P(a)} e^{d(a,i)}}$$



$$w_n^a = \frac{e^{-d(a,n)}}{\sum_{j \in N(a)} e^{-d(a,j)}}$$

- ✔ Imbalance of positives/negatives
- ✔ Robust to outliers
- ✔ Emphasize hard examples

Triplet Loss

Schroff et al. CVPR 2015

$$L_U = [m + d(a,p) - d(a,n)]_+$$

Uniform Weights

$$w_p^a = 1$$

Hard Triplet Loss

Hermans et al. arXiv 2017

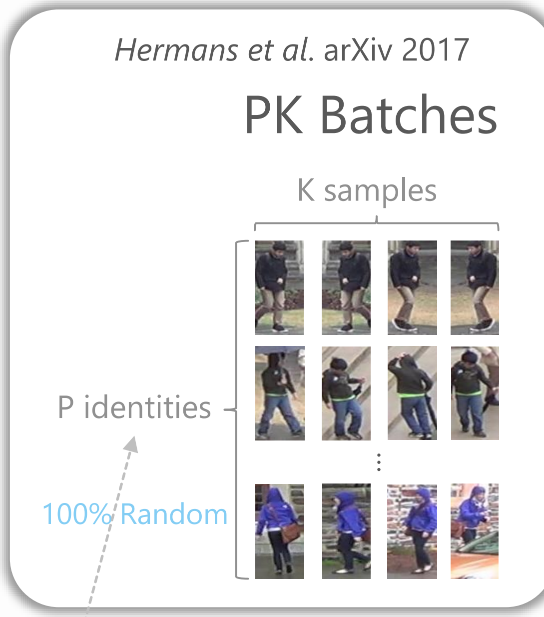
Mishchuk et al. NIPS 2017

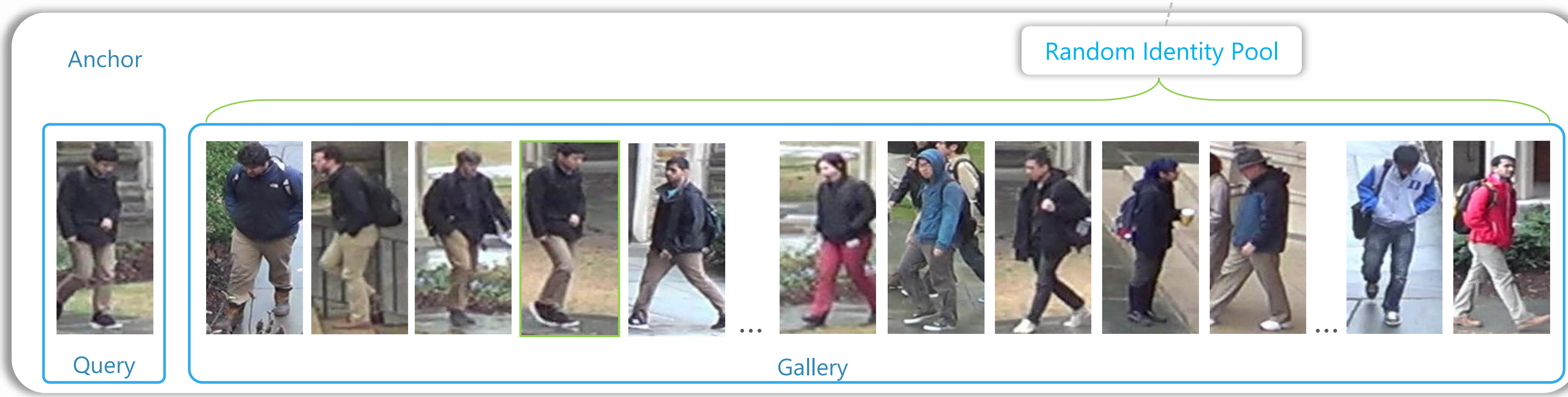
$$L_H = \left[m + \max_{p \in P(a)} d(a,p) - \min_{n \in N(a)} d(a,n) \right]_+$$

Binary (Hard) Weights

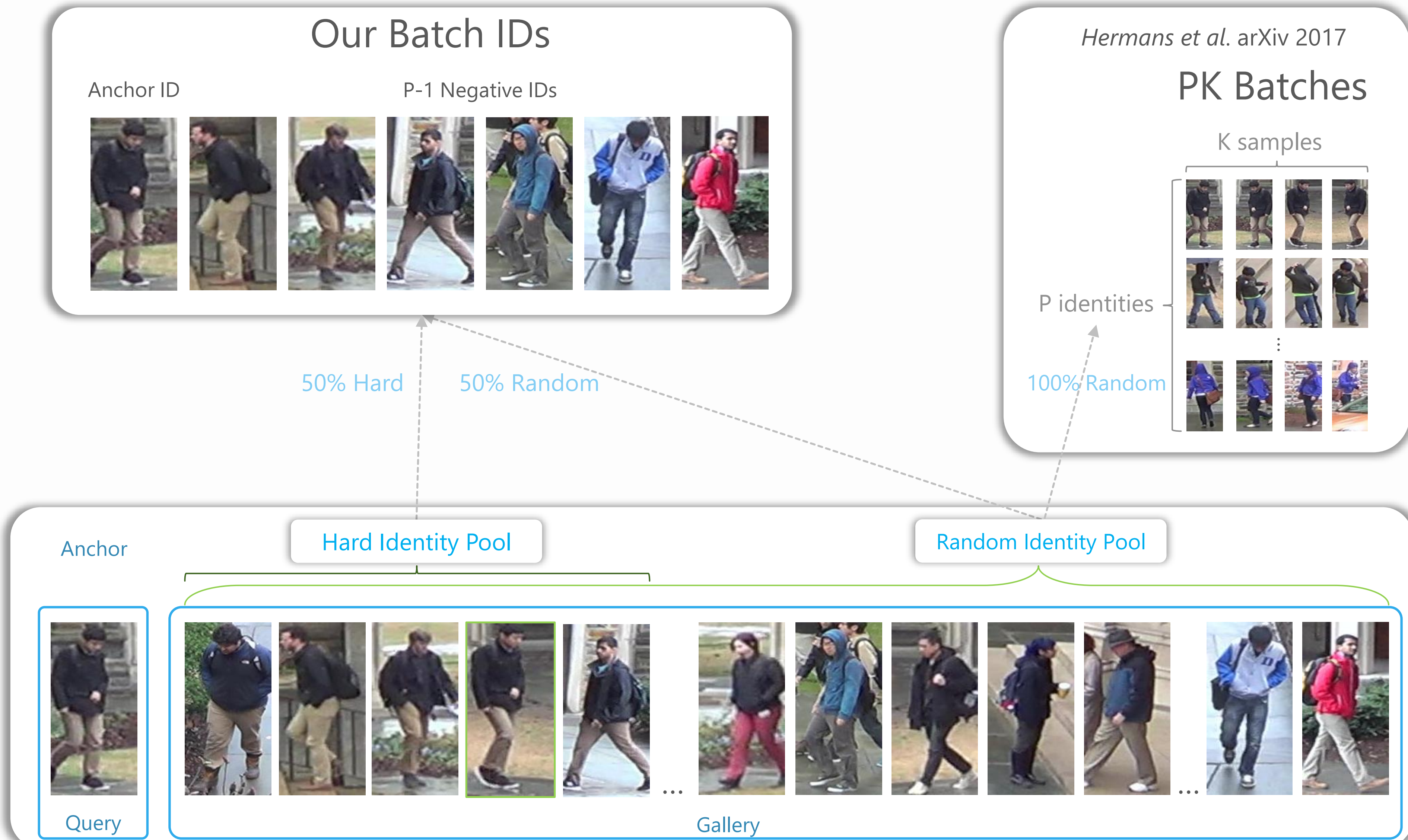
$$w_p^a = \begin{cases} 1, & p = \arg\max_{i \in P(a)} d(a,i) \\ 0, & \text{otherwise} \end{cases}$$

Hard Negative ID Mining





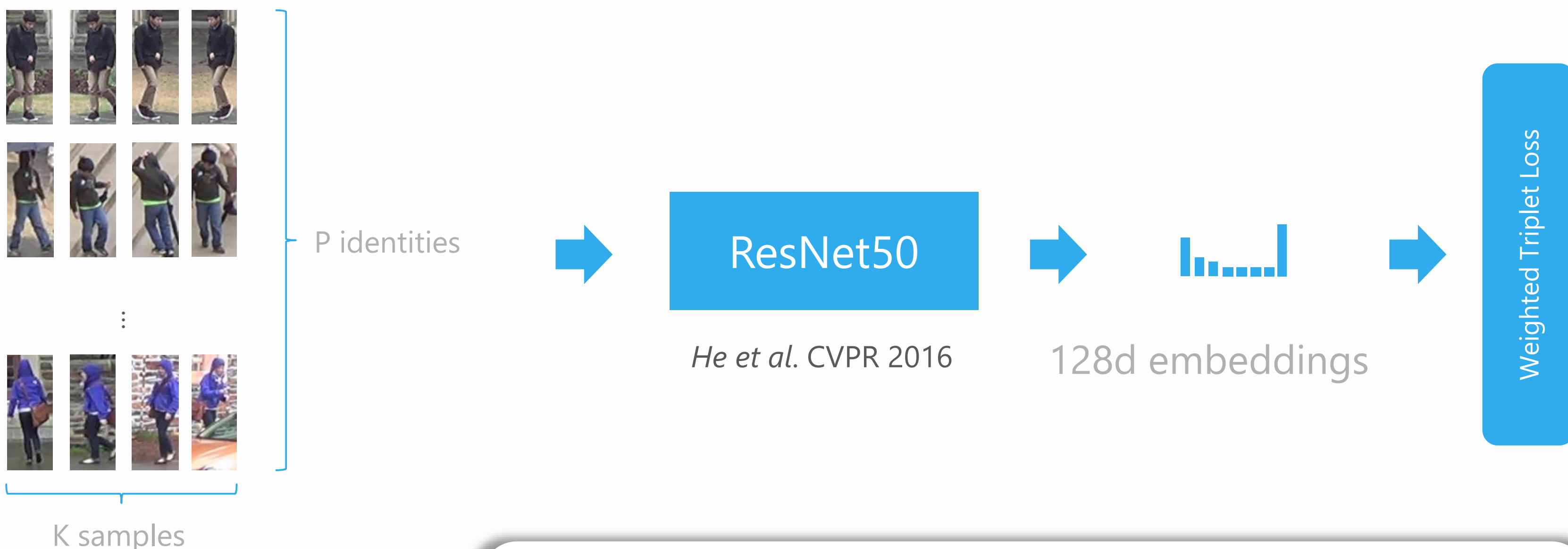
Hard Negative ID Mining

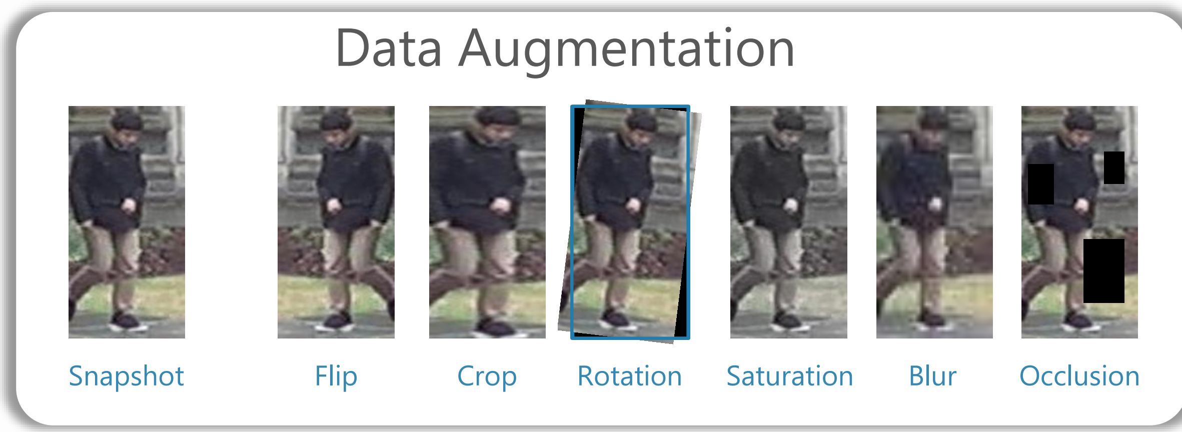


Appearance Evaluation

Training Setup

PK Batches





Qualitative Results

Qualitative Results

Adaptive vs Hard Weights

	Euclidean	
	mAP	rank-1
HardTriplet	54.60	73.24
Ours	**54.97**	**74.23**
HardTriplet (+Aug)	56.65	74.91
Ours (+Aug)	**57.28**	**75.31**
HardTriplet (+Aug+HNM)	54.90	74.23
Ours (+Aug+HNM)	**58.74**	**77.69**

Re-ID results on DukeMTMC-ReID

Hard Triplet Loss
Hermans et al. arXiv 2017
Mishchuk et al. NIPS 2017

Our adaptive weights are superior to the hard binary weights

Adaptive vs Hard Weights

	Euclidean	
	mAP	rank-1
HardTriplet	54.60	73.24
Ours	**54.97**	**74.23**
HardTriplet (+Aug)	56.65	74.91
Ours (+Aug)	**57.28**	**75.31**
HardTriplet (+Aug+HNM)	54.90	74.23
Ours (+Aug+HNM)	**58.74**	**77.69**

Re-ID results on DukeMTMC-ReID

Hard Triplet Loss
Hermans et al. arXiv 2017
Mishchuk et al. NIPS 2017

Hard Negative ID Mining shows the robustness of adaptive weights

State Of The Art Comparison

	Euclidean	
	mAP	rank-1
BoW+kissme	12.17	25.13
LOMO+XQDA	17.04	30.75
Baseline	44.99	65.22
DPFL	48.90	70.10
PAN	51.51	71.59
HardTriplet	56.65	74.91
SVDNet	56.80	76.70
Ours	**58.74**	**77.69**
DPFL (2-stream)	60.60	79.20
Ours (2-stream)	**63.40**	**79.80**

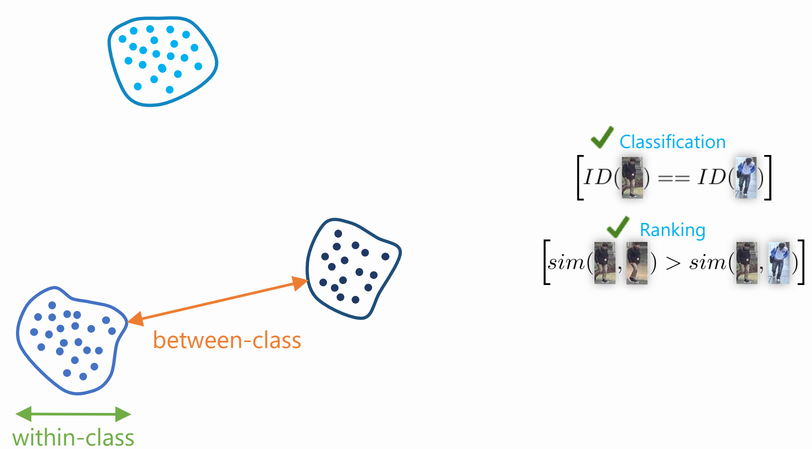
Re-ID results on DukeMTMC-ReID

Our method outperforms recent state of the art DPFL (ICCV 2017)

Summary

- Hard Negative ID Mining

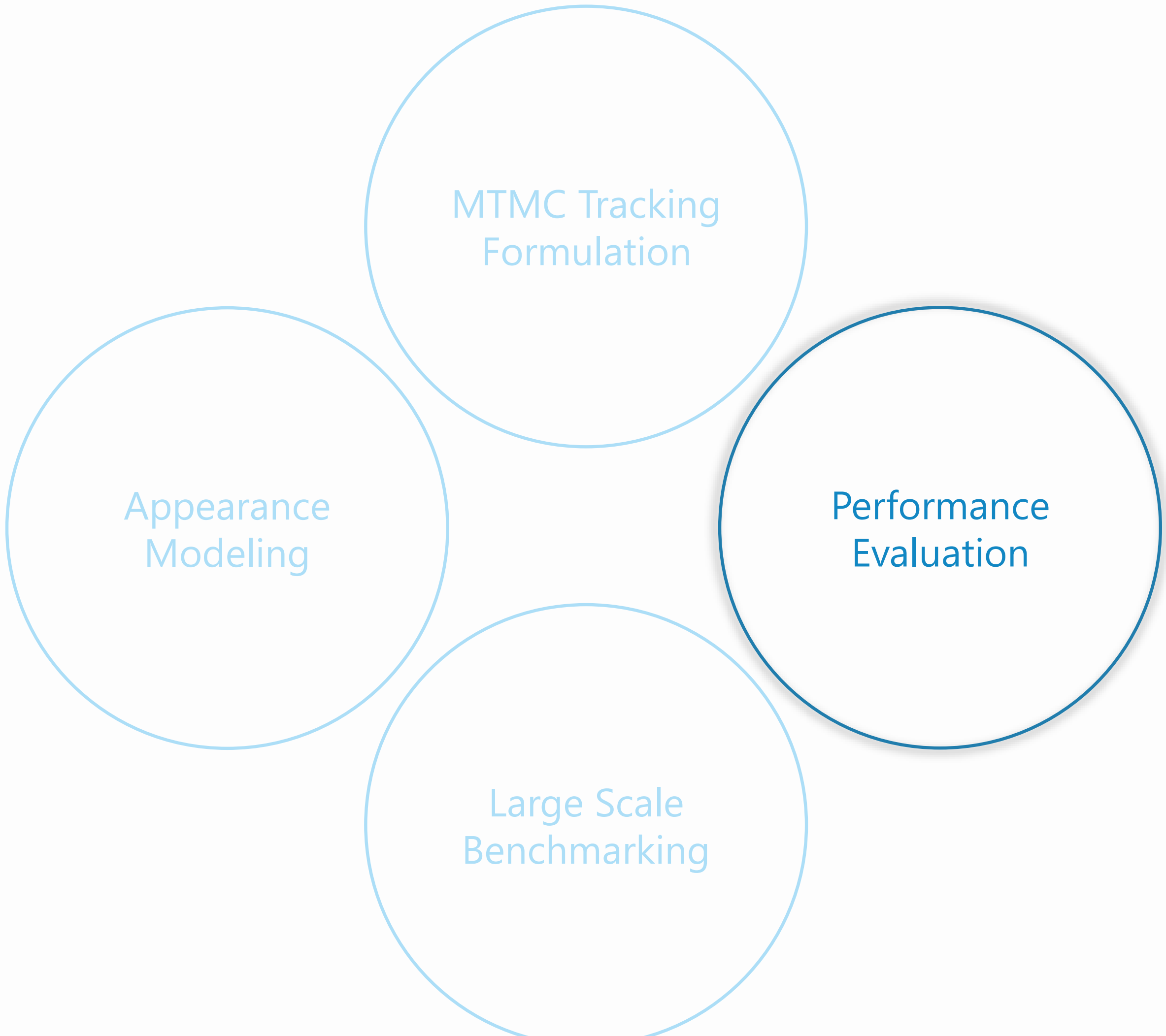
- Weighted triplet loss with adaptive weights for ReID and MTMCT



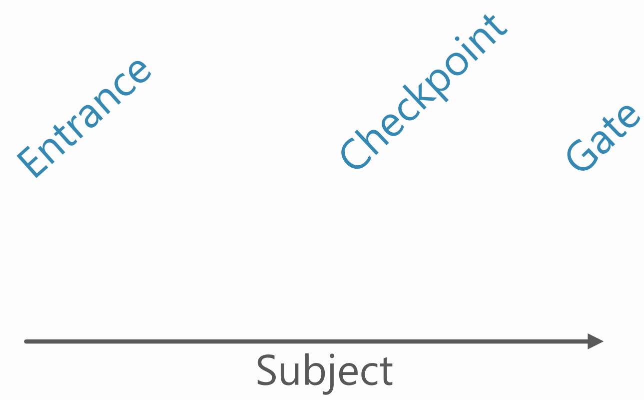
$$\overbrace{\max_{a\sim p} d(a,p)}^{\text{largest within-class distance}} < \overbrace{\min_{a\nsim n} d(a,n)}^{\text{smallest between-class distance}}$$

[1] Features for Multi-Target Multi-Camera Tracking and Re-Identification. Ristani and Tomasi. CVPR 2018

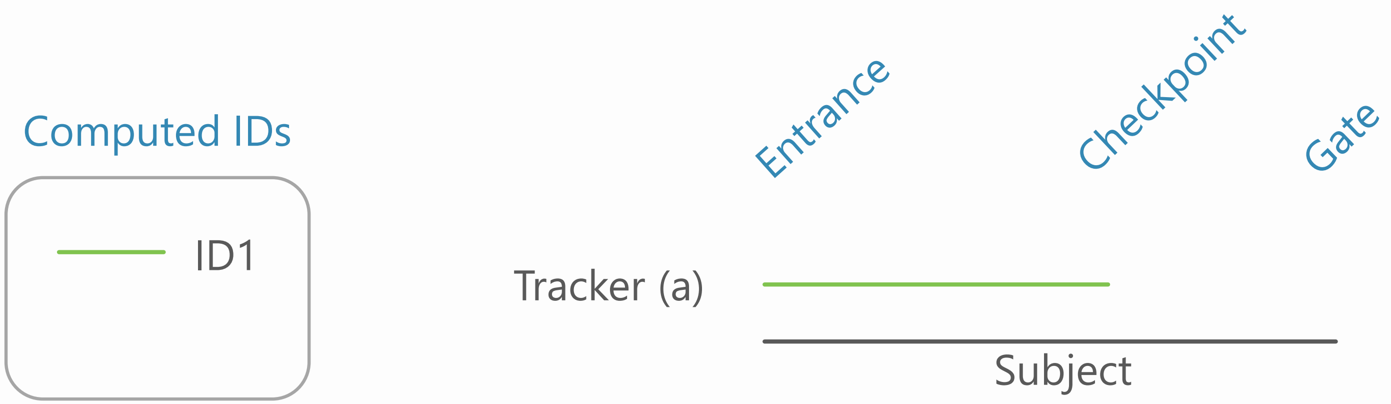
My Research


MTMC Tracking Formulation
Appearance Modeling
Performance Evaluation
Large Scale Benchmarking

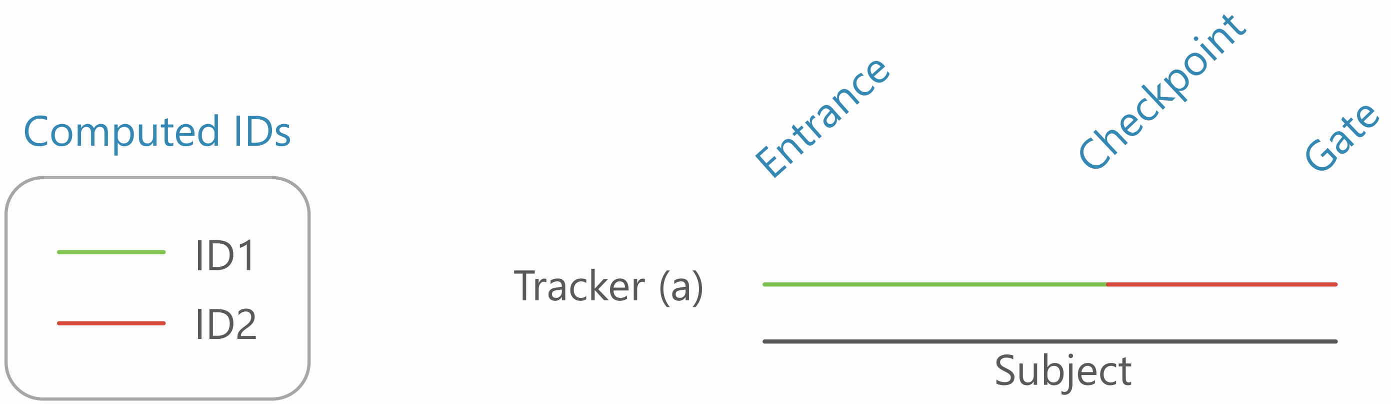
Evaluation Paradigms



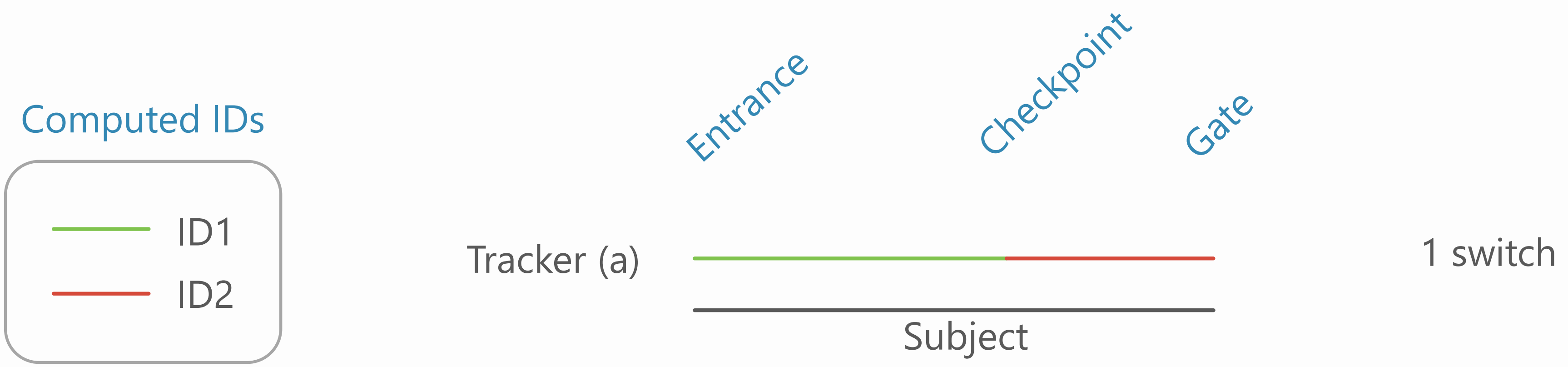
Evaluation Paradigms



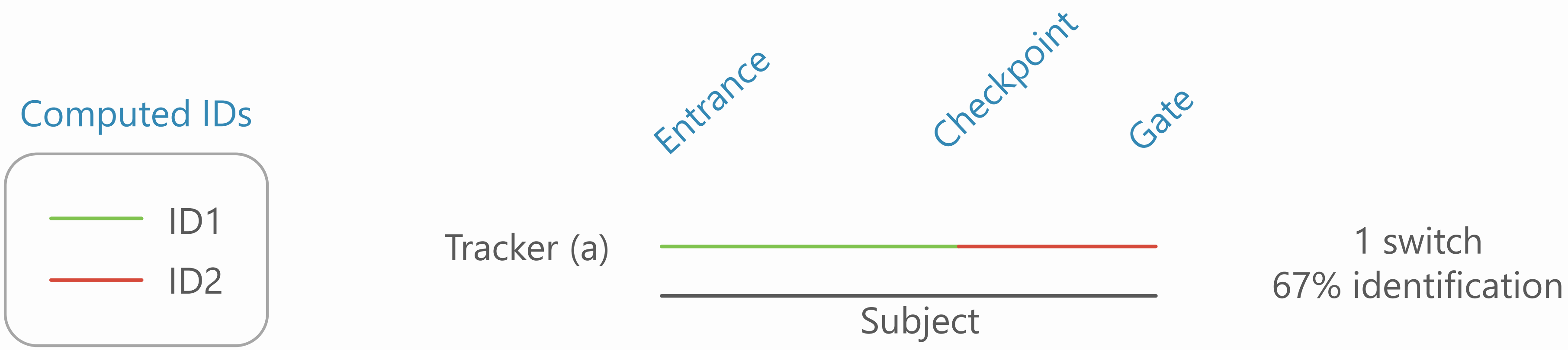
Evaluation Paradigms



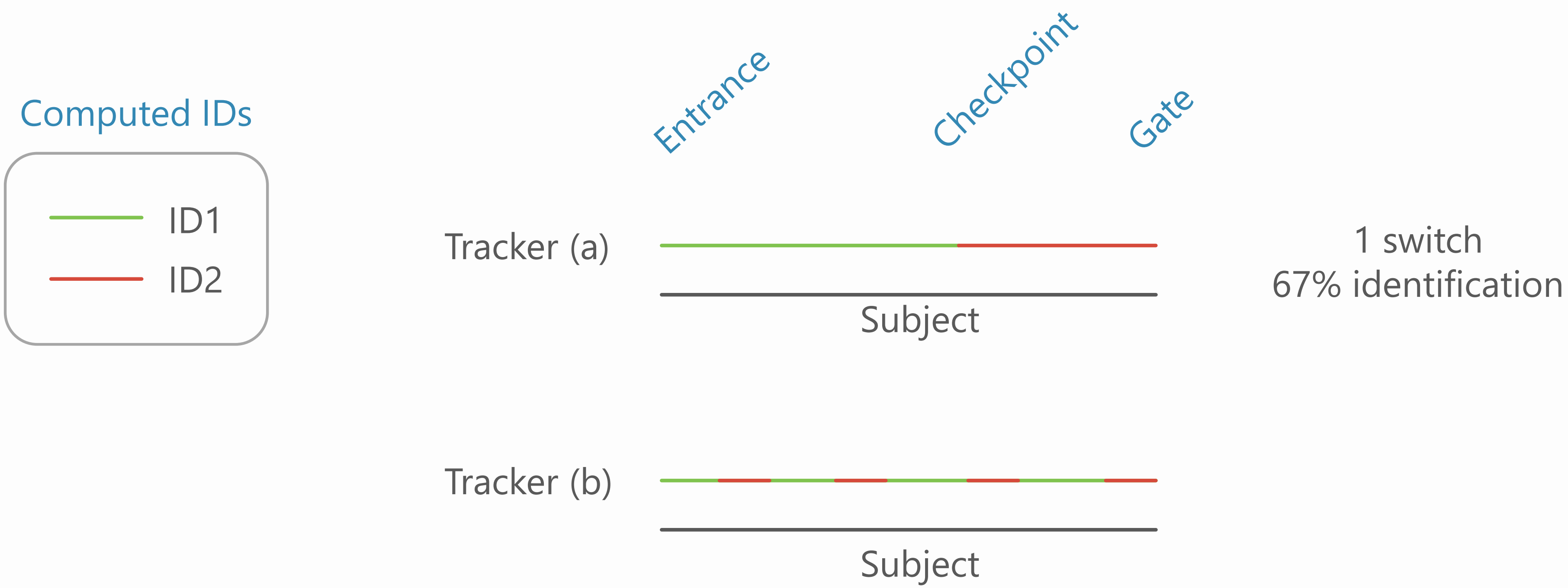
Evaluation Paradigms



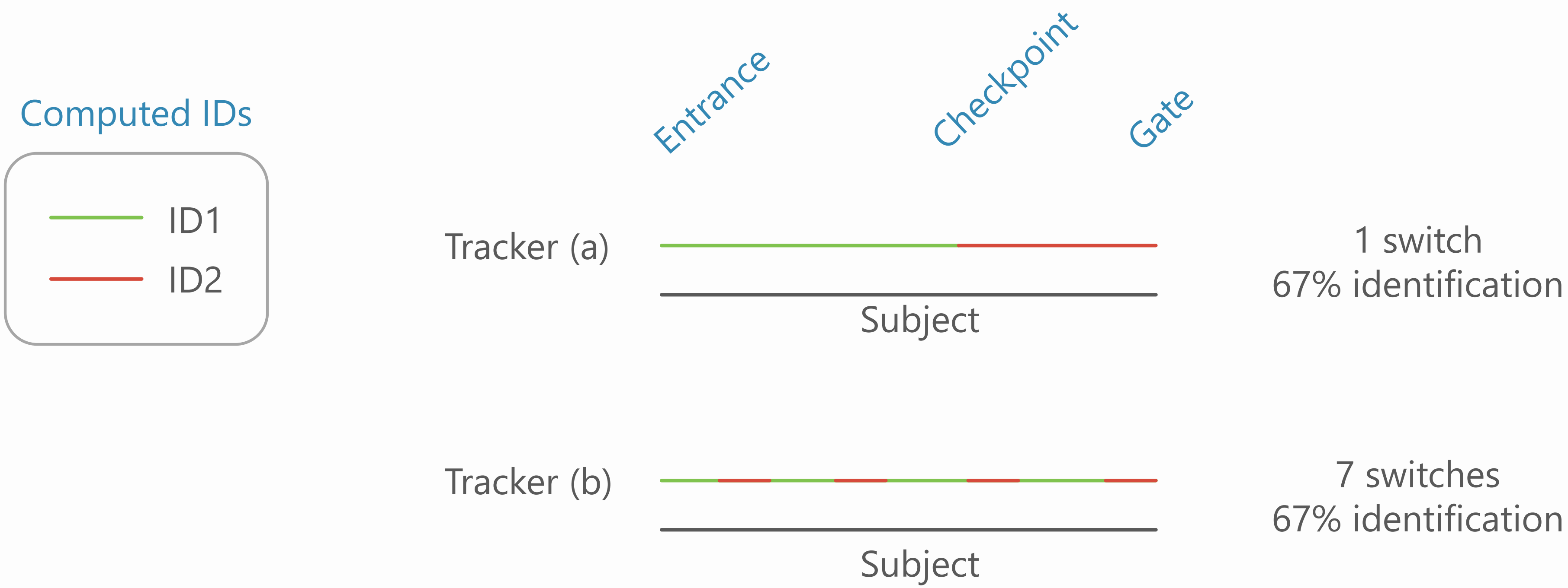
Evaluation Paradigms



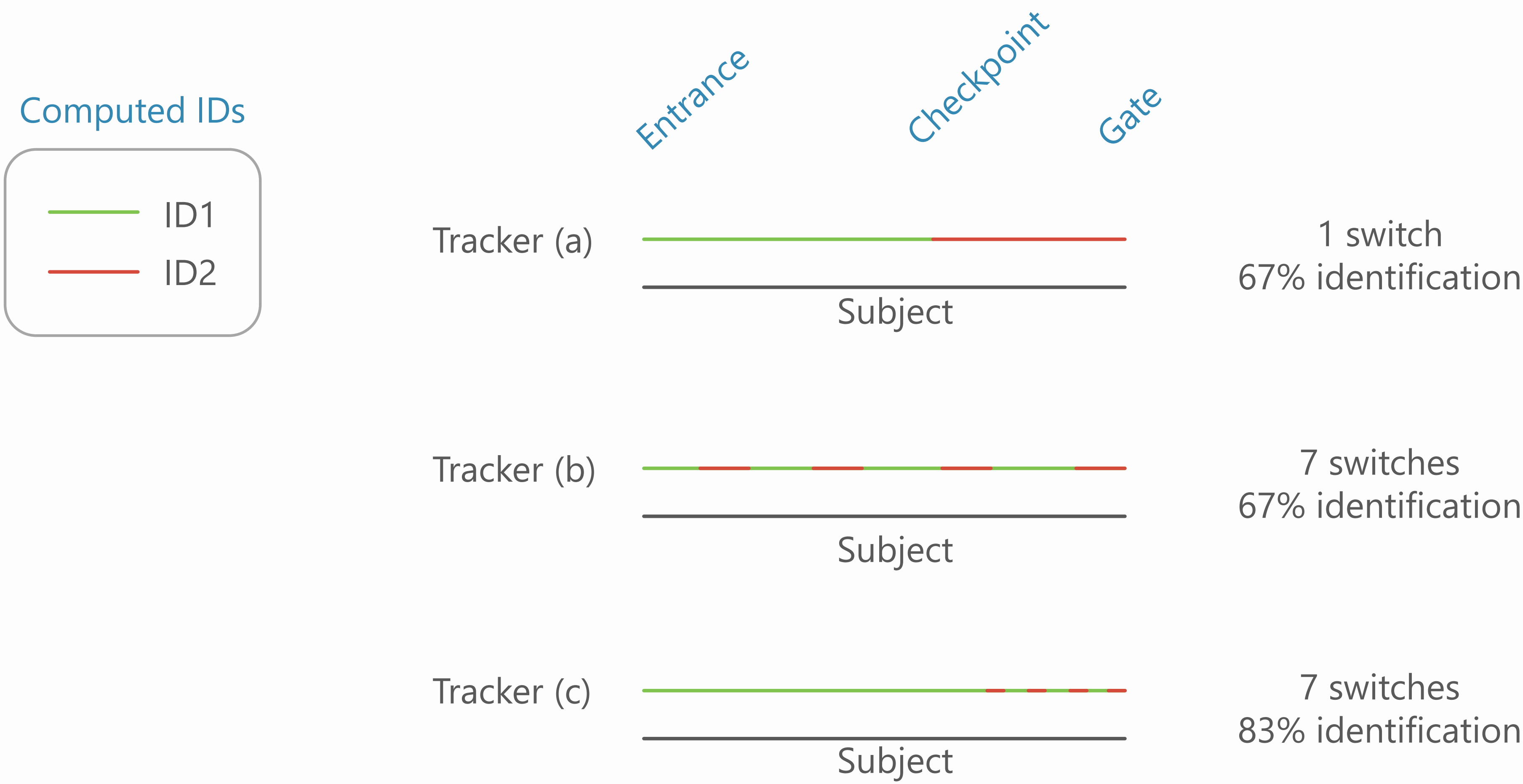
Evaluation Paradigms



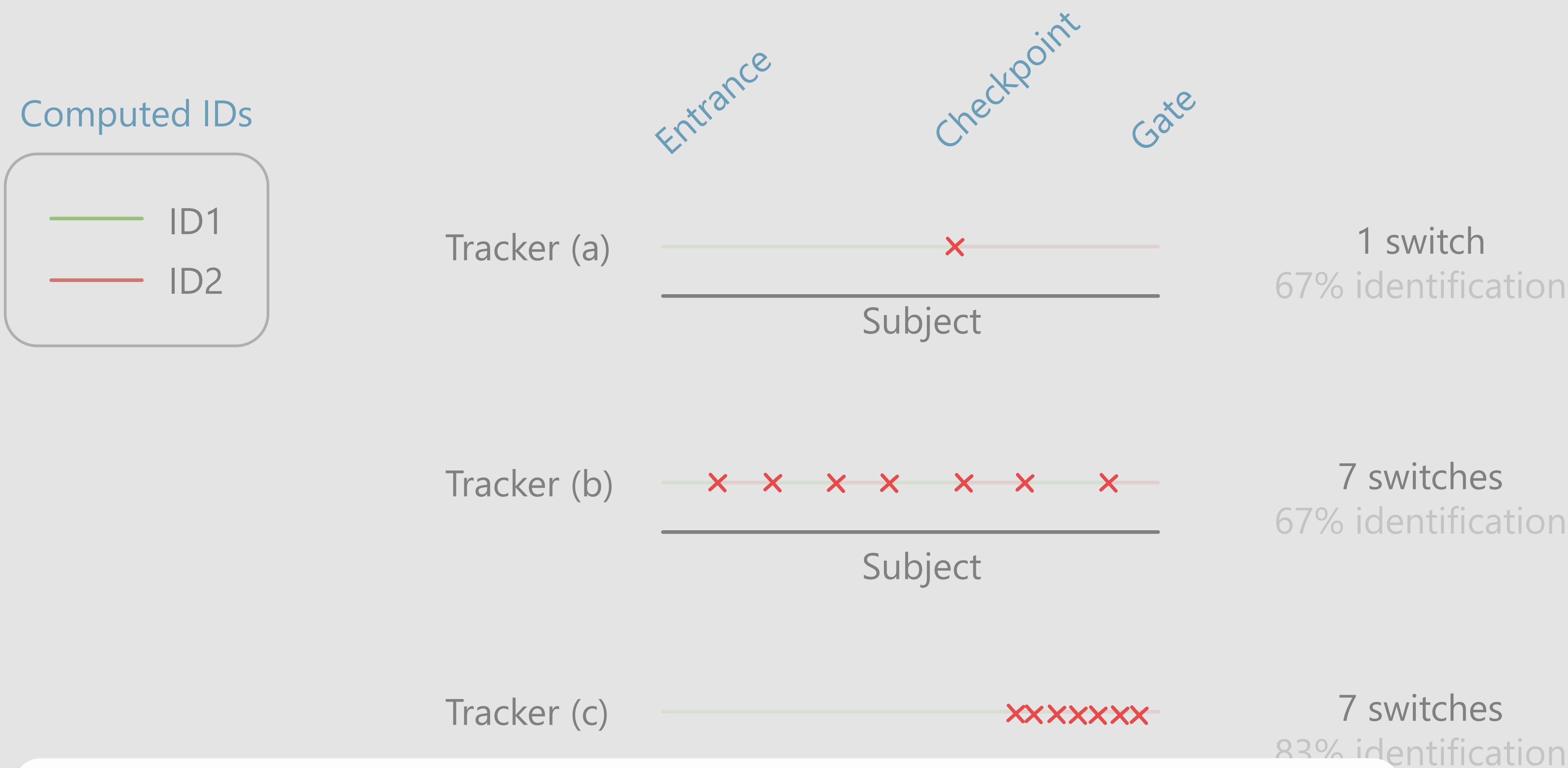
Evaluation Paradigms



Evaluation Paradigms

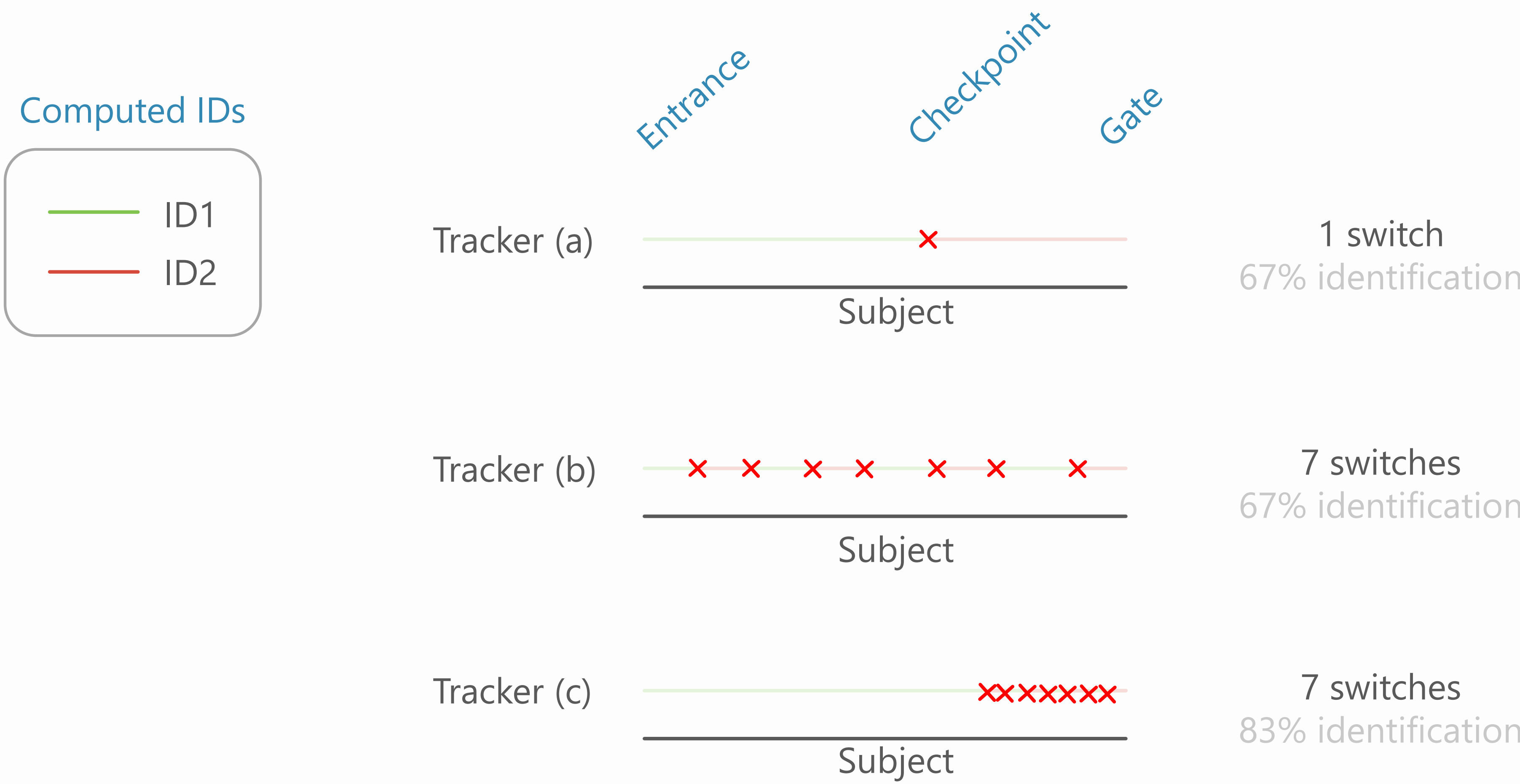


Evaluation Paradigms

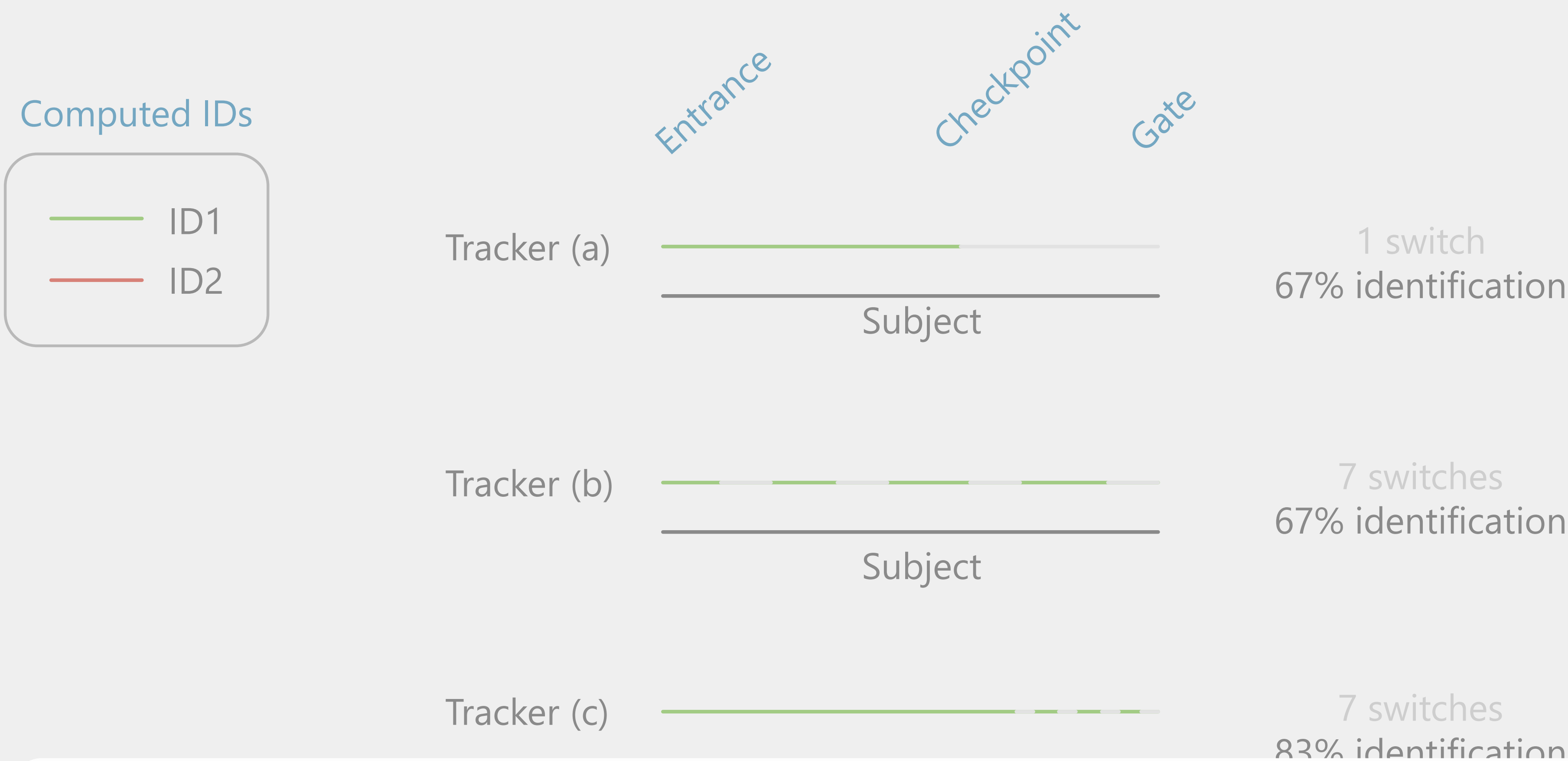


- Researchers: How often does the tracker switch identities?

Evaluation Paradigms

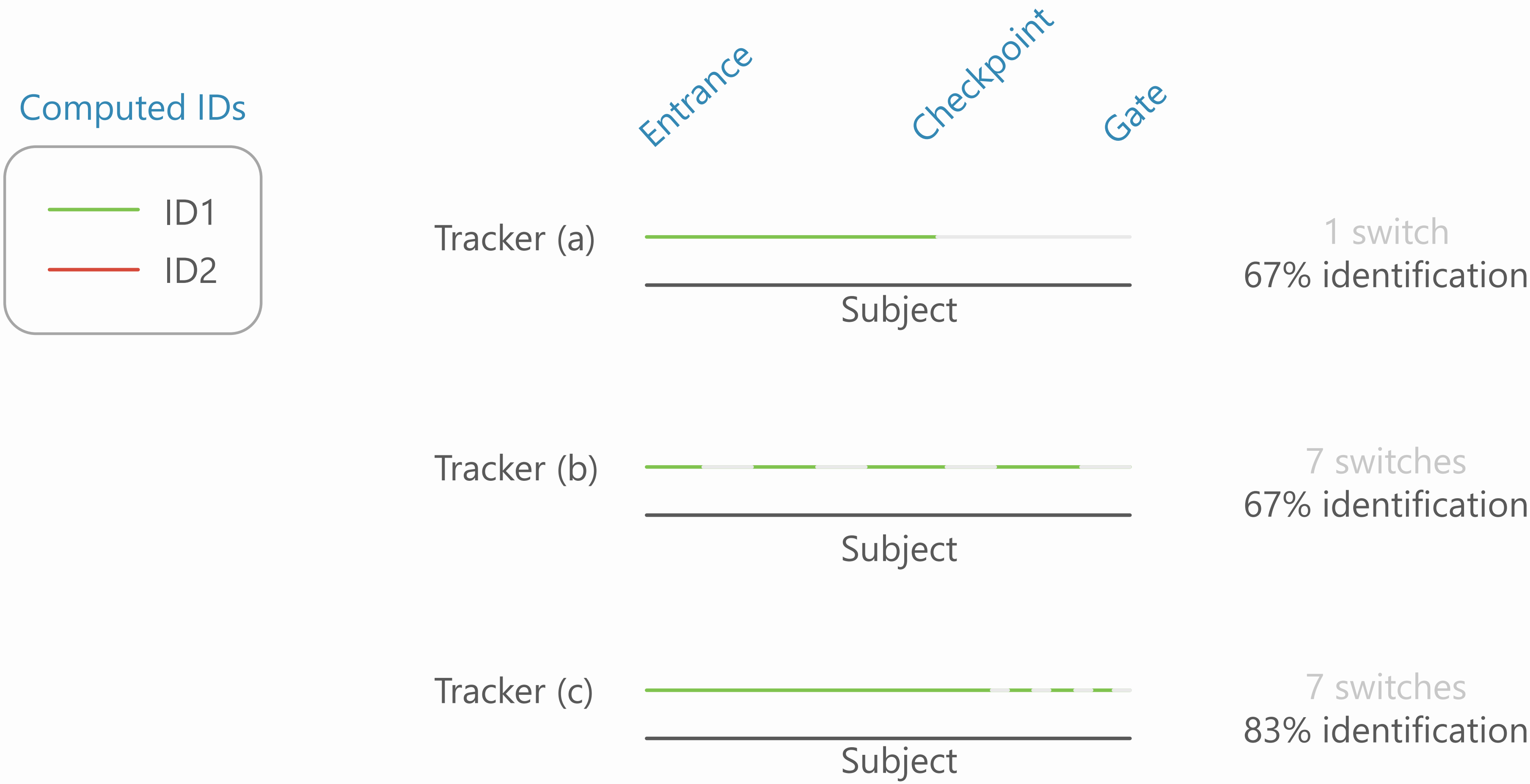


Evaluation Paradigms

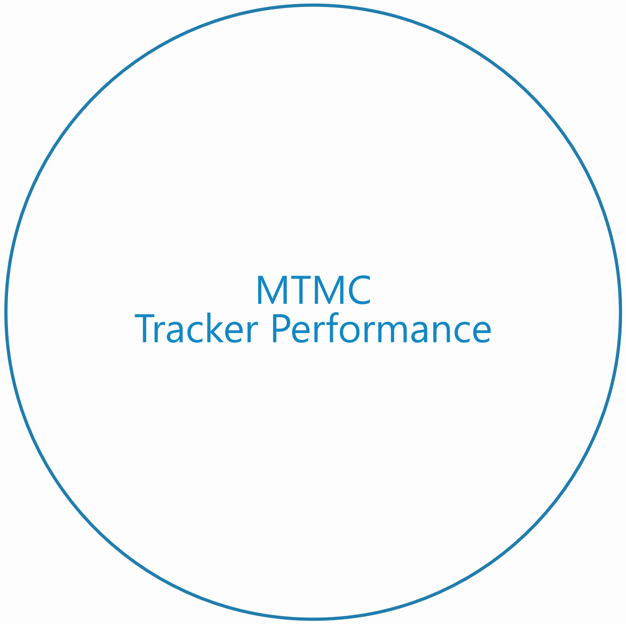


- End-users: How often does the tracker know the position of each identity?

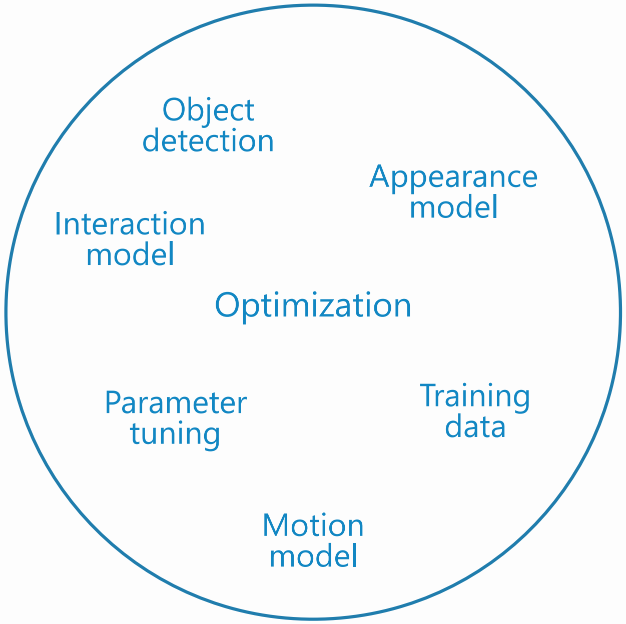
Evaluation Paradigms



Performance Measures



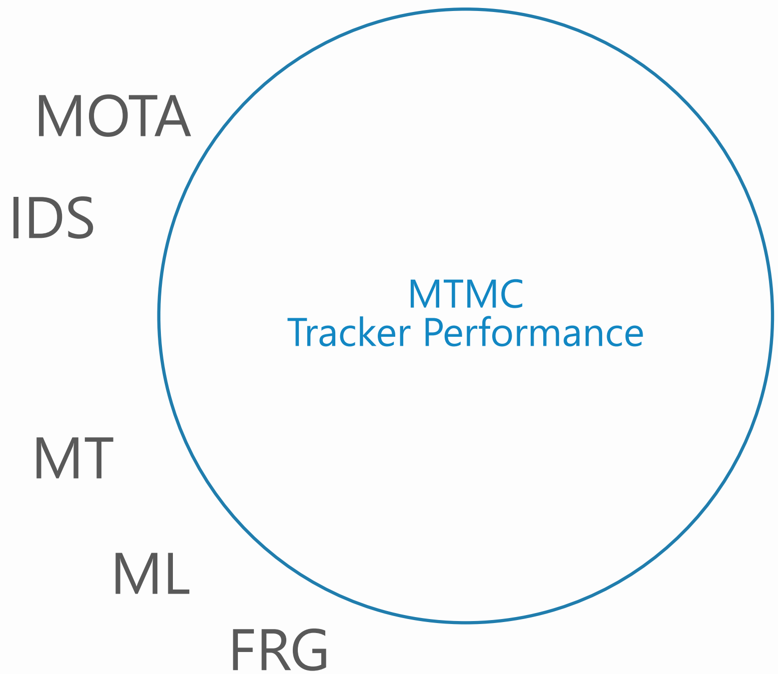
Performance Measures



Performance Measures

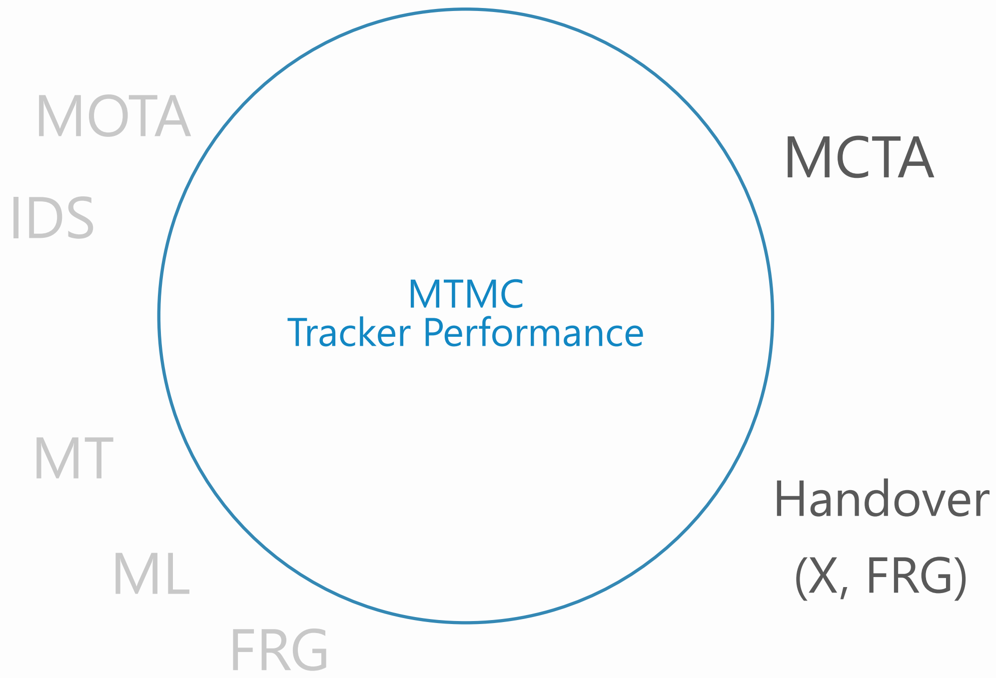
MTMC
Tracker Performance

Performance Measures



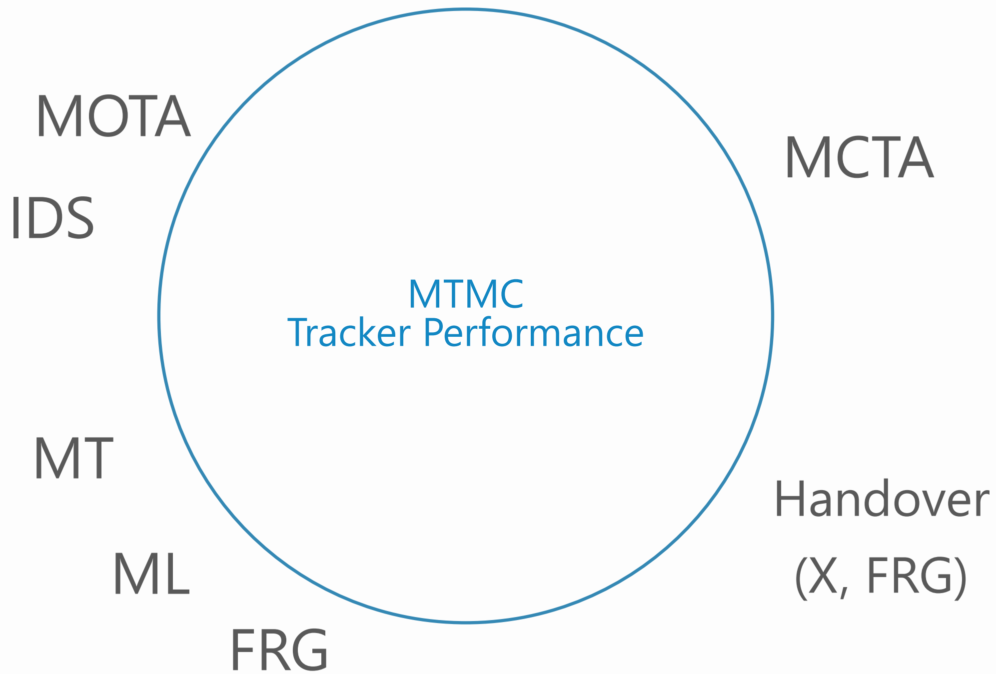
[1] Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics. Image Video Proc. 2008
[2] Learning to Associate: HybridBoosted Multi-Target Tracker for Crowded Scene. CVPR 2009

Performance Measures

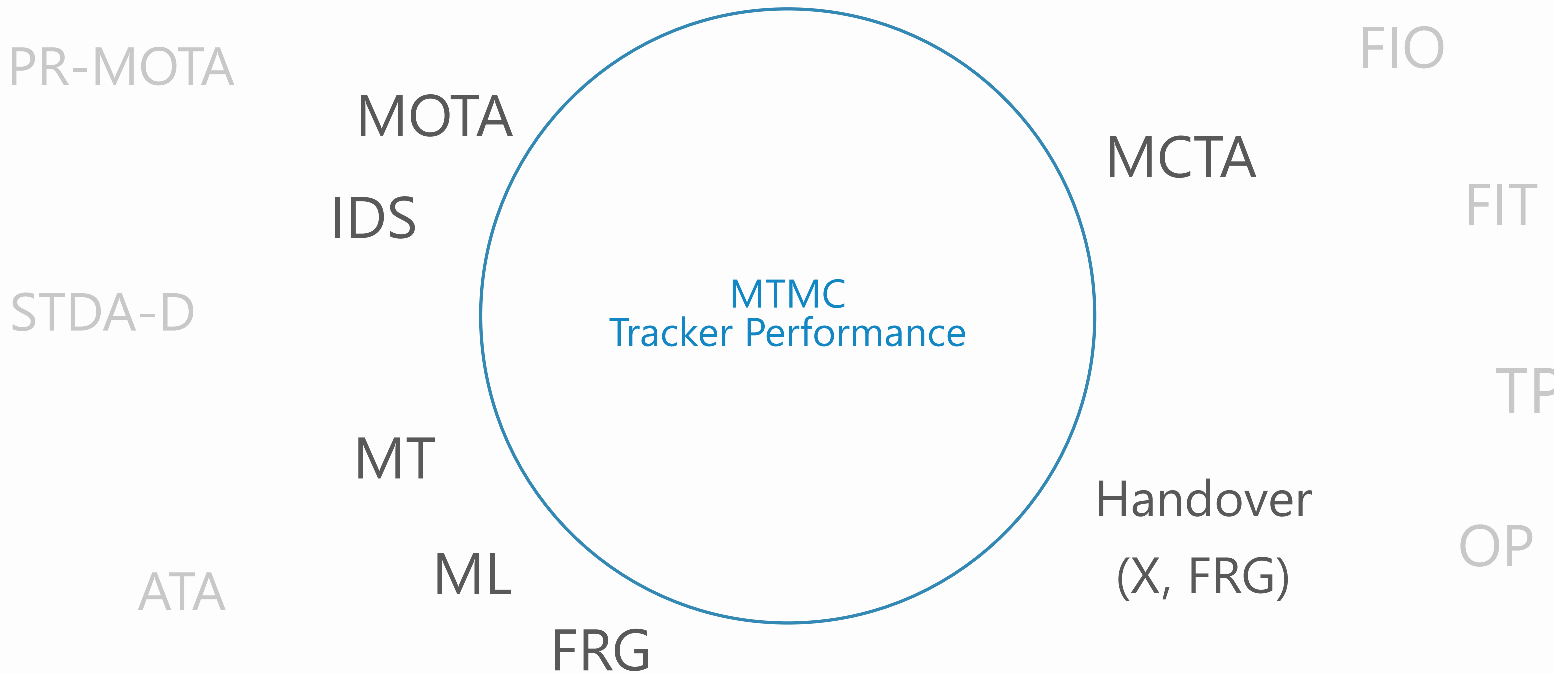


[3] An Equalized Global Graph Model-Based Approach for Multi-Camera Object Tracking. IEEE TCAS 2016
[4] Inter-camera Association of Multi-target Tracks by On-Line Learned Appearance Affinity Models. ECCV 2010

Performance Measures

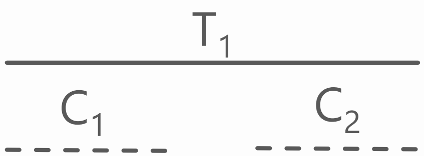


Performance Measures



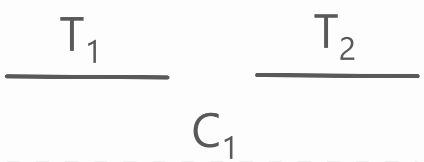
Errors in Tracking

- Mismatches
  - Fragmentations



$$\Phi$$

  - Merges



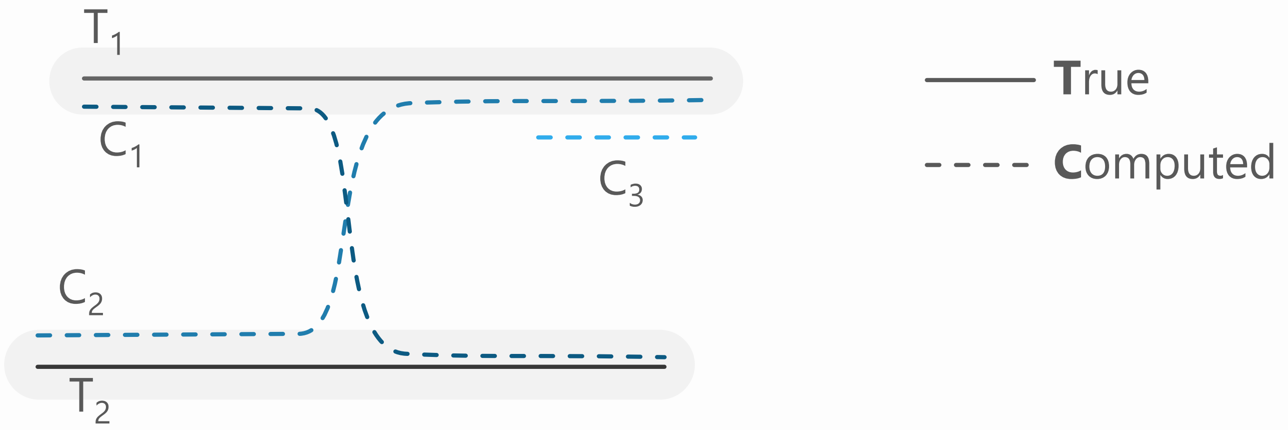
$$M$$

- False Positives/Negatives

$$FN$$

$$FP$$

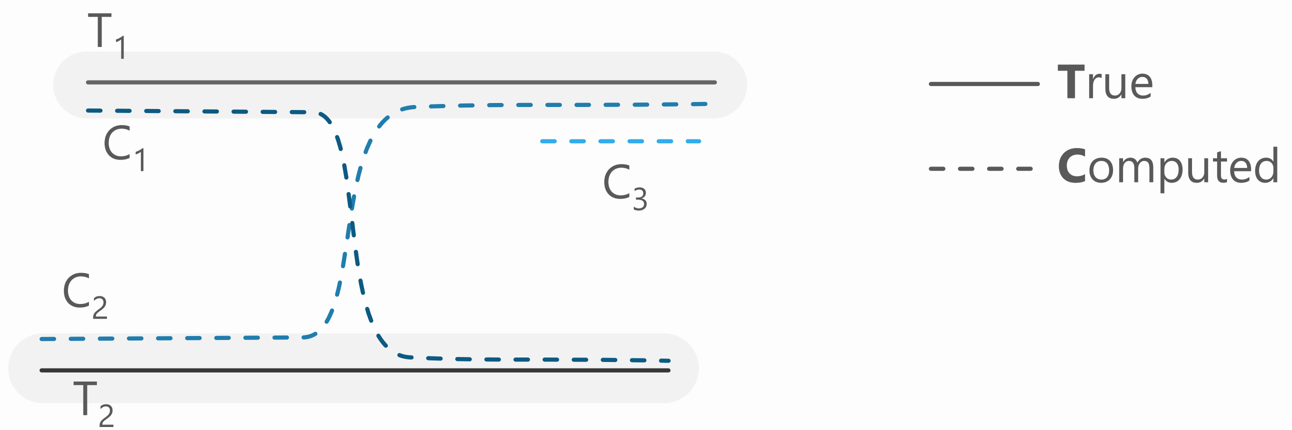
Key Evaluation Steps



Key Evaluation Steps

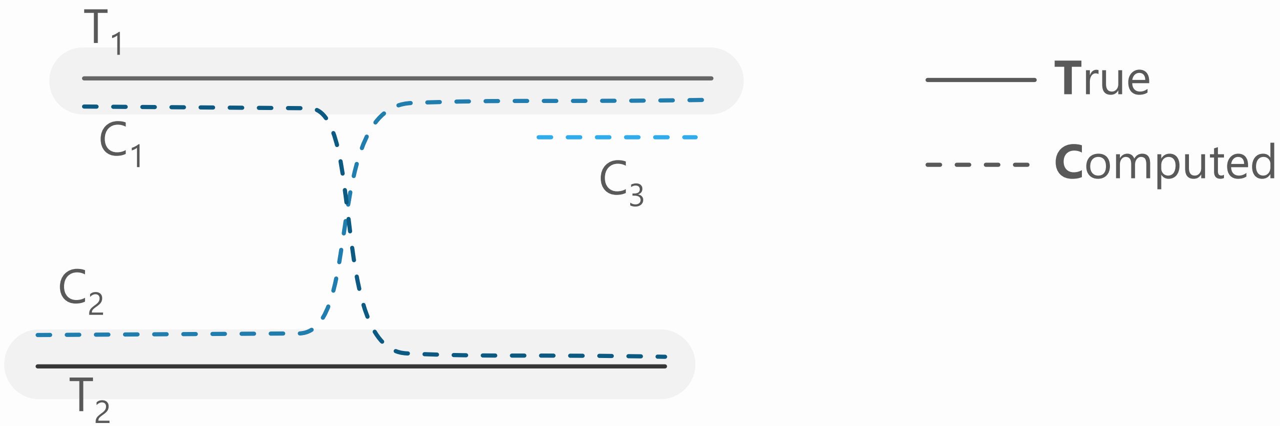
- Truth-to-Result Matching

$$\mu : T \to C$$



Key Evaluation Steps

- Truth-to-Result Matching



$$\mu : T \rightarrow C$$

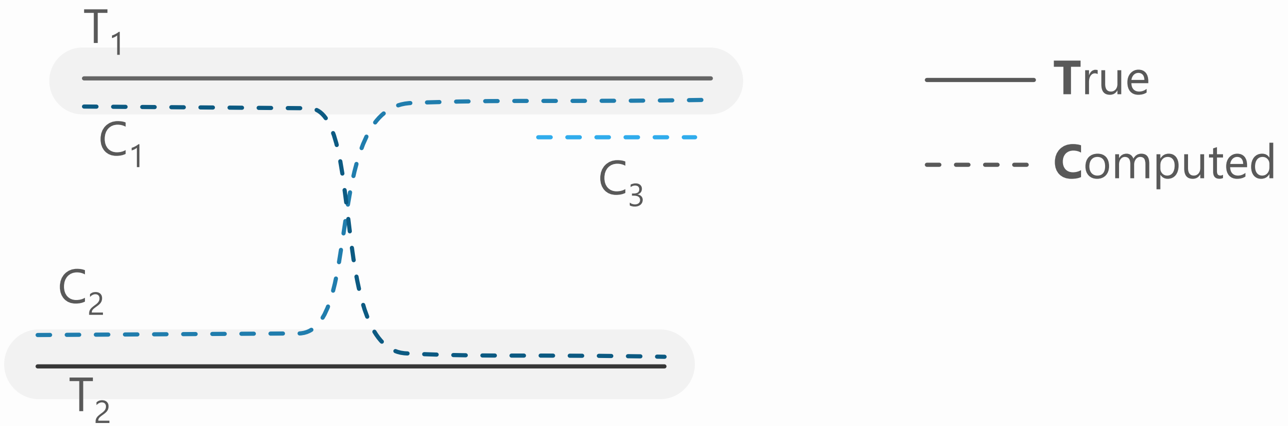
- Scoring Function

$$f(T, C, \mu)$$

Key Evaluation Steps

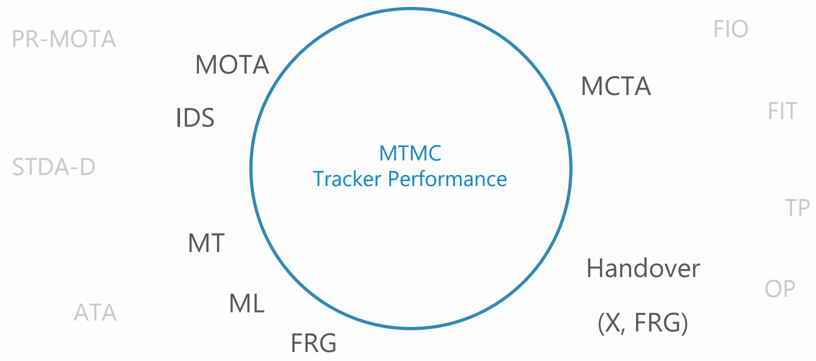
- Truth-to-Result Matching

$$\mu : T \rightarrow C$$



- Scoring Function

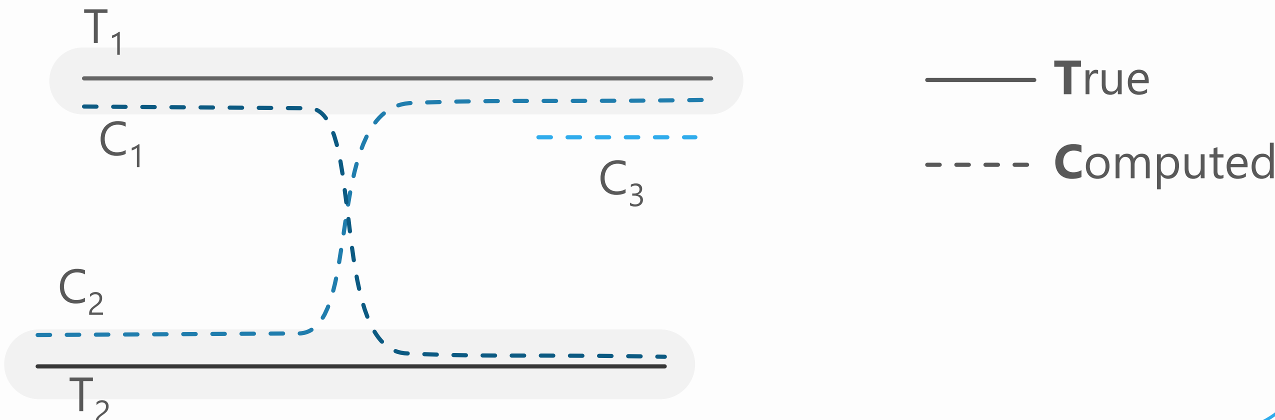
$$f(T, C, \mu)$$



Key Evaluation Steps

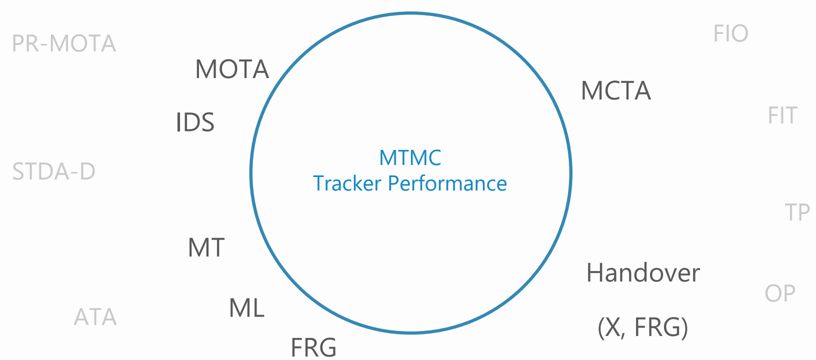
- Truth-to-Result Matching

$$\mu : T \to C$$



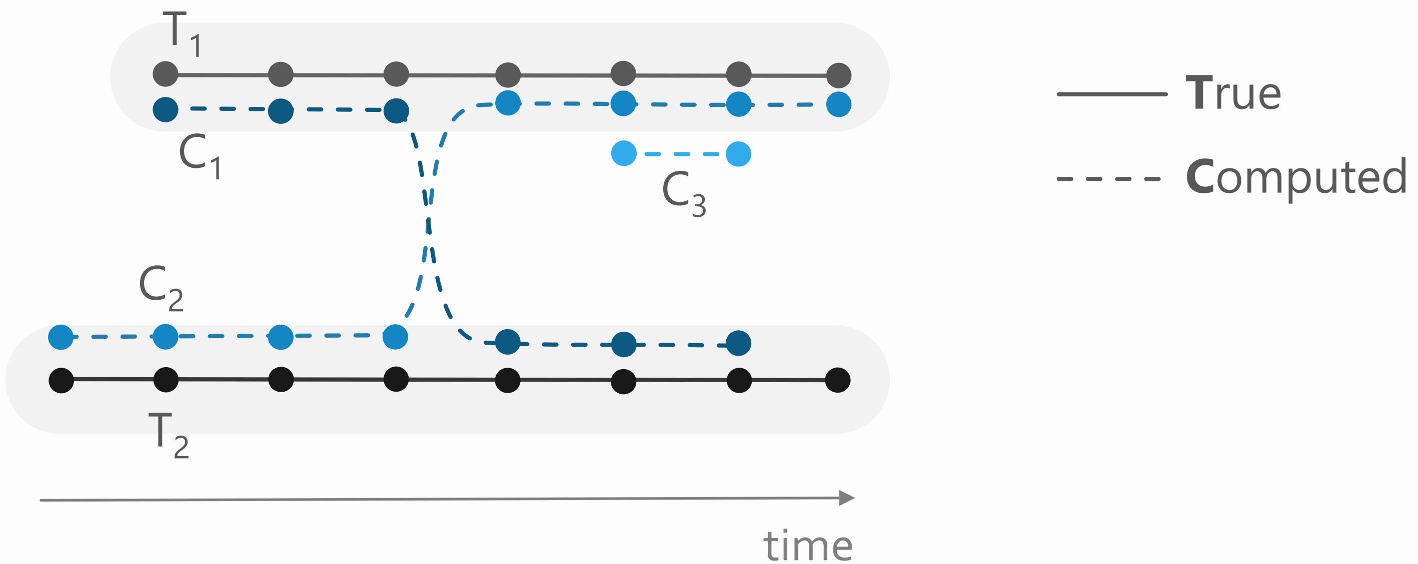
- Scoring Function

$$f(T, C, \mu)$$



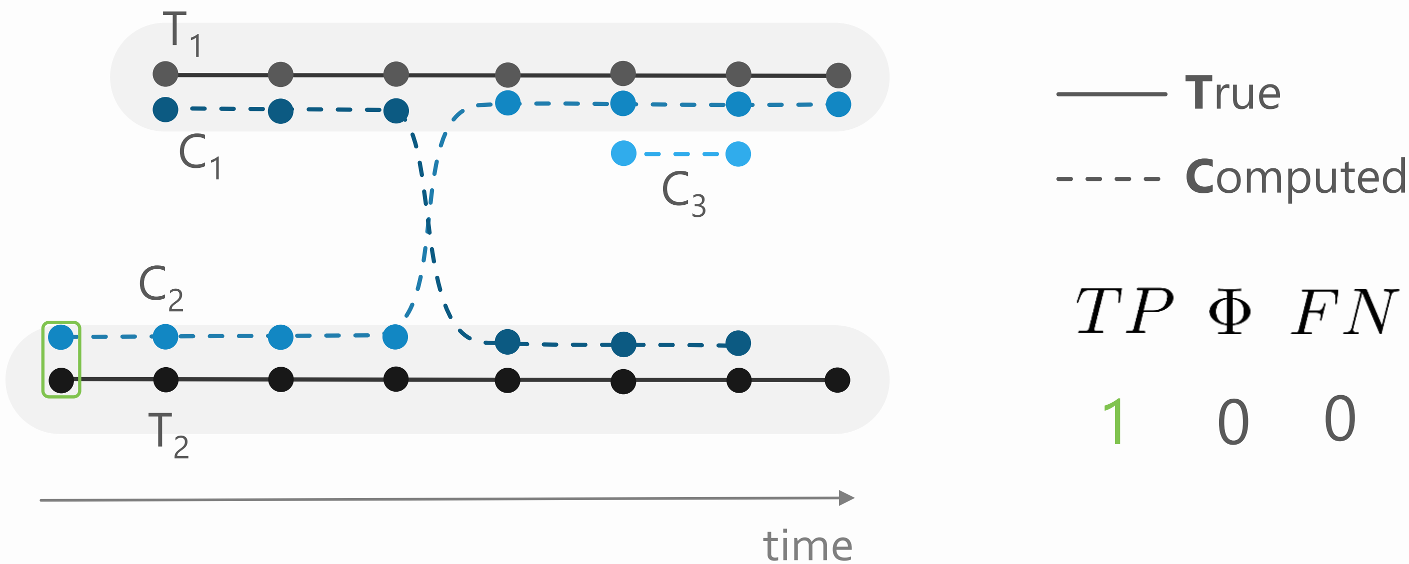
CLEAR MOT Mapping

- At each frame *t*
  - Preserve mapping of detections from frame *t*-1 if still valid
  - Solve bipartite matching at frame *t* for unmapped detections



CLEAR MOT Mapping

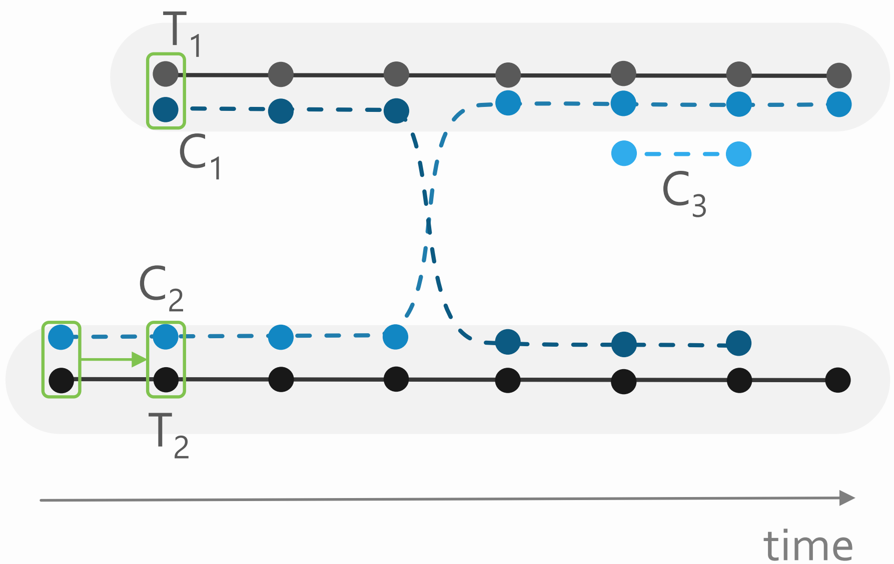
- At each frame *t*
  - Preserve mapping of detections from frame *t*-1 if still valid
  - Solve bipartite matching at frame *t* for unmapped detections



TP	Φ	FN	FP
1	0	0	0

CLEAR MOT Mapping

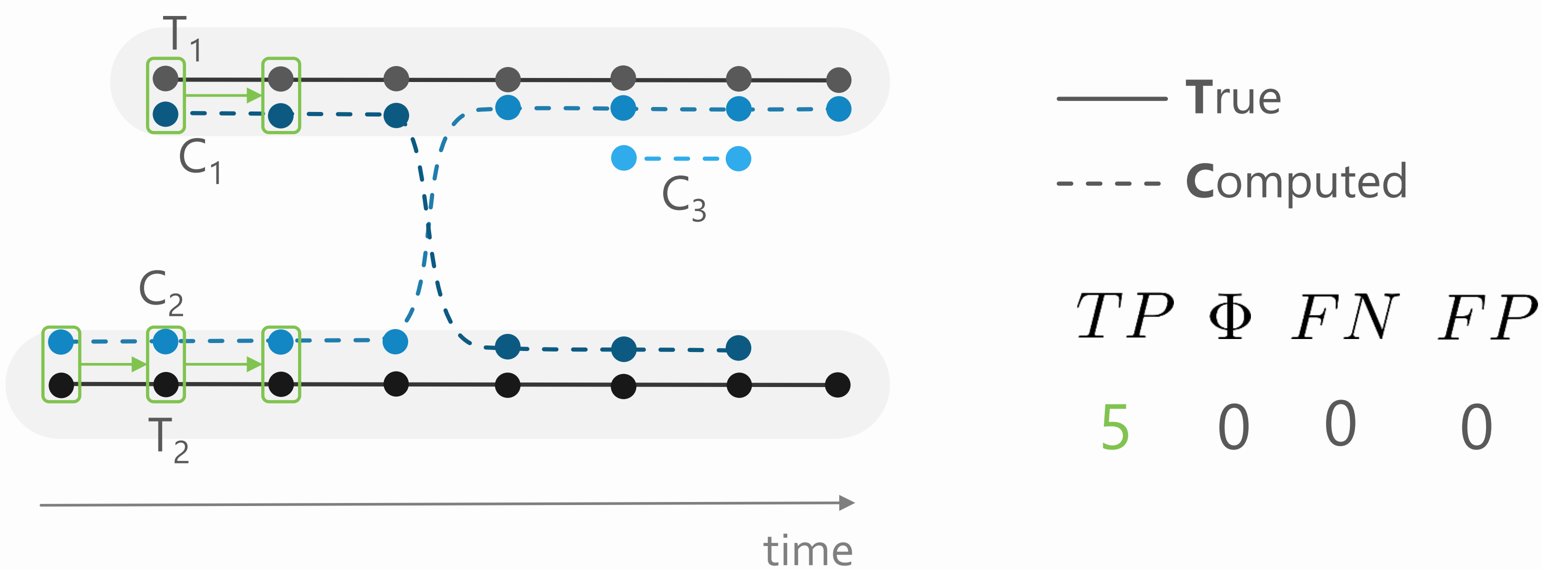
- At each frame *t*
  - Preserve mapping of detections from frame *t*-1 if still valid
  - Solve bipartite matching at frame *t* for unmapped detections



TP	Φ	FN	FP
3	0	0	0

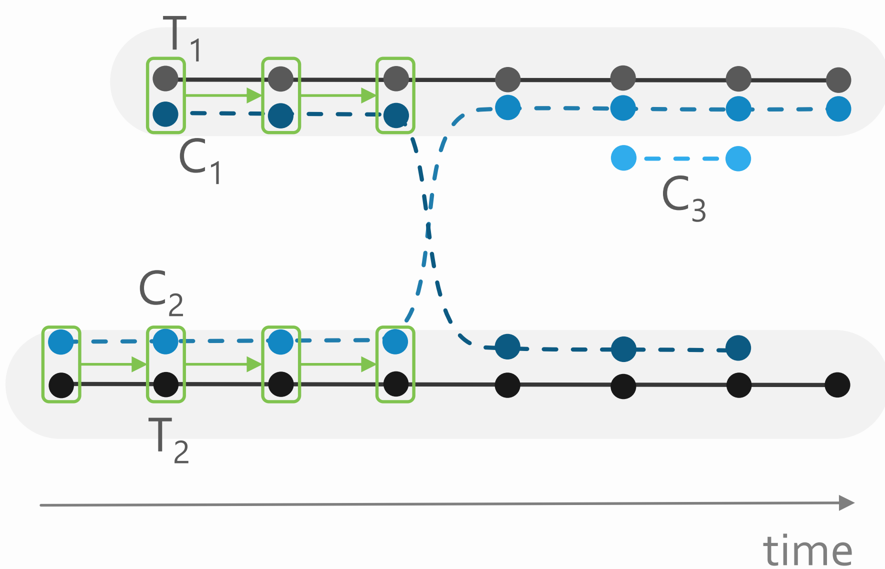
CLEAR MOT Mapping

- At each frame *t*
  - Preserve mapping of detections from frame *t*-1 if still valid
  - Solve bipartite matching at frame *t* for unmapped detections



CLEAR MOT Mapping

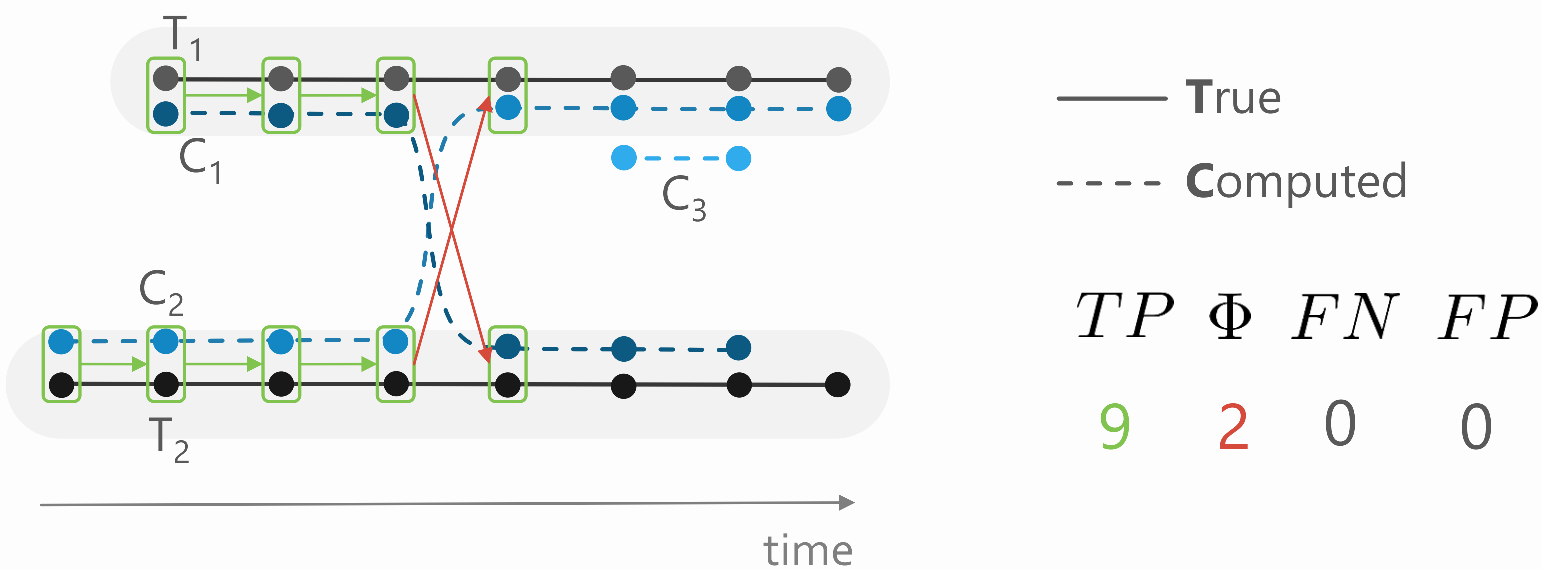
- At each frame *t*
  - Preserve mapping of detections from frame *t*-1 if still valid
  - Solve bipartite matching at frame *t* for unmapped detections



TP	Φ	FN	FP
7	0	0	0

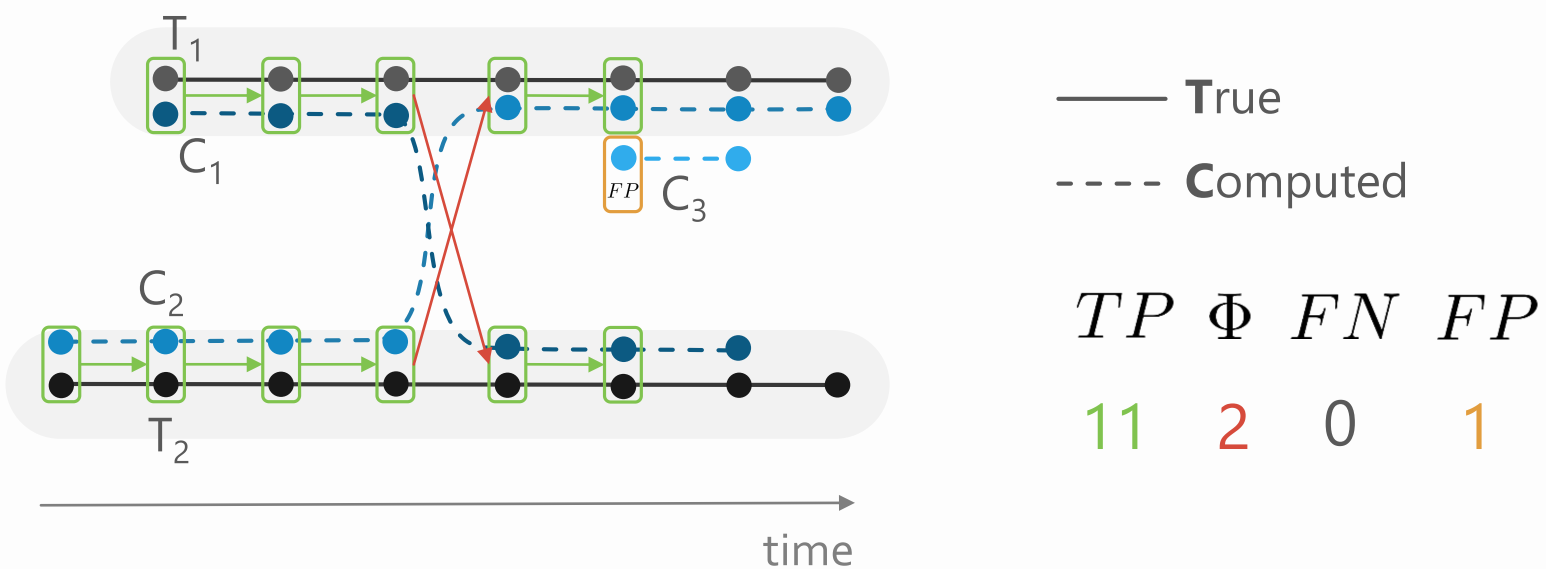
CLEAR MOT Mapping

- At each frame *t*
  - Preserve mapping of detections from frame *t*-1 if still valid
  - Solve bipartite matching at frame *t* for unmapped detections



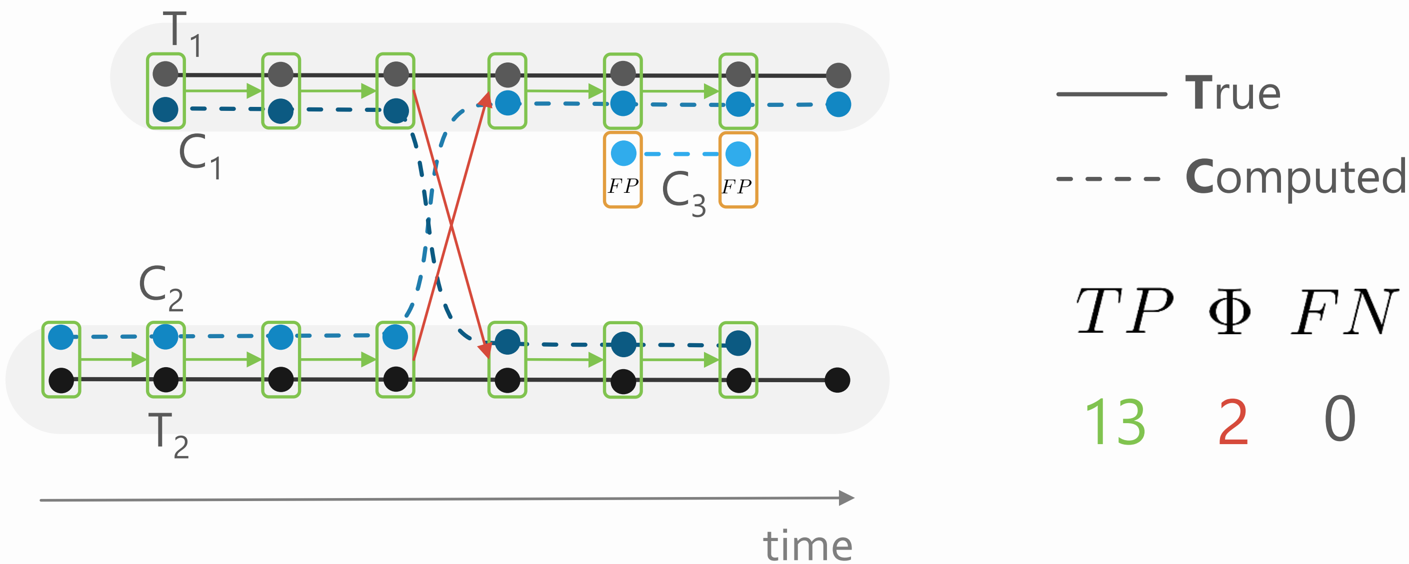
CLEAR MOT Mapping

- At each frame *t*
  - Preserve mapping of detections from frame *t*-1 if still valid
  - Solve bipartite matching at frame *t* for unmapped detections



CLEAR MOT Mapping

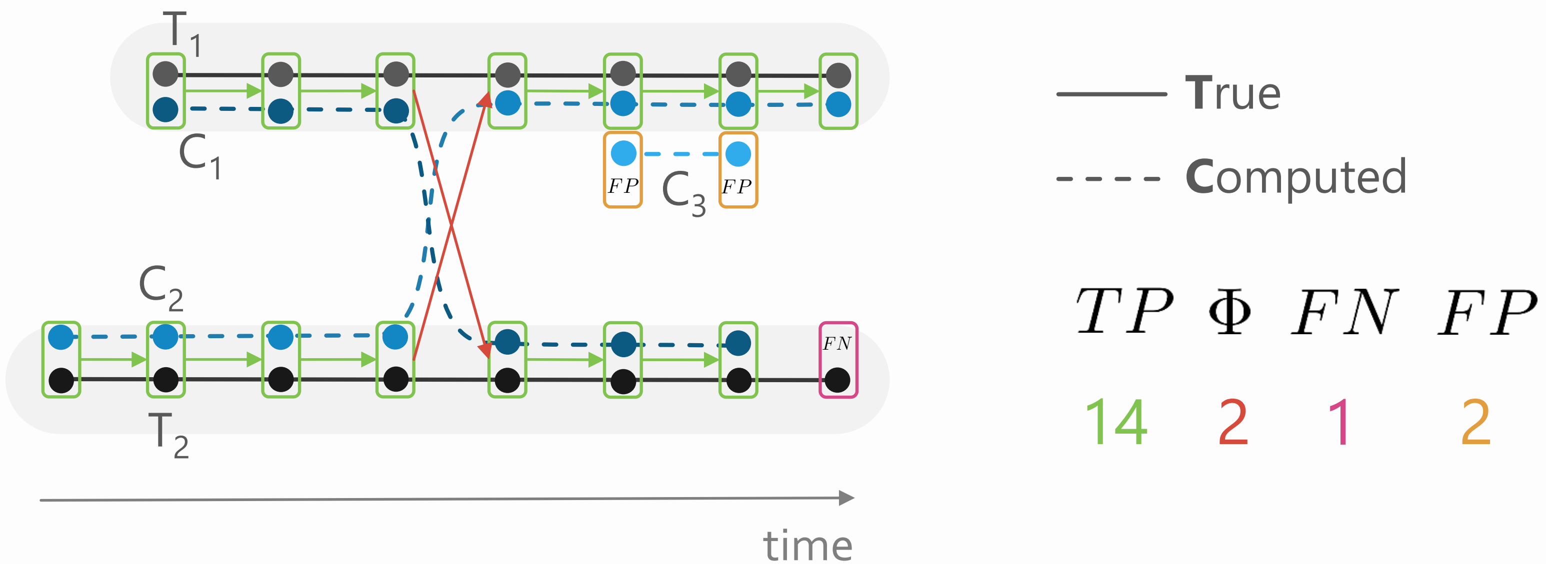
- At each frame *t*
  - Preserve mapping of detections from frame *t*-1 if still valid
  - Solve bipartite matching at frame *t* for unmapped detections



TP	Φ	FN	FP
13	2	0	2

CLEAR MOT Mapping

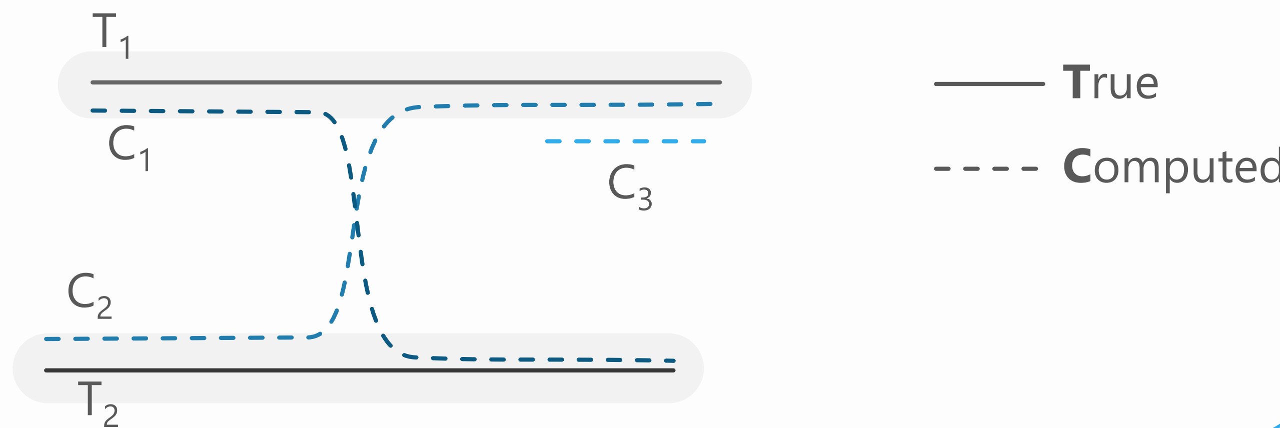
- At each frame *t*
  - Preserve mapping of detections from frame *t*-1 if still valid
  - Solve bipartite matching at frame *t* for unmapped detections



Key Evaluation Steps

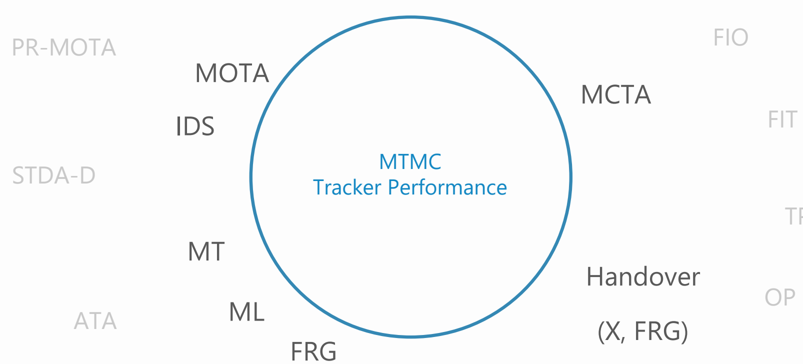
- Truth-to-Result Matching

$$\mu : T \rightarrow C$$



- Scoring Function

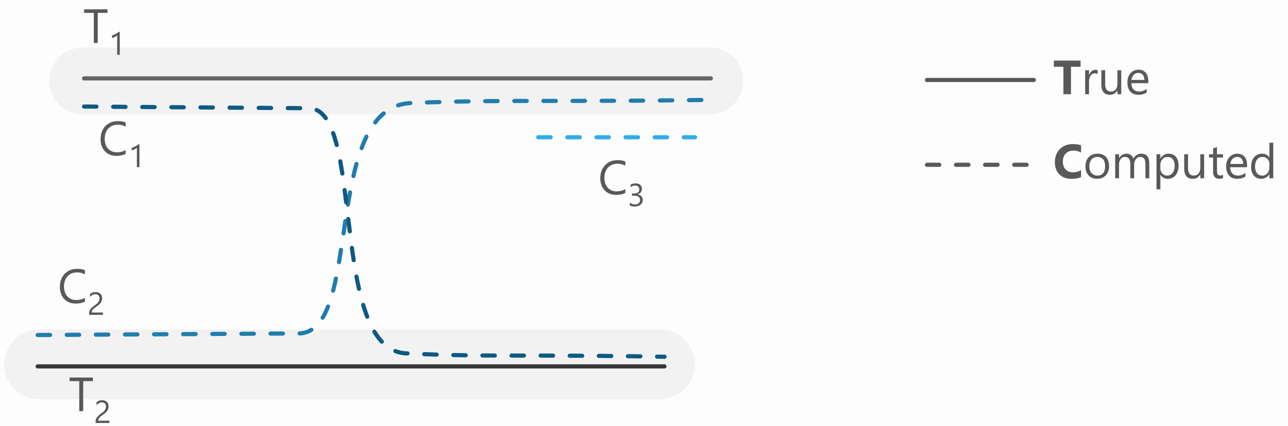
$$f(T, C, \mu)$$



Key Evaluation Steps

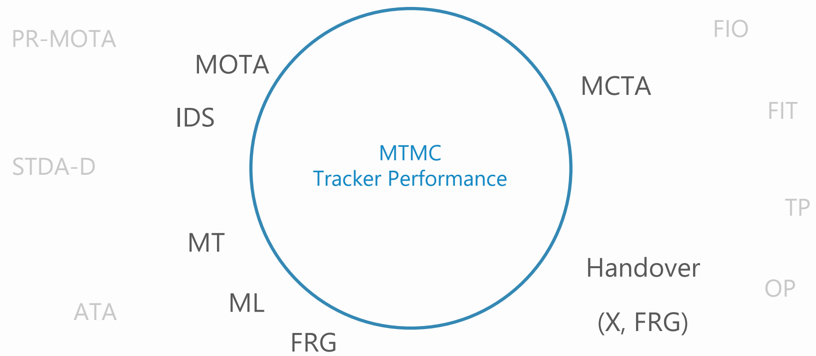
- Truth-to-Result Matching

$$\mu : T \to C$$



- Scoring Function

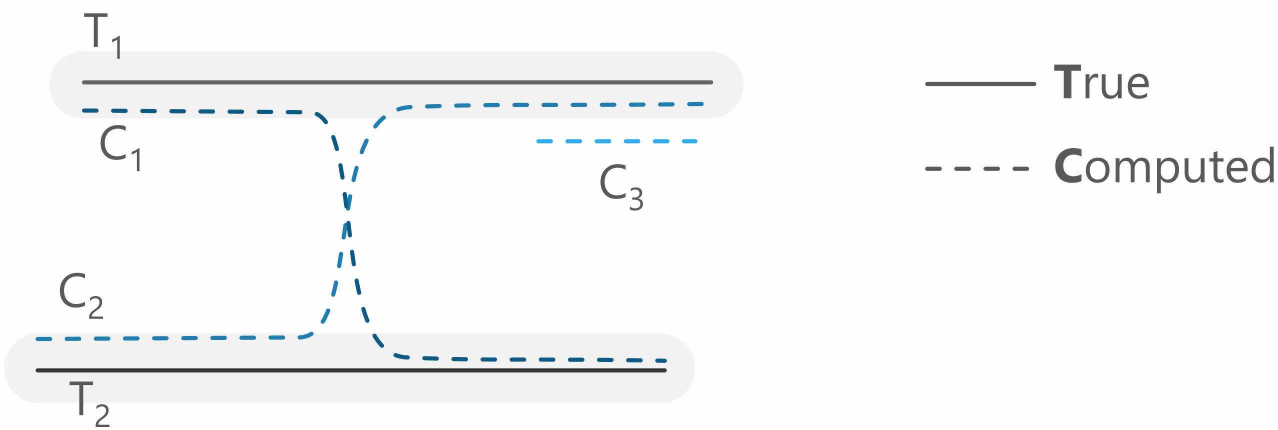
$$f(T, C, \mu)$$



Key Evaluation Steps

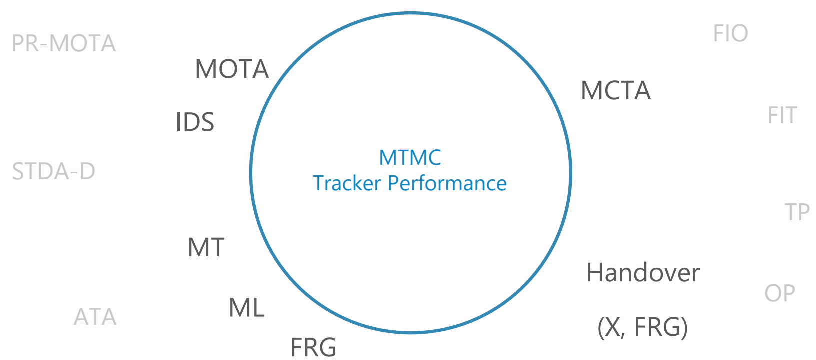
- Truth-to-Result Matching

$$\mu : T \rightarrow C$$



- Scoring Function

$$f(T, C, \mu)$$



Common Scoring Functions

- Multiple Object Tracking Accuracy (MOTA)

$$\mathrm{MOTA} = 1 - \frac{FN + FP + \Phi + M}{T} \qquad [-\infty, 1]$$

- Multiple Camera Tracking Accuracy (MCTA)

$$\mathrm{MCTA} = \underbrace{\frac{2PR}{P+R}}_{F_1} \underbrace{\left(1 - \frac{\Phi^w + M^w}{T^w}\right)}_{\text{within camera}} \underbrace{\left(1 - \frac{\Phi^h + M^h}{T^h}\right)}_{\text{handover}} \qquad [0, 1]$$

- Trajectory Scores

MT, PT, ML $\qquad \mathbb{Z}^*$

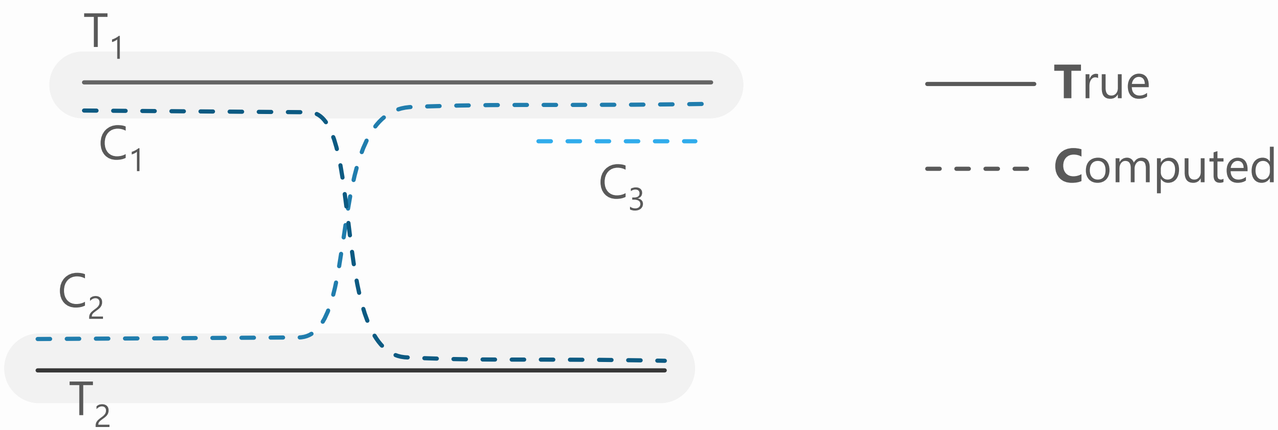
- Handover Errors

$\Phi^h, M^h \qquad \mathbb{Z}^*$

Issues

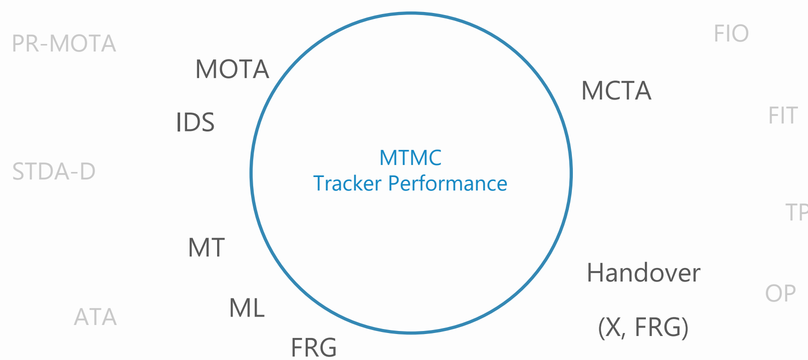
- Truth-to-Result Matching

$$\mu : T \to C$$



- Scoring Function

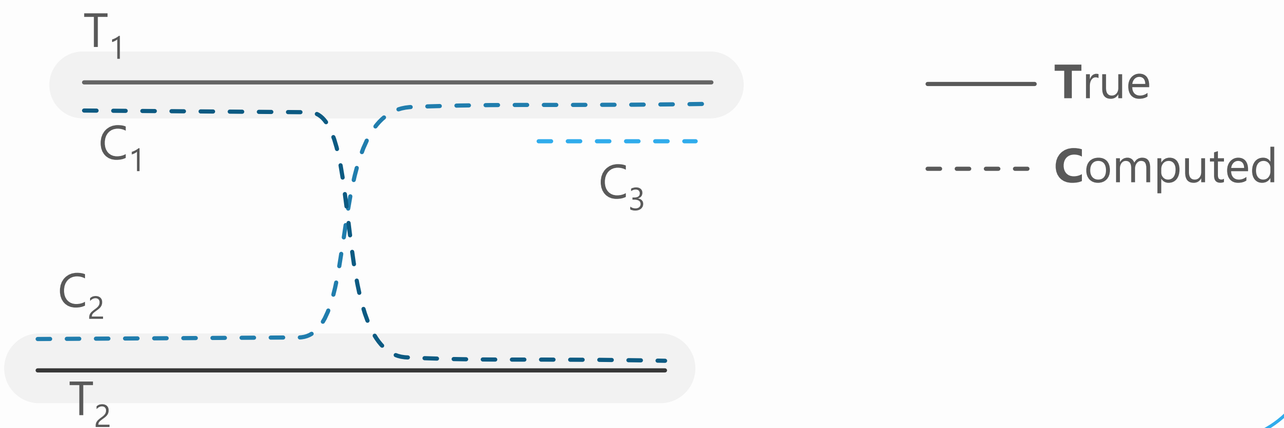
$$f(T, C, \mu)$$



Issues

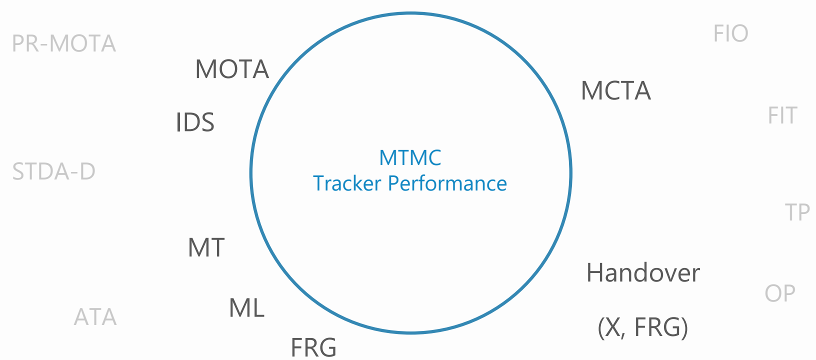
- Truth-to-Result Matching

$$\mu : T \to C$$



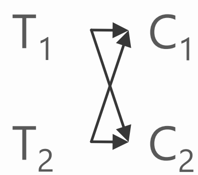
- Scoring Function

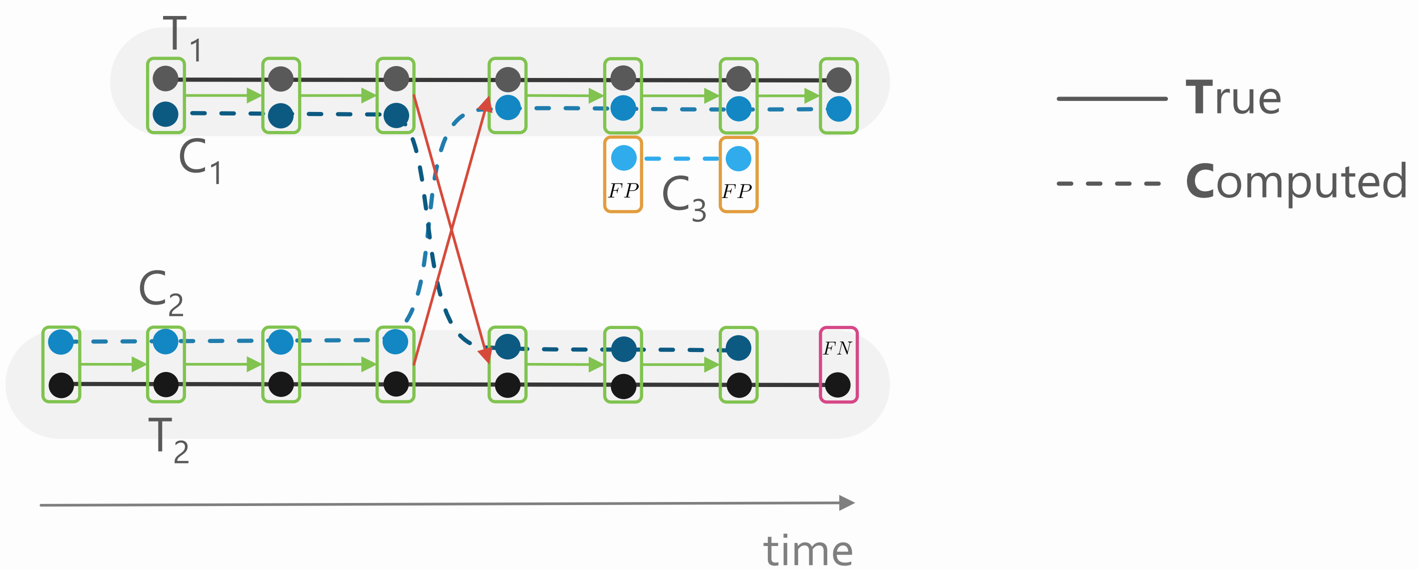
$$f(T, C, \mu)$$



CLEAR MOT Mapping

- Identity mapping bijective at each frame
- Not bijective overall

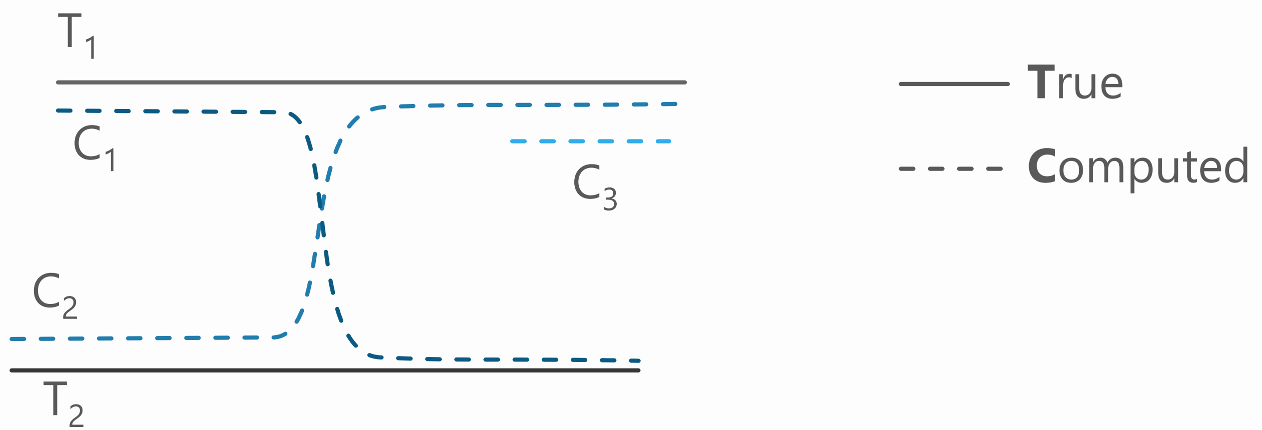




Issues

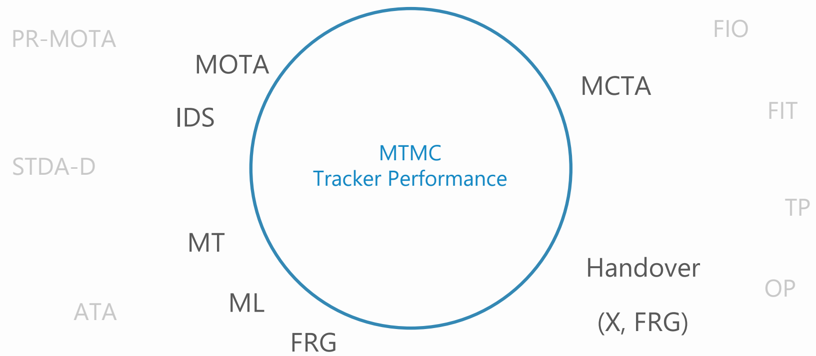
- Truth-to-Result Matching

$$\mu : T \rightarrow C$$



- Scoring Function

$$f(T, C, \mu)$$



Scoring Functions

- Multiple Object Tracking Accuracy (MOTA)

$$\mathrm{MOTA} = 1 - \frac{FN + FP + \Phi + M}{T} \qquad [-\infty, 1]$$

- Multiple Camera Tracking Accuracy (MCTA)

$$\mathrm{MCTA} = \underbrace{\frac{2PR}{P+R}}_{F_1} \underbrace{\left(1 - \frac{\Phi^w + M^w}{T^w}\right)}_{\text{within camera}} \underbrace{\left(1 - \frac{\Phi^h + M^h}{T^h}\right)}_{\text{handover}} \qquad [0, 1]$$

- Trajectory Scores

MT, PT, ML $\qquad \mathbb{Z}^*$

- Handover Errors

$\Phi^h, M^h \qquad \mathbb{Z}^*$

Scoring Functions

- Multiple Object Tracking Accuracy (MOTA)

$$\mathrm{MOTA} = 1 - \frac{FN + FP + \Phi + M}{T} \qquad [-\infty, 1]$$

- Multiple Camera Tracking Accuracy (MCTA)

$$\mathrm{MCTA} = \underbrace{\frac{2PR}{P+R}}_{F_1} \underbrace{\left(1 - \frac{\Phi^w + M^w}{T^w}\right)}_{\text{within camera}} \underbrace{\left(1 - \frac{\Phi^h + M^h}{T^h}\right)}_{\text{handover}} \qquad [0, 1]$$

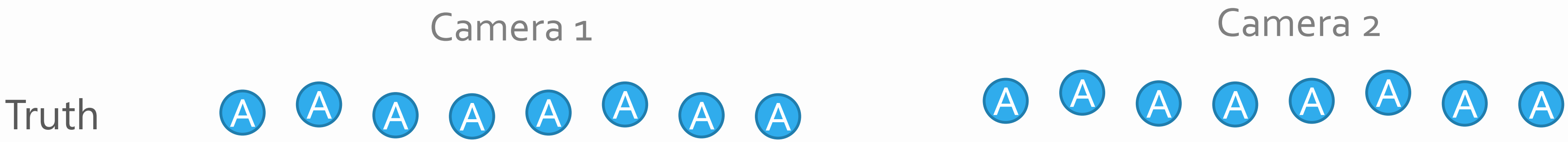
- Trajectory Scores

MT, PT, ML $\qquad \mathbb{Z}^*$

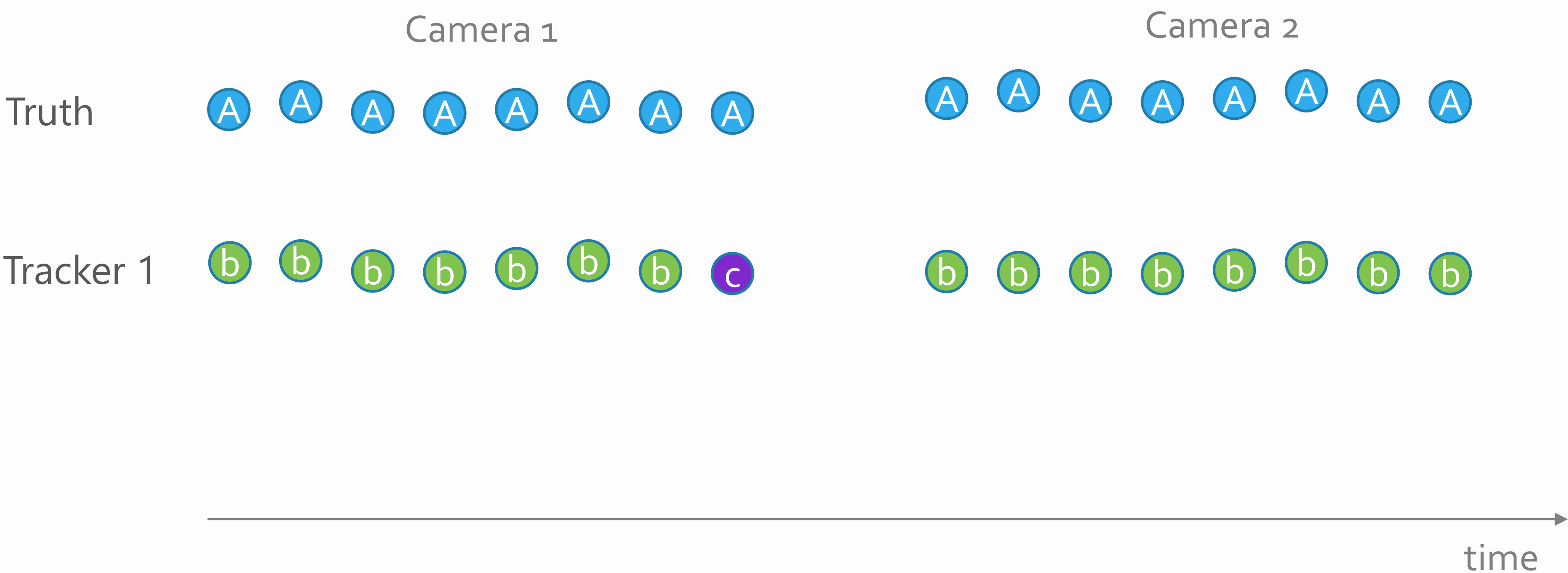
- Handover Errors

Φ^h, M^h Brittle due to mapping $\qquad \mathbb{Z}^*$

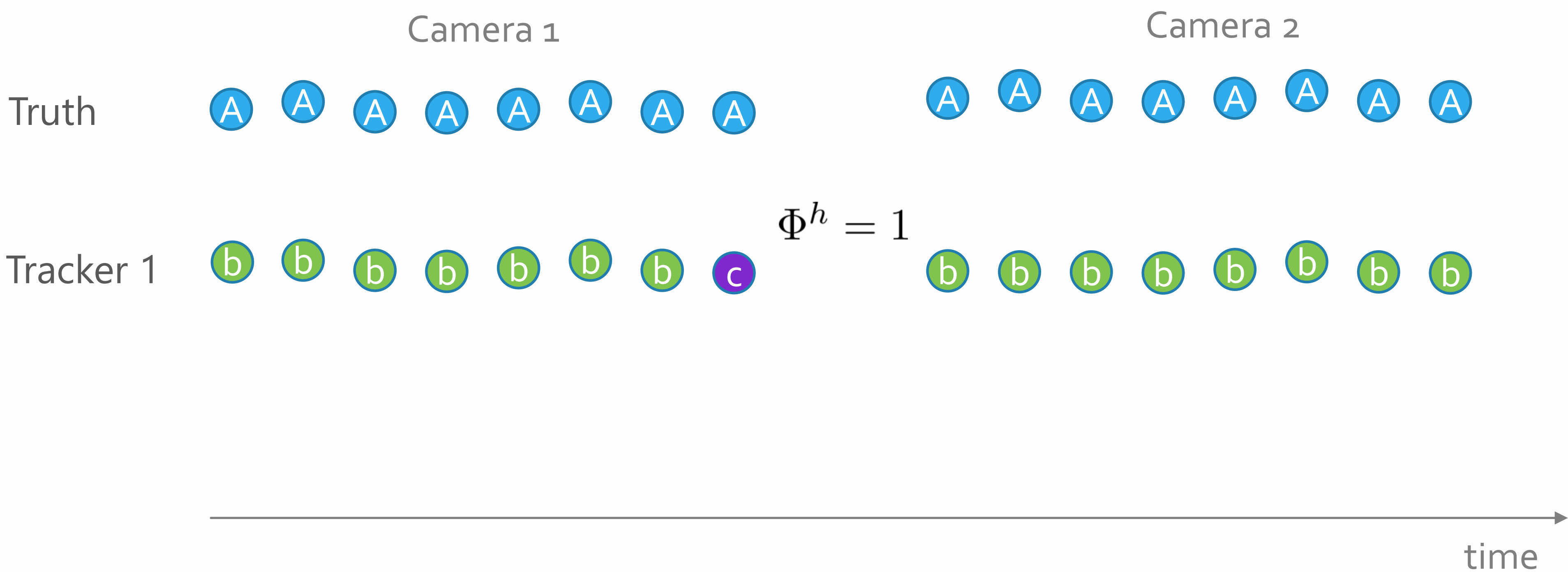
Handover Errors



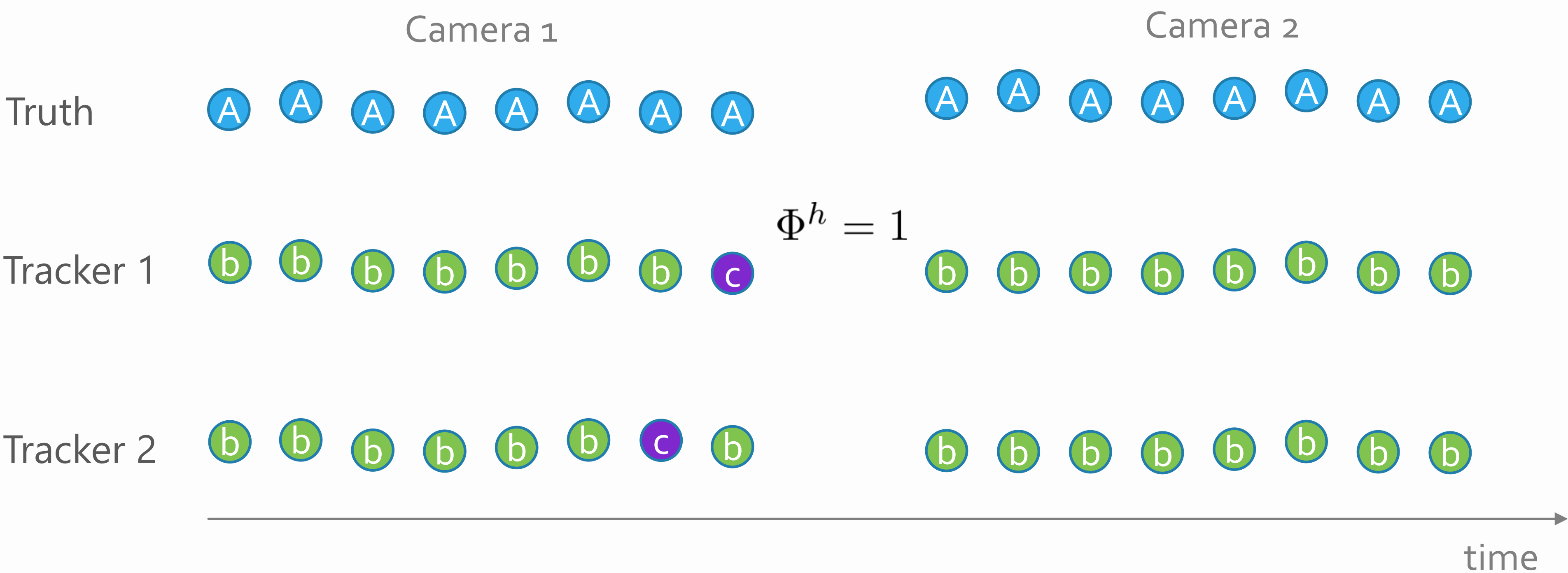
Handover Errors



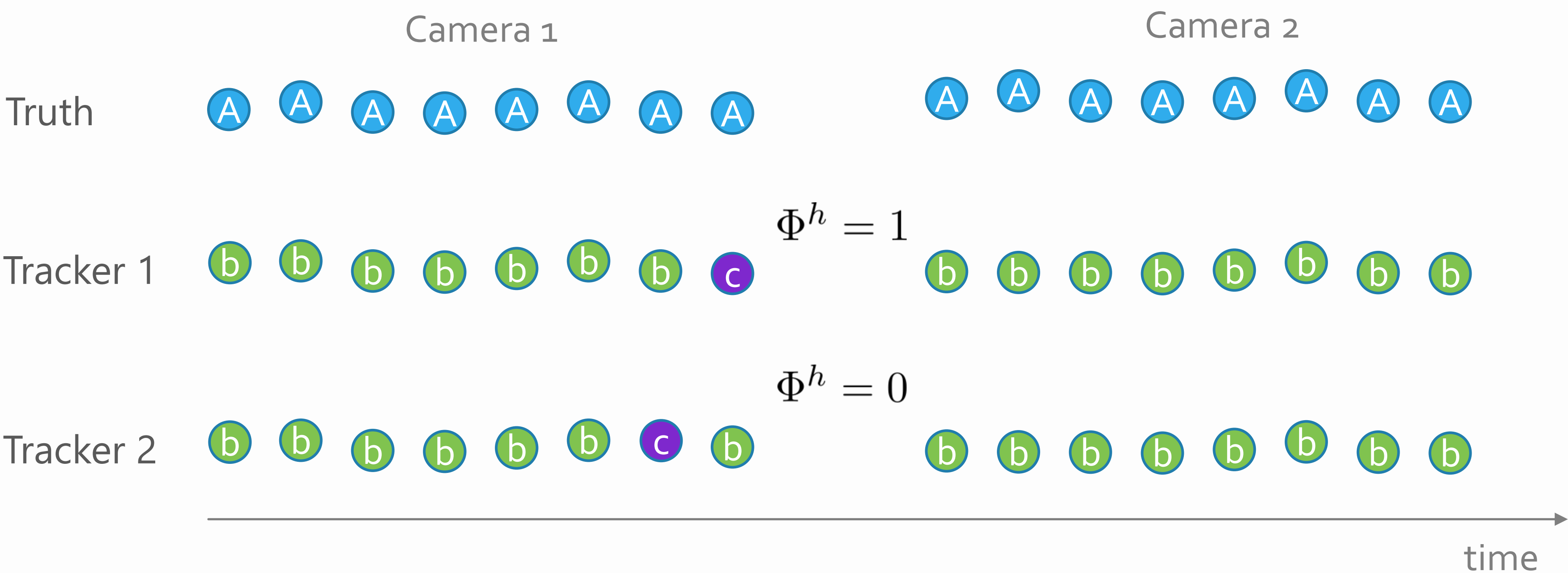
Handover Errors



Handover Errors



Handover Errors



Scoring Functions

- Multiple Object Tracking Accuracy (MOTA)

$$\text{MOTA} = 1 - \frac{FN + FP + \Phi + M}{T} \qquad [-\infty, 1]$$

- Multiple Camera Tracking Accuracy (MCTA)

$$\text{MCTA} = \underbrace{\frac{2PR}{P+R}}_{F_1} \underbrace{\left(1 - \frac{\Phi^w + M^w}{T^w}\right)}_{\text{within camera}} \underbrace{\left(1 - \frac{\Phi^h + M^h}{T^h}\right)}_{\text{handover}} \qquad [0, 1]$$

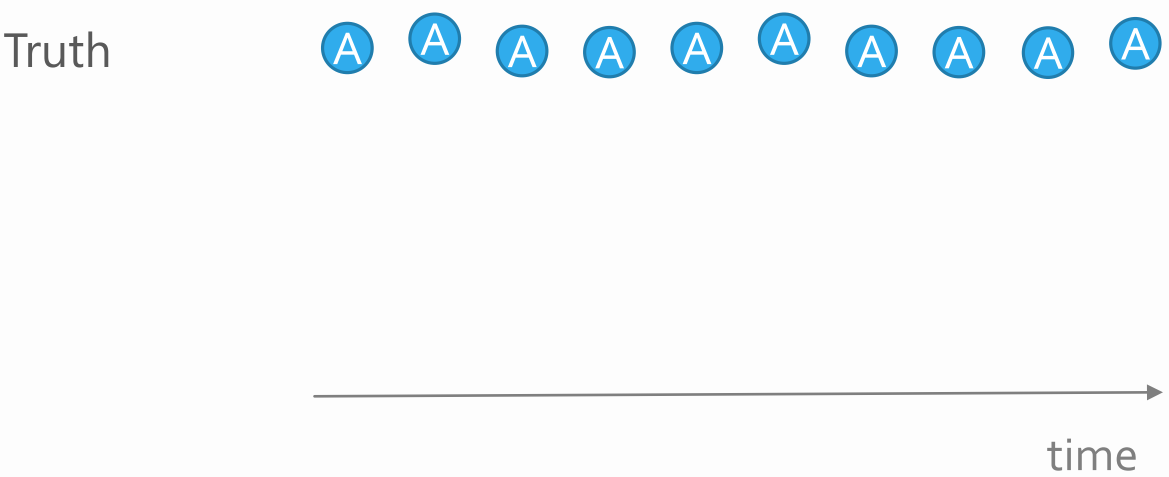
- Trajectory Scores

MT, PT, ML Ignore identity $\mathbb{Z}^*$

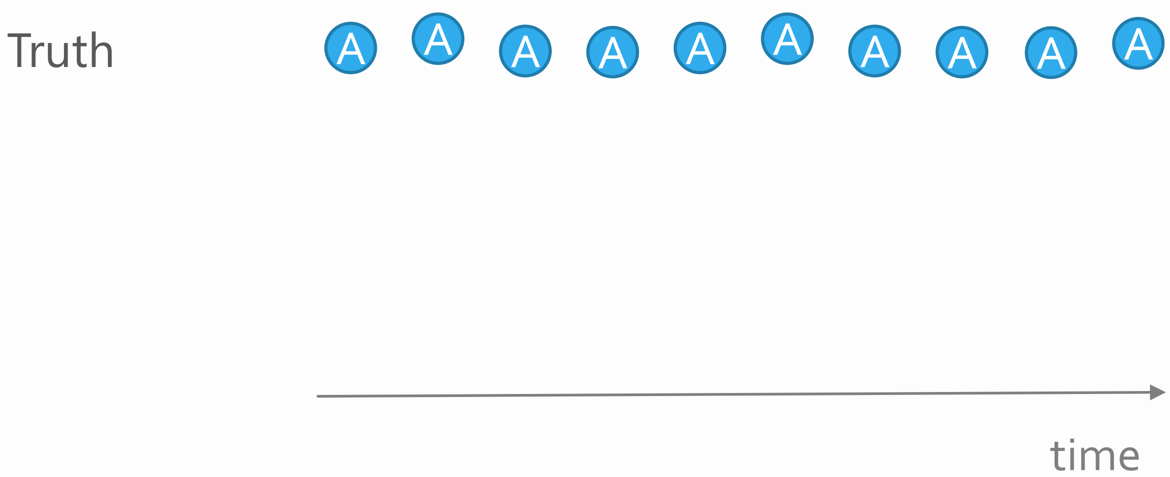
- Handover Errors

Φ^h, M^h $\mathbb{Z}^*$

Trajectory Scores

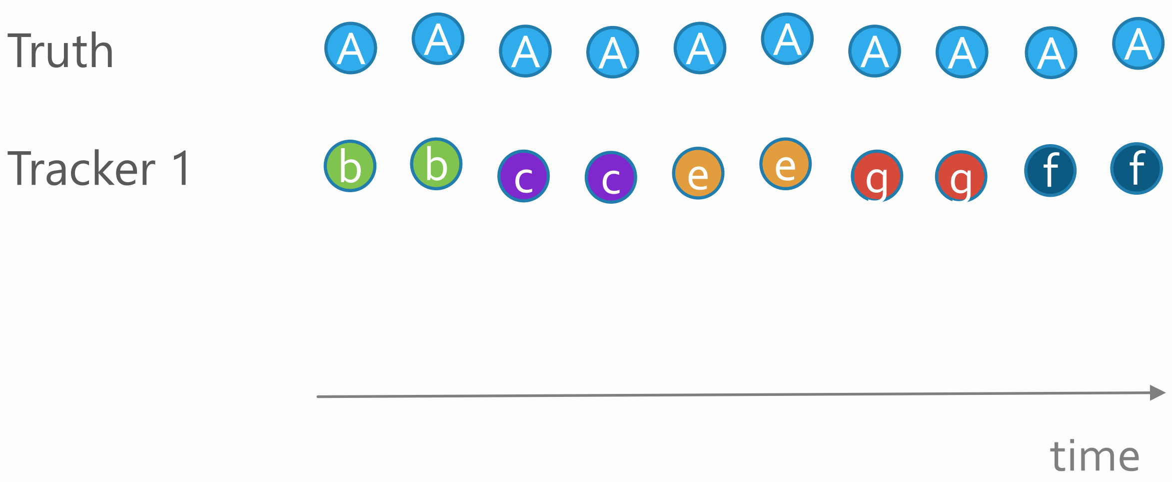


Trajectory Scores



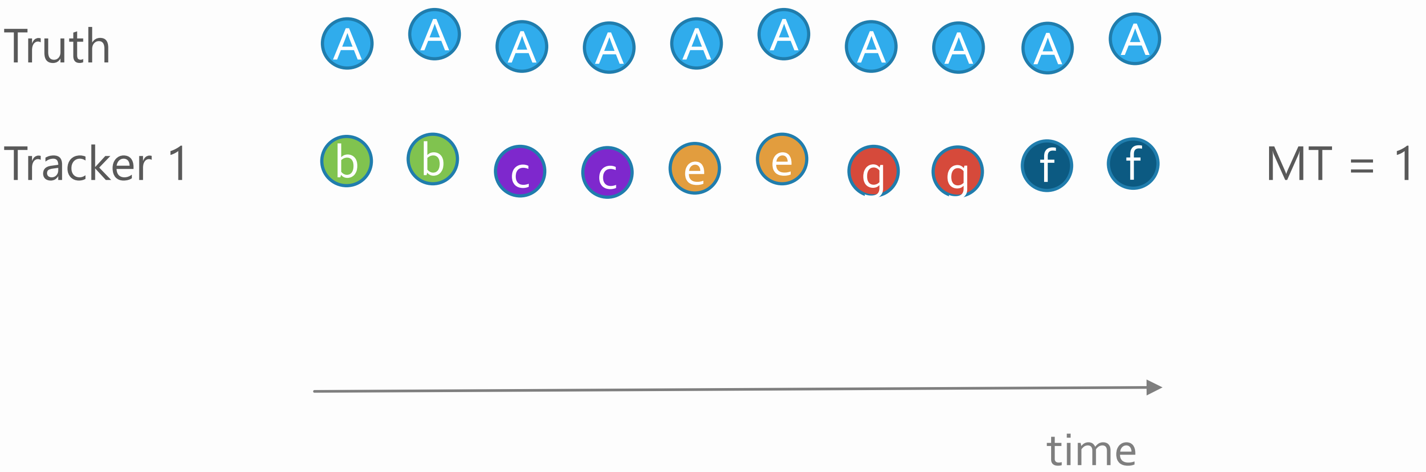
Mostly Tracked: GT trajectories which are covered by tracker output for more than 80% in length

Trajectory Scores



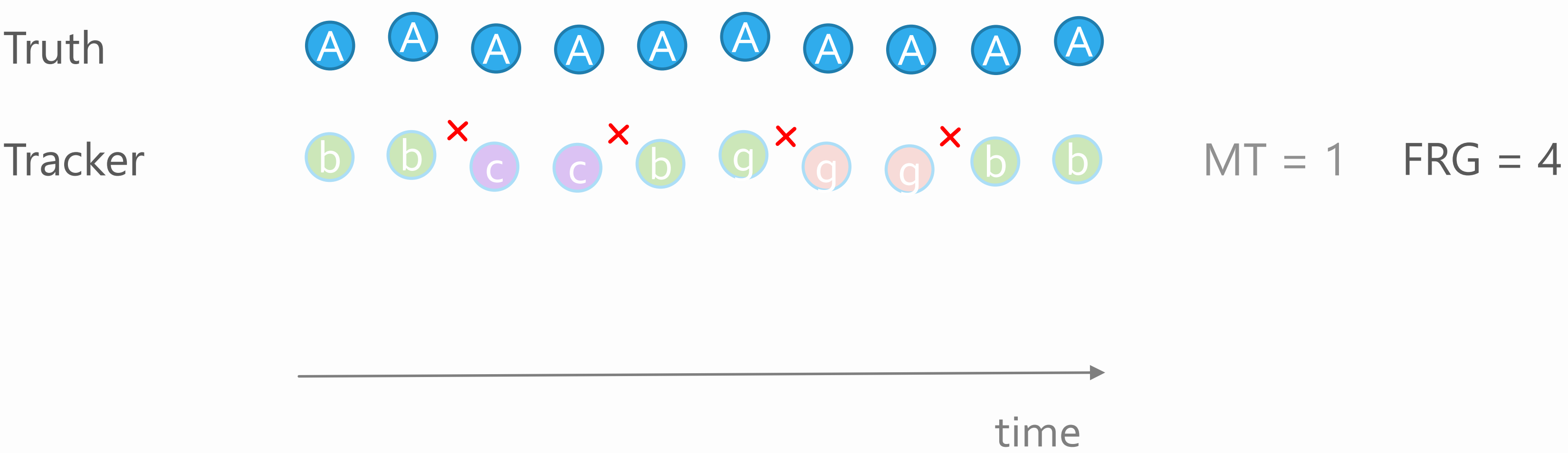
Mostly Tracked: GT trajectories which are covered by tracker output for more than 80% in length

Trajectory Scores



Mostly Tracked: GT trajectories which are covered by tracker output for more than 80% in length

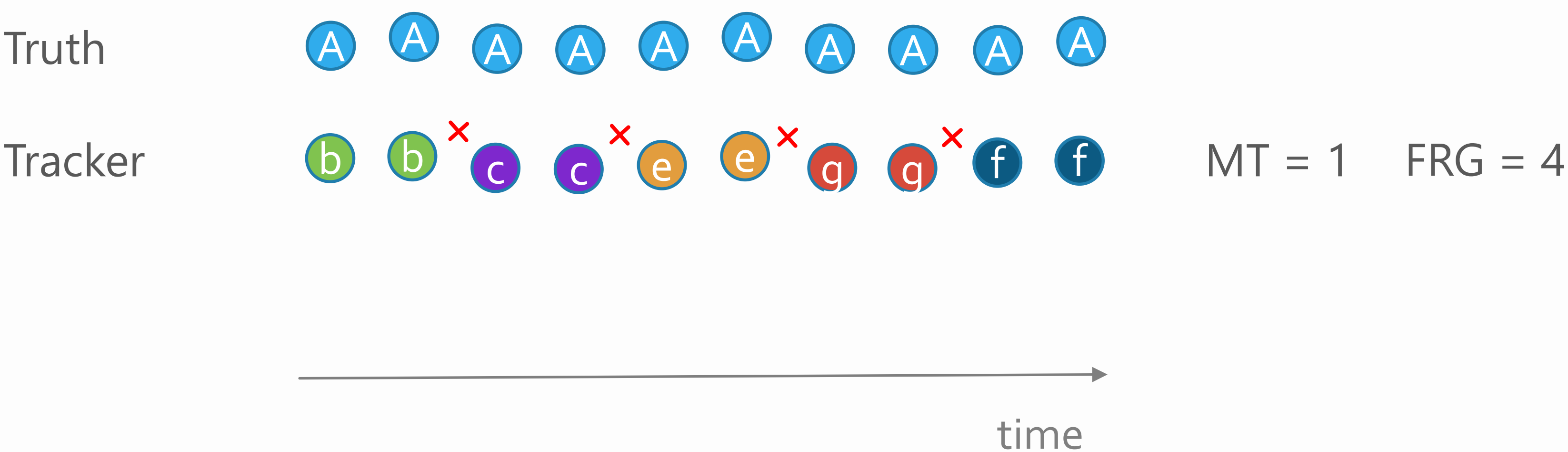
Trajectory Scores



Mostly Tracked: Percentage of GT trajectories which are covered by tracker output for more than 80% in length

Fragments: The number of times that a GT trajectory is interrupted in tracking result

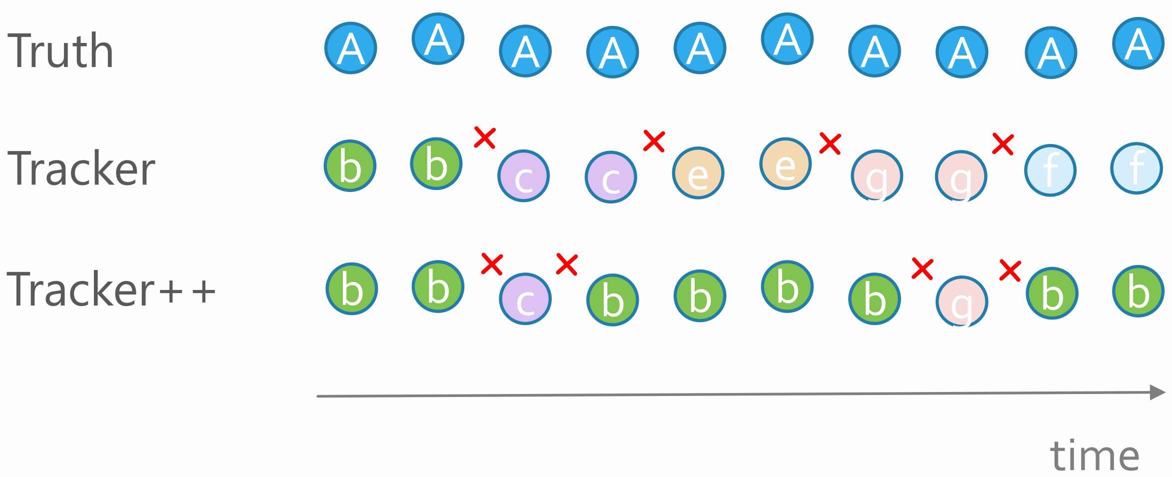
Trajectory Scores



Mostly Tracked: Percentage of GT trajectories which are covered by tracker output for more than 80% in length

Fragments: The number of times that a GT trajectory is interrupted in tracking result

Trajectory Scores

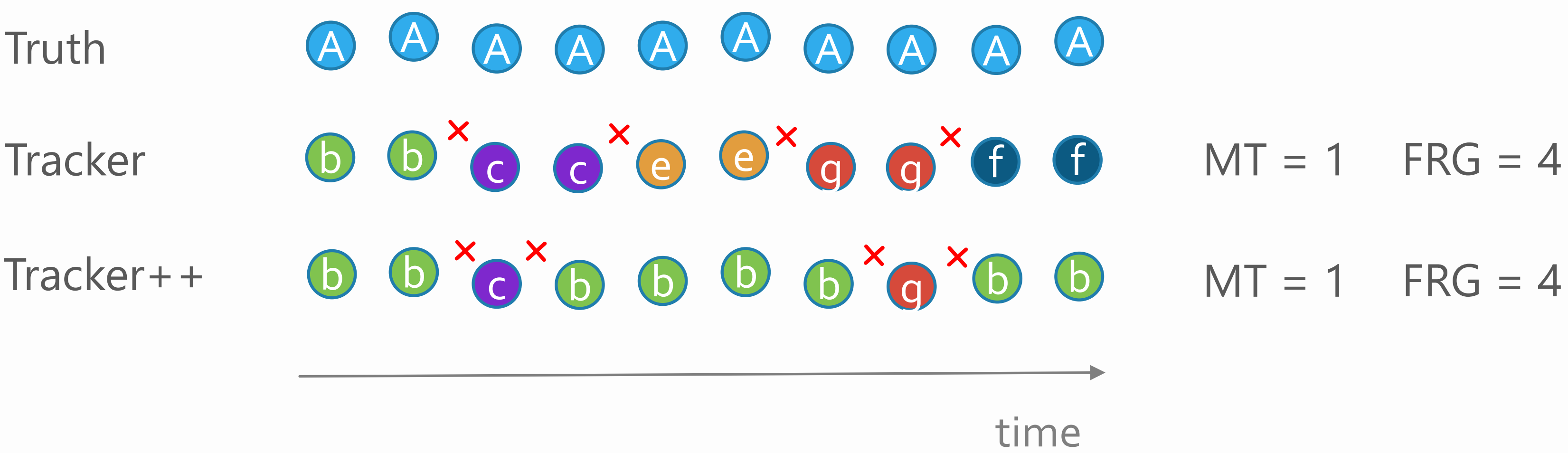


MT = 1   FRG = 4

Mostly Tracked: Percentage of GT trajectories which are covered by tracker output for more than 80% in length

Fragments: The number of times that a GT trajectory is interrupted in tracking result

Trajectory Scores



Mostly Tracked: Percentage of GT trajectories which are covered by tracker output for more than 80% in length

Fragments: The number of times that a GT trajectory is interrupted in tracking result

Scoring Functions

- Multiple Object Tracking Accuracy (MOTA)

$$\text{MOTA} = 1 - \frac{FN + FP + \Phi + M}{T} \qquad [-\infty, 1]$$

- Multiple Camera Tracking Accuracy (MCTA)

$$\text{MCTA} = \underbrace{\frac{2PR}{P+R}}_{F_1} \underbrace{\left(1 - \frac{\Phi^w + M^w}{T^w}\right)}_{\text{within camera}} \underbrace{\left(1 - \frac{\Phi^h + M^h}{T^h}\right)}_{\text{handover}} \qquad [0, 1]$$

Bizarre combination

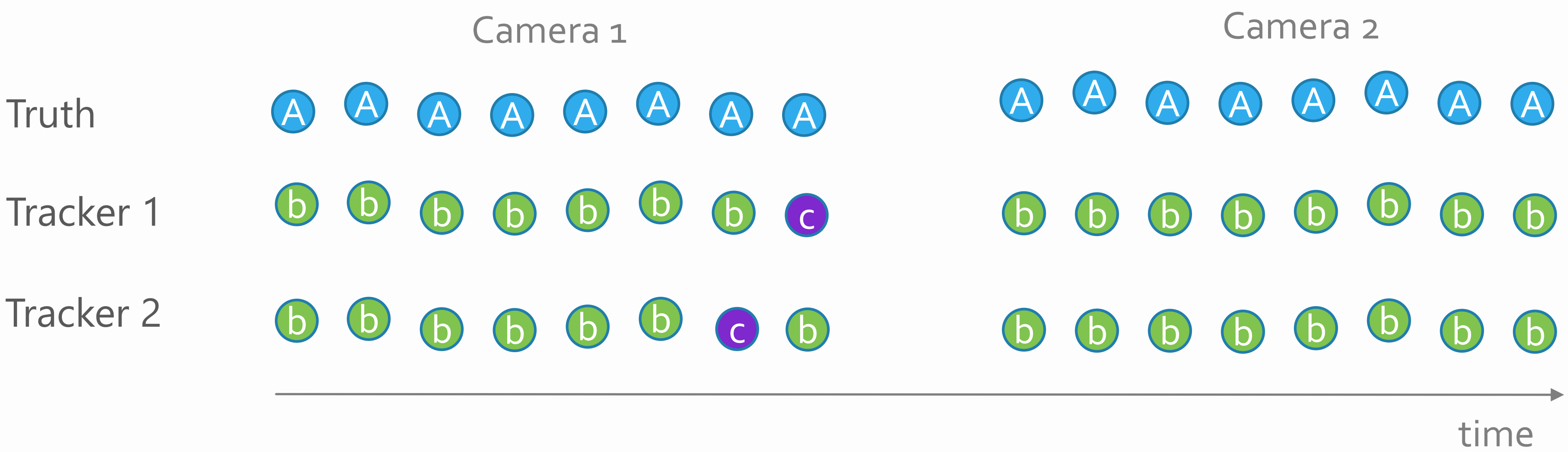
- Trajectory Scores

MT, PT, ML $\qquad \mathbb{Z}^*$

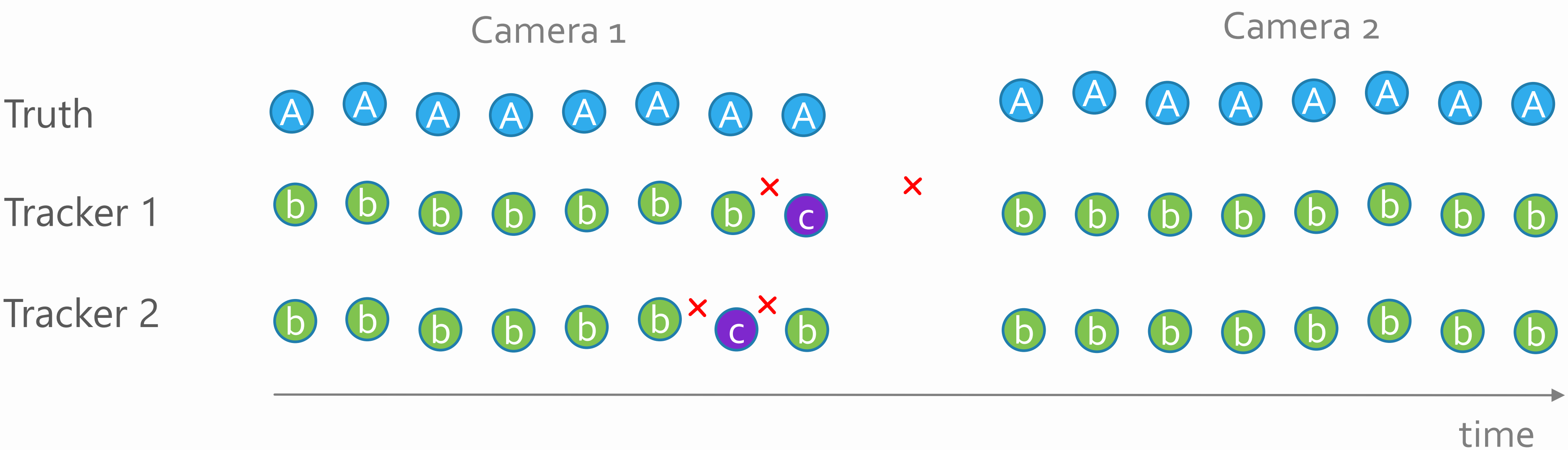
- Handover Errors

$\Phi^h, M^h \qquad \mathbb{Z}^*$

MCTA


Camera 1
Camera 2
Truth
Tracker 1
Tracker 2
A A A A A A A A A A A A A A A A
b b b b b b b c b b b b b b b b
b b b b b b c b b b b b b b b b
time

MCTA



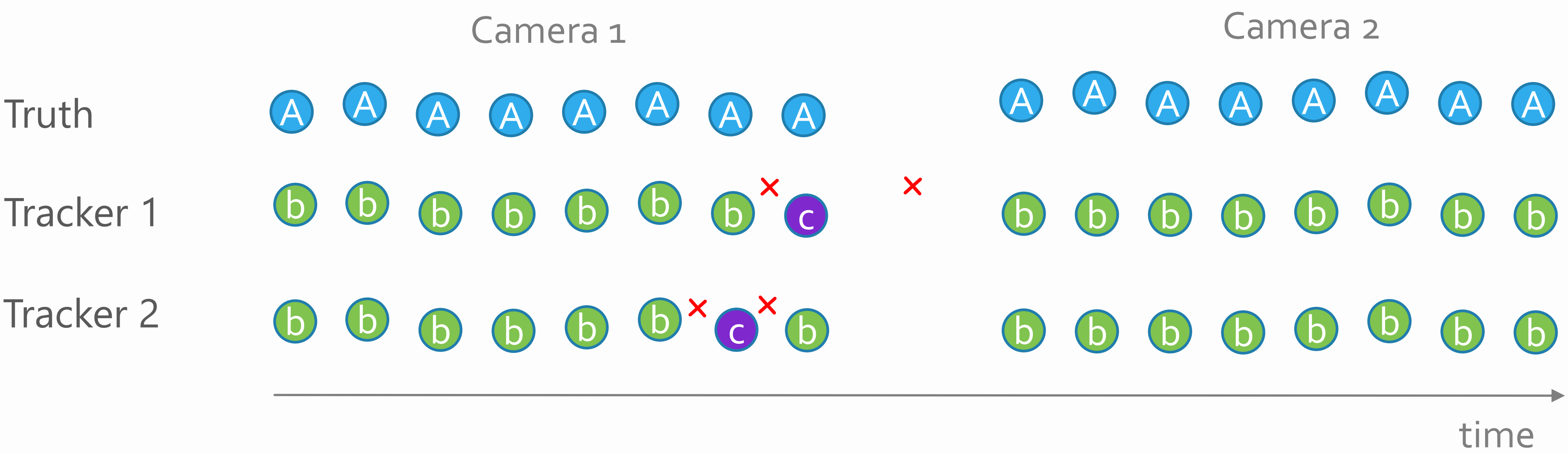
Fragmentations

Handover | Within camera

Tracker 1 $\Phi^h = 1, \Phi^w = 1$

Tracker 2 $\Phi^h = 0, \Phi^w = 2$

MCTA



Fragmentations

Handover | Within camera

Tracker 1 $\Phi^h = 1, \Phi^w = 1$

$$\mathrm{MCTA} = 1 \left(1 - \frac{1}{16}\right)\left(1 - \frac{1}{2}\right) = .46875$$

Tracker 2 $\Phi^h = 0, \Phi^w = 2$

$$\mathrm{MCTA} = \underbrace{1}_{F_1} \underbrace{\left(1 - \frac{2}{16}\right)}_{\text{within camera}} \underbrace{\left(1 - \frac{0}{2}\right)}_{\text{handover}} = .875$$

Scoring Functions

- Multiple Object Tracking Accuracy (MOTA)

$$\mathrm{MOTA} = 1 - \frac{FN + FP + \Phi + M}{T}$$ Mapping not one-to-one $[-\infty, 1]$

- Multiple Camera Tracking Accuracy (MCTA)

$$\mathrm{MCTA} = \underbrace{\frac{2PR}{P+R}}_{F_1} \underbrace{\left(1 - \frac{\Phi^w + M^w}{T^w}\right)}_{\text{within camera}} \underbrace{\left(1 - \frac{\Phi^h + M^h}{T^h}\right)}_{\text{handover}}$$ $[0, 1]$

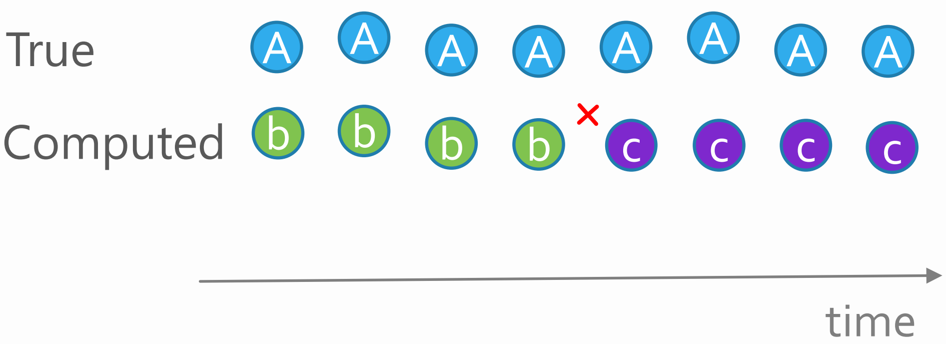
- Trajectory Scores

MT, PT, ML $\mathbb{Z}^*$

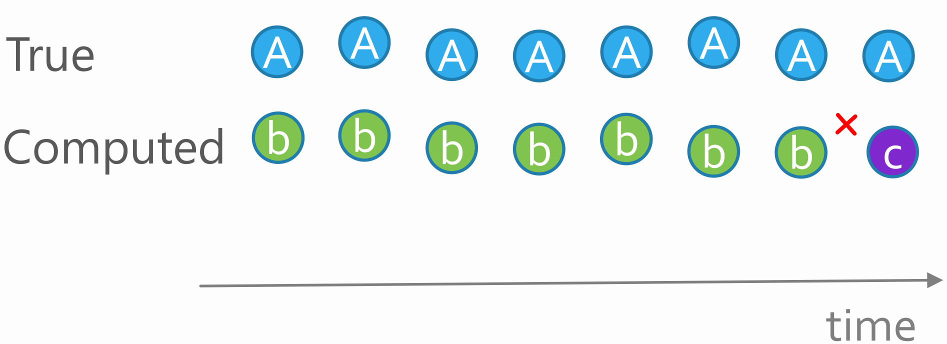
- Handover Errors

Φ^h, M^h $\mathbb{Z}^*$

MOTA

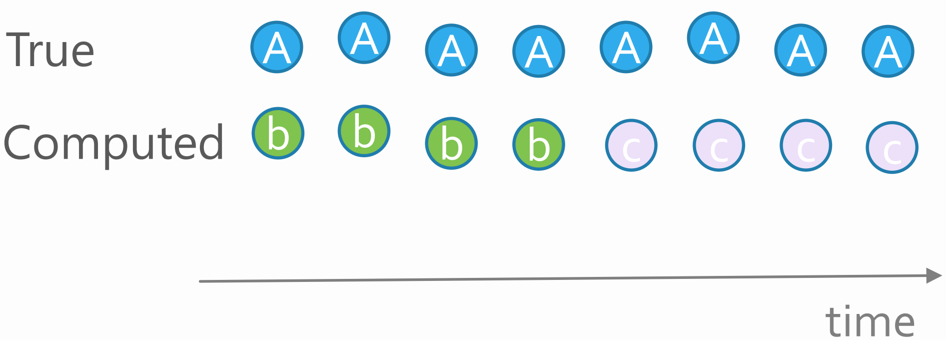


$$\mathrm{MOTA} \approx 1$$
$$\Phi = 1$$



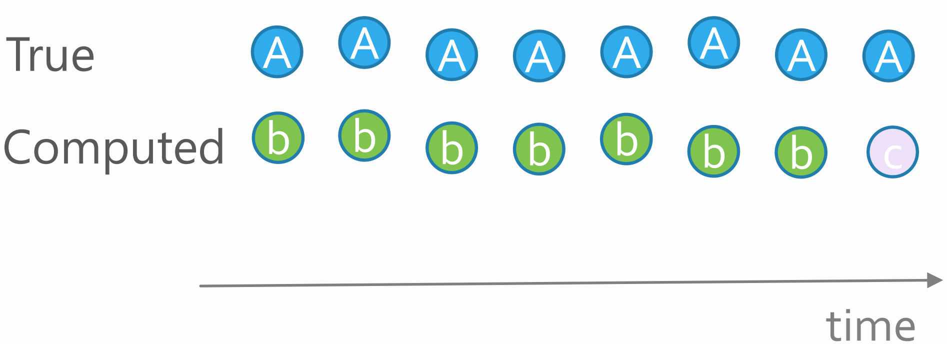
$$\mathrm{MOTA} \approx 1$$
$$\Phi = 1$$

MOTA



$$\mathrm{MOTA} \approx 1$$
$$\Phi = 1$$

Half explanation



$$\mathrm{MOTA} \approx 1$$
$$\Phi = 1$$

Nearly full explanation

Issues Summary

- Multiple Object Tracking Accuracy (MOTA)

$$\mathrm{MOTA} = 1 - \frac{FN + FP + \Phi + M}{T}$$

Mapping not one-to-one

$[-\infty, 1]$

- Multiple Camera Tracking Accuracy (MCTA)

$$\mathrm{MCTA} = \underbrace{\frac{2PR}{P+R}}_{F_1} \underbrace{\left(1 - \frac{\Phi^w + M^w}{T^w}\right)}_{\text{within camera}} \underbrace{\left(1 - \frac{\Phi^h + M^h}{T^h}\right)}_{\text{handover}}$$

Bizarre combination

$[0, 1]$

- Trajectory Scores

MT, PT, ML

Ignore identity

$\mathbb{Z}^*$

- Handover Errors

Φ^h, M^h

Brittle due to mapping

$\mathbb{Z}^*$

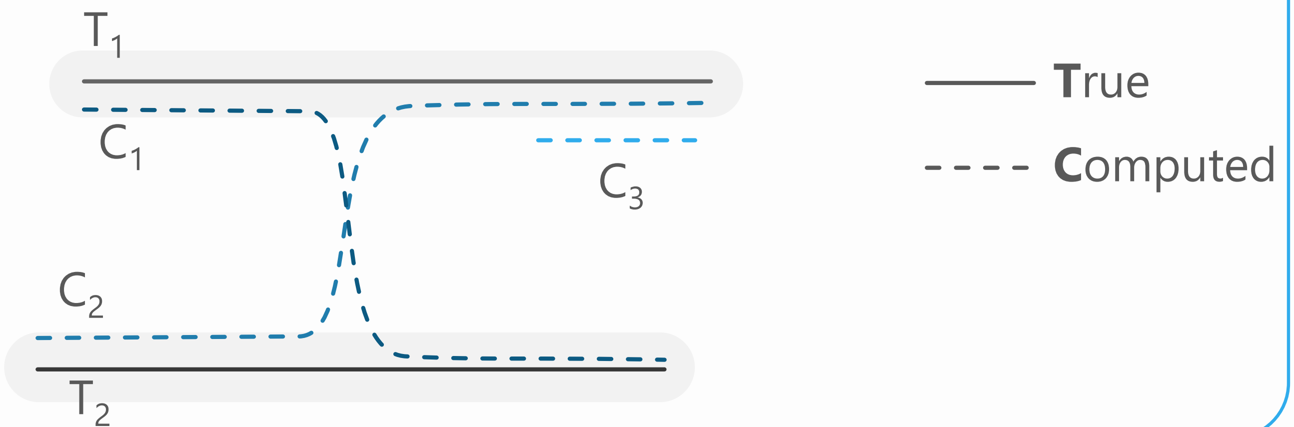
- All scores build on top of the CLEAR MOT mapping (not bijective globally)

Identity Measures

Identity Measures

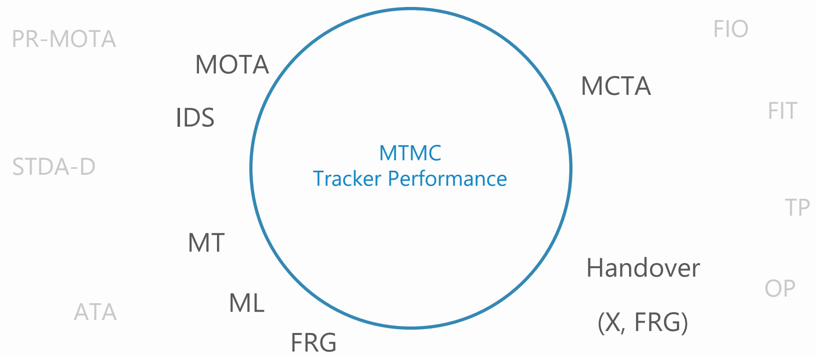
- Truth-to-Result Matching

$$\mu : T \rightarrow C$$



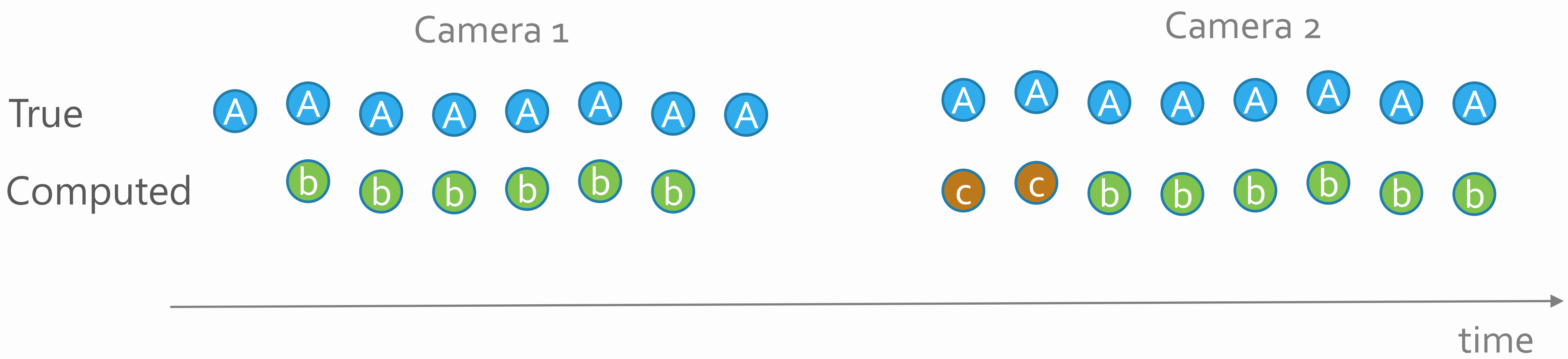
- Scoring Function

$$f(T, C, \mu)$$



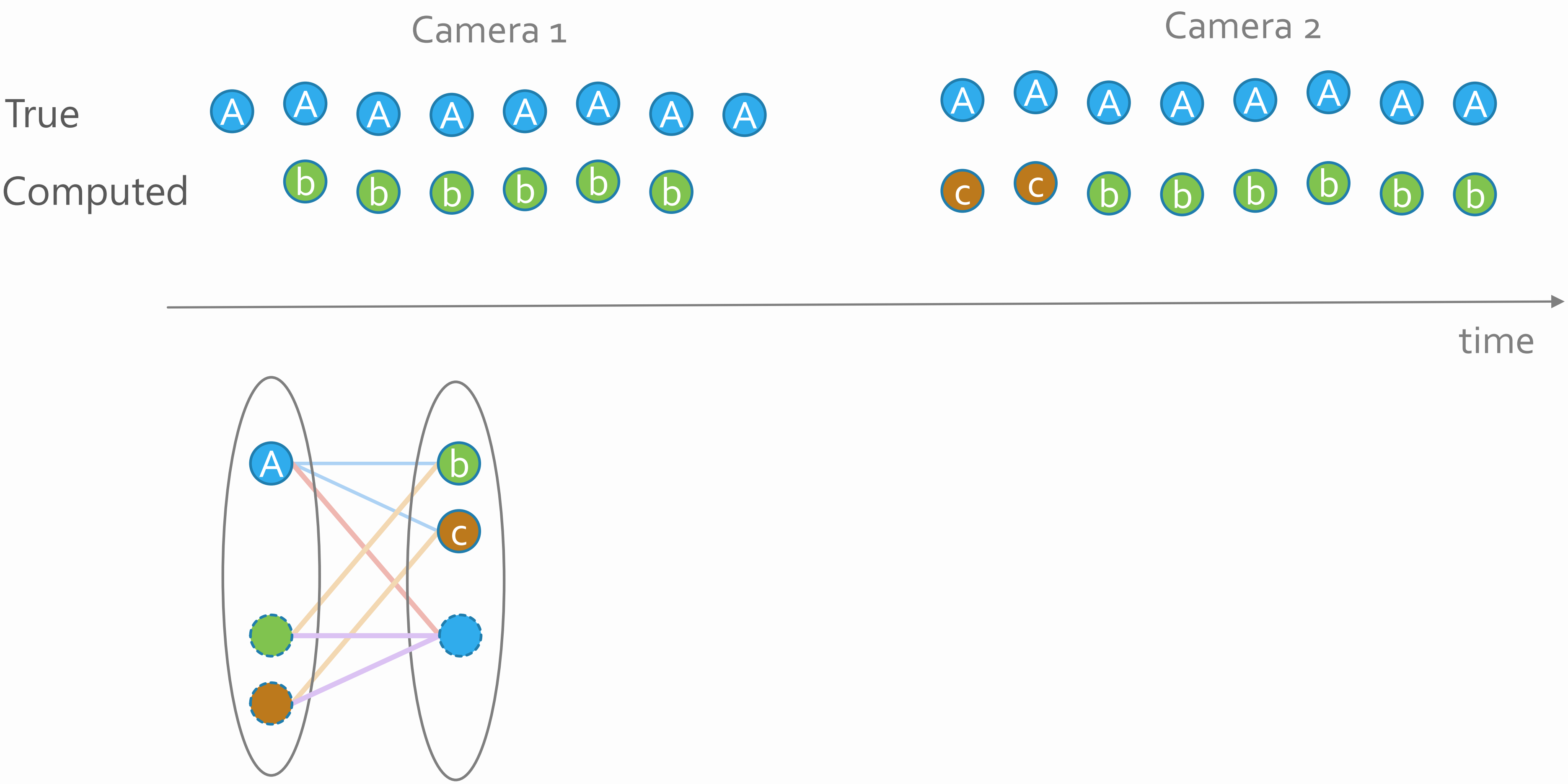
Identity Measures

- Proposed Truth-to-Result Matching



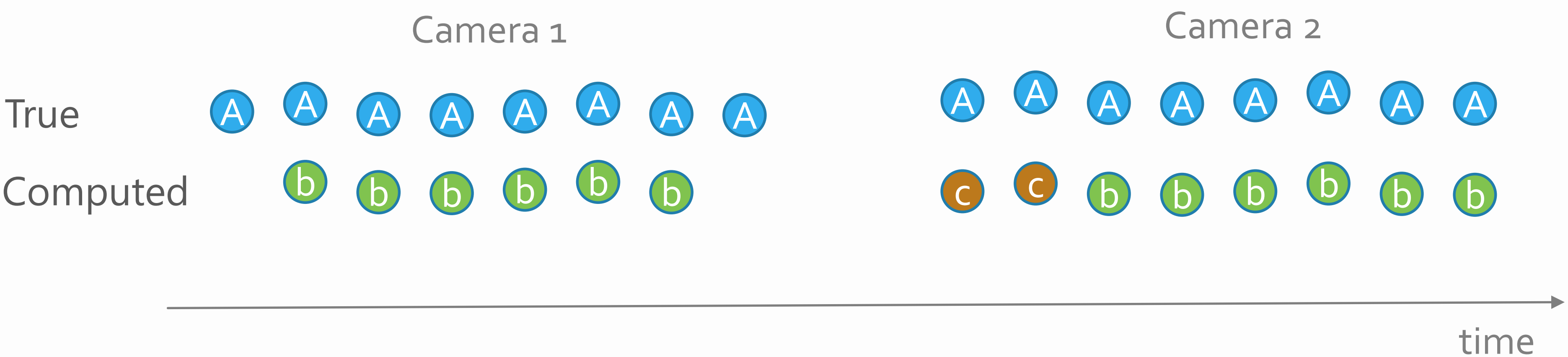
Identity Measures

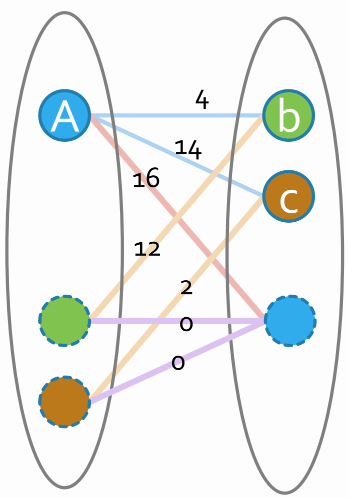
- Proposed Truth-to-Result Matching



Identity Measures

- Proposed Truth-to-Result Matching

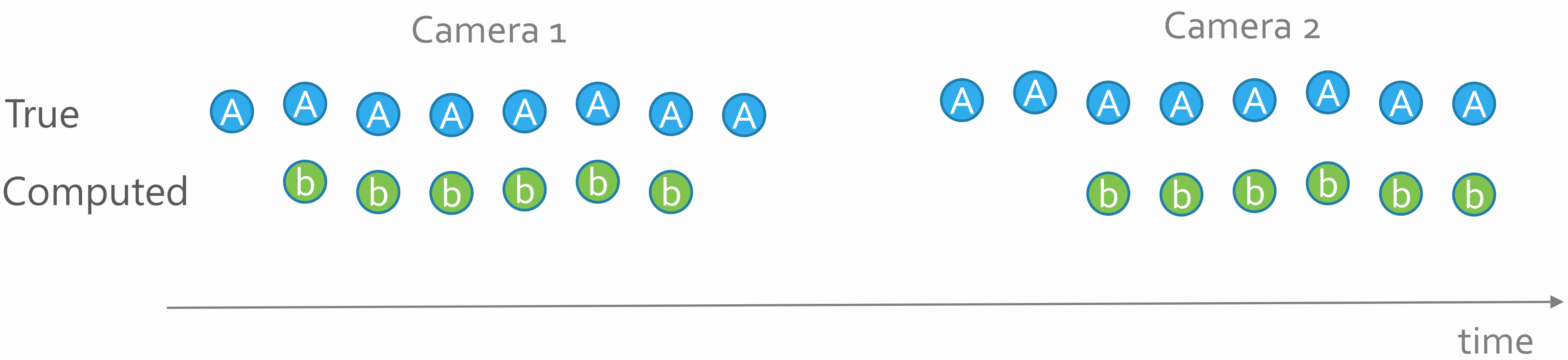


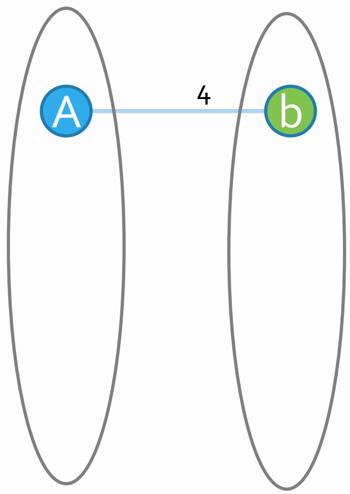


Cost on each edge is number of mis-assigned frames between two trajectories

Identity Measures

- Proposed Truth-to-Result Matching

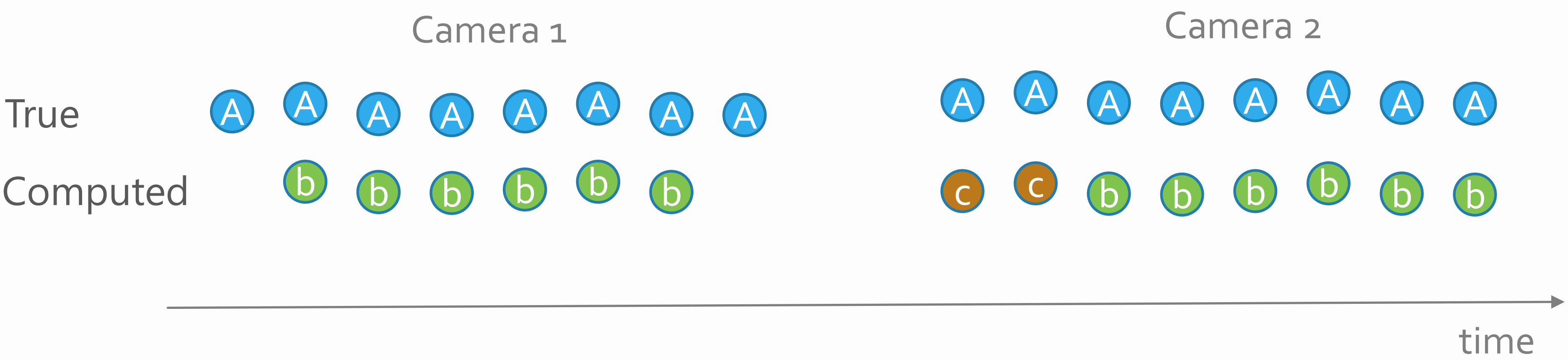


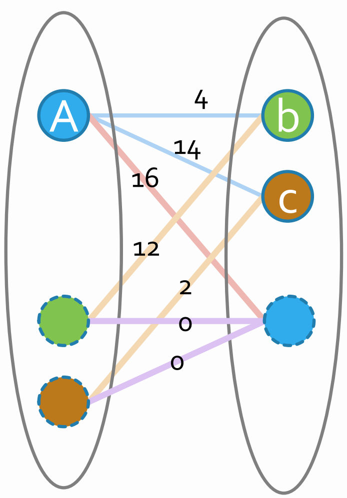


Cost on each edge is number of mis-assigned frames between two trajectories

Identity Measures

- Proposed Truth-to-Result Matching

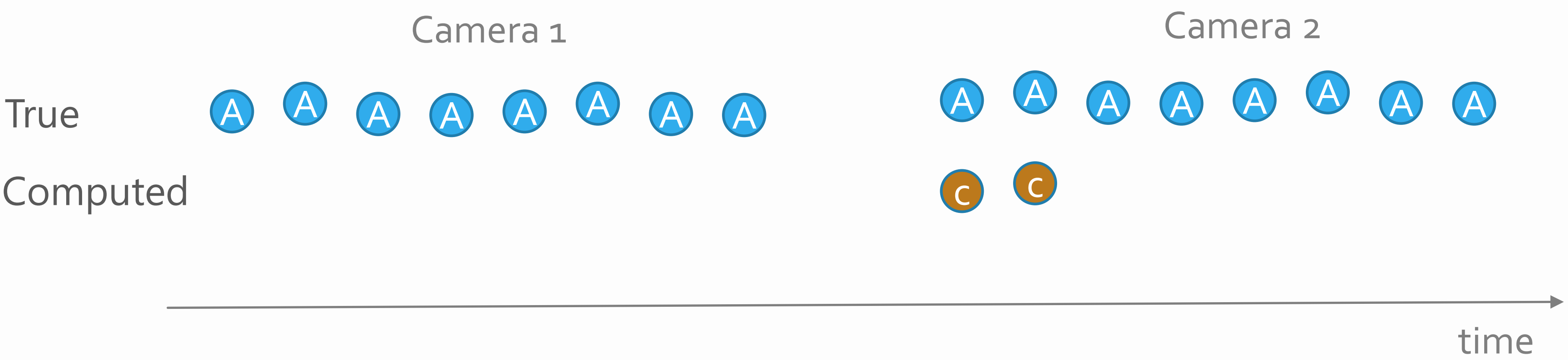


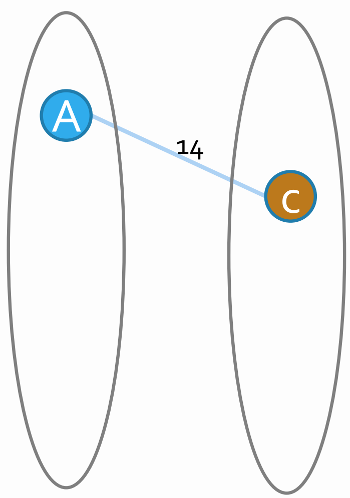


Cost on each edge is number of mis-assigned frames between two trajectories

Identity Measures

- Proposed Truth-to-Result Matching

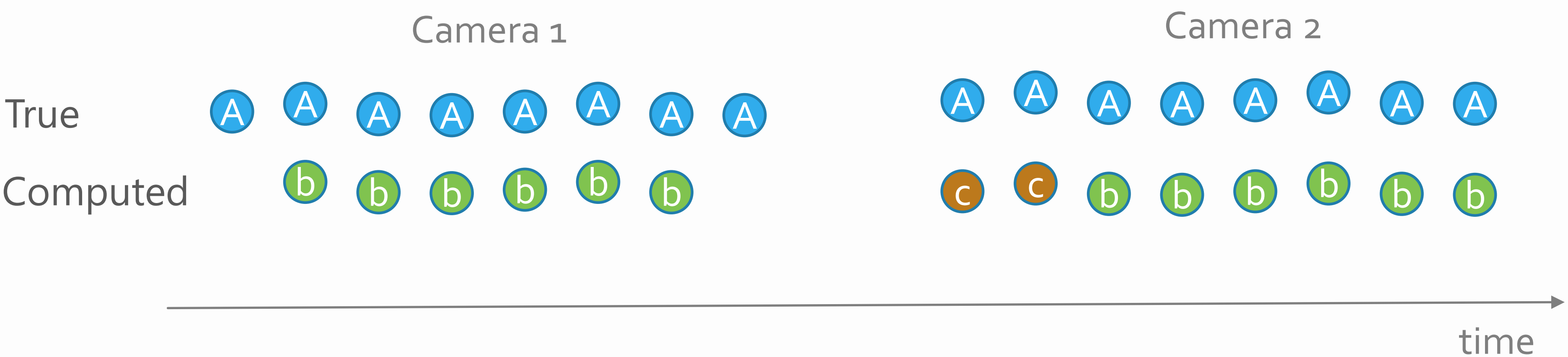


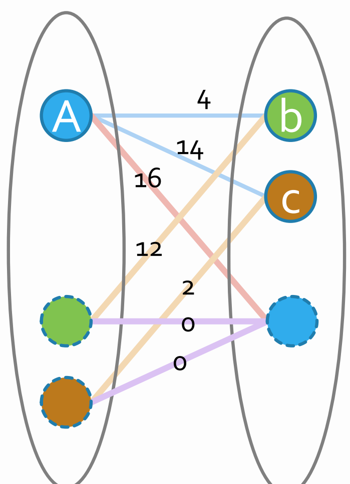


Cost on each edge is number of mis-assigned frames between two trajectories

Identity Measures

- Proposed Truth-to-Result Matching

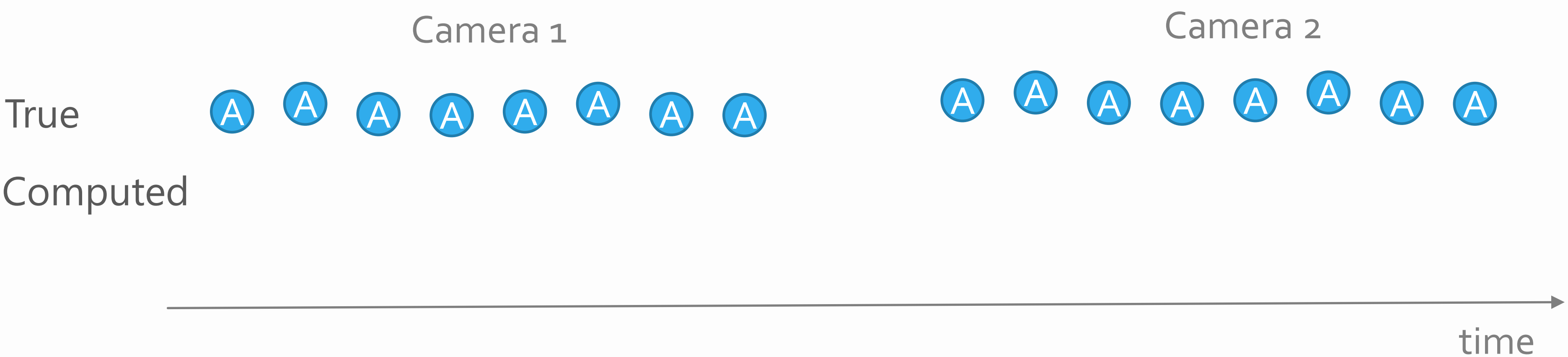


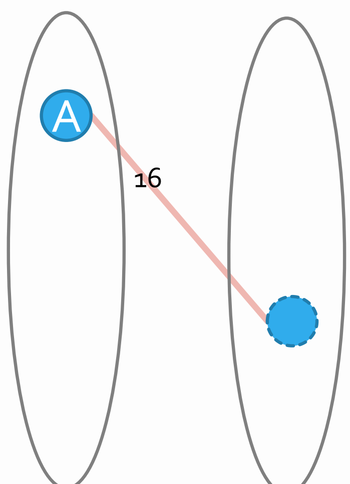


Cost on each edge is number of mis-assigned frames between two trajectories

Identity Measures

- Proposed Truth-to-Result Matching

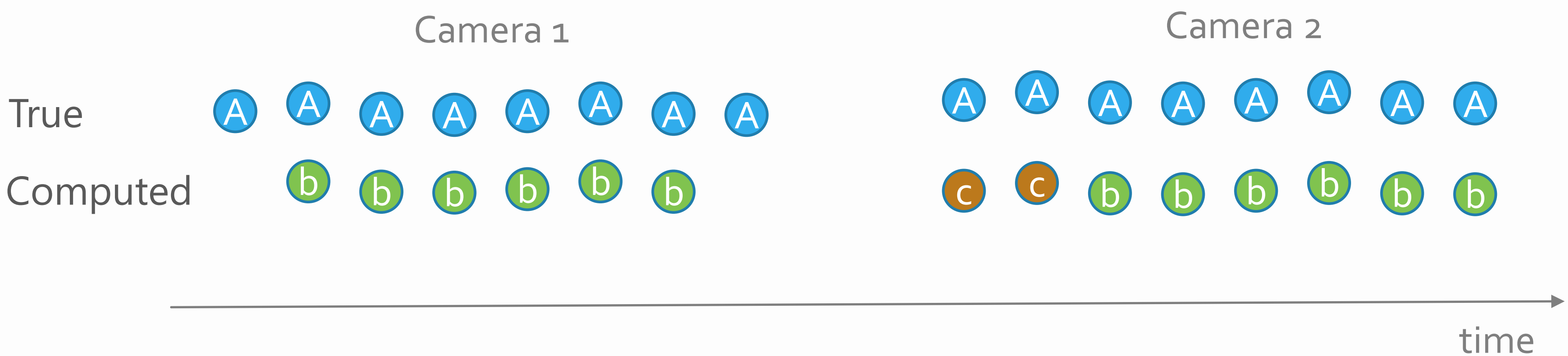


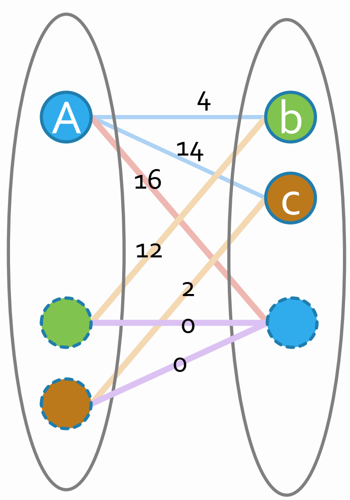


Cost on each edge is number of mis-assigned frames between two trajectories

Identity Measures

- Proposed Truth-to-Result Matching

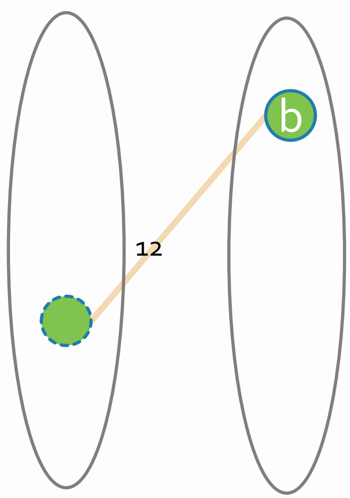




Cost on each edge is number of mis-assigned frames between two trajectories

Identity Measures

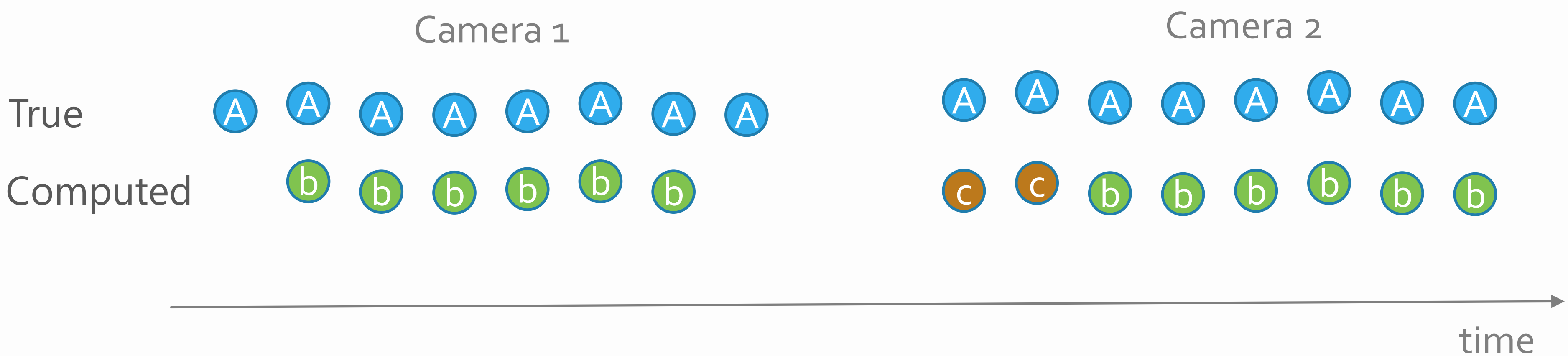
- Proposed Truth-to-Result Matching

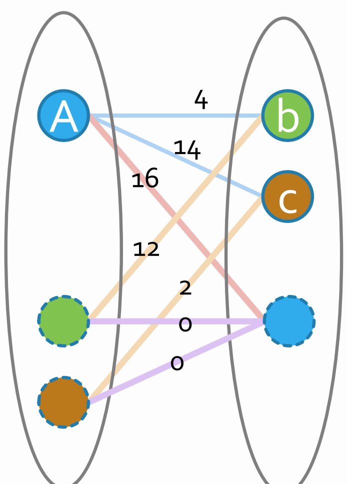


Cost on each edge is number of mis-assigned frames between two trajectories

Identity Measures

- Proposed Truth-to-Result Matching

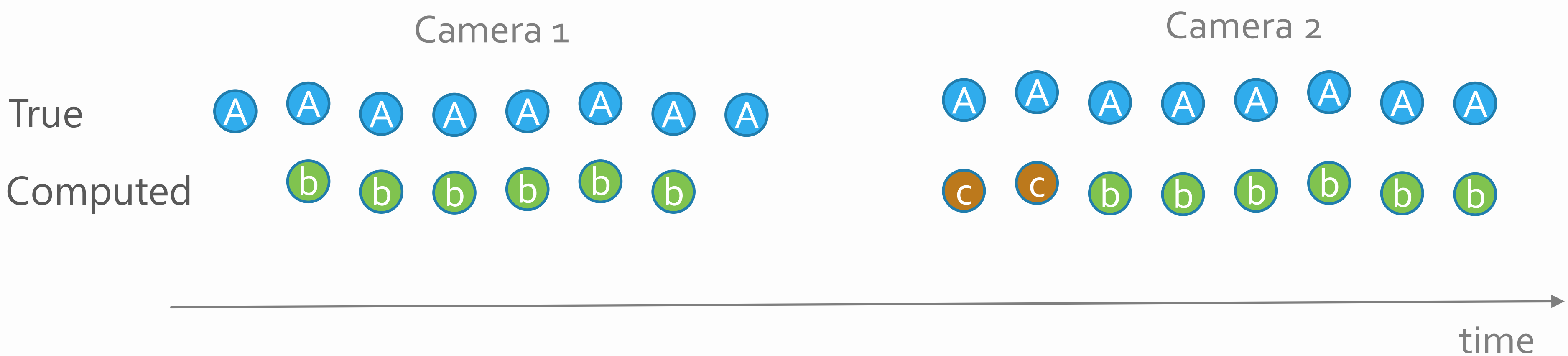


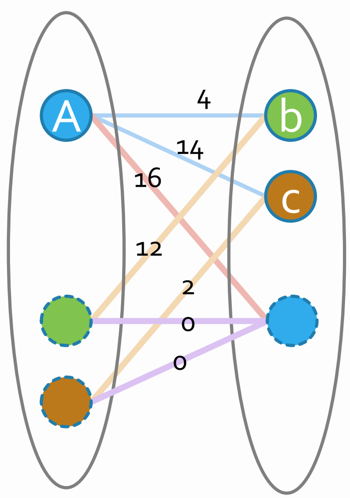


Cost on each edge is number of mis-assigned frames between two trajectories

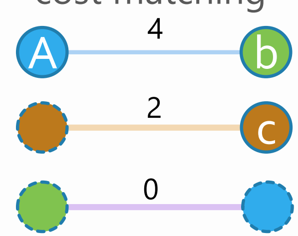
Identity Measures

- Proposed Truth-to-Result Matching





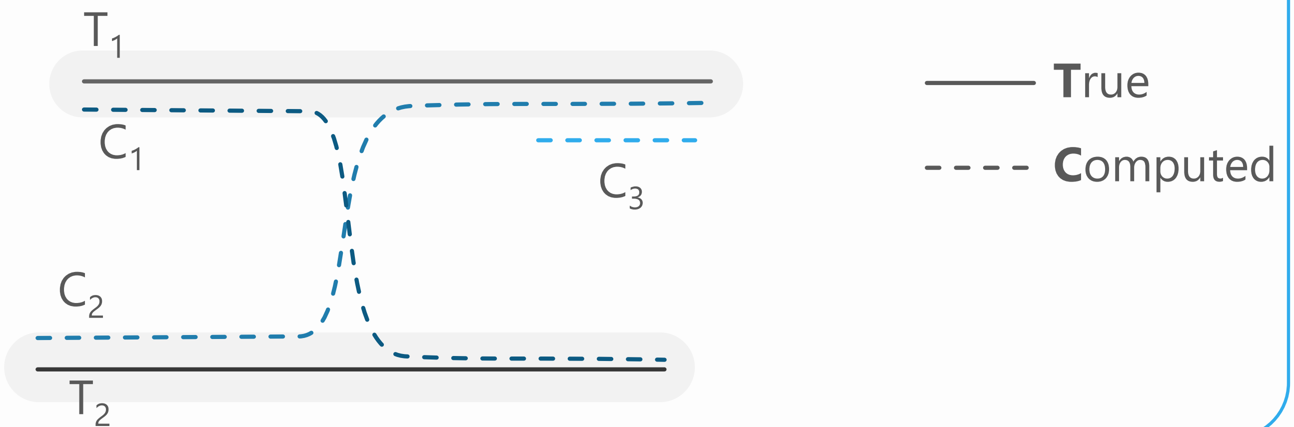
ID correspondence determined through min cost matching



Identity Measures

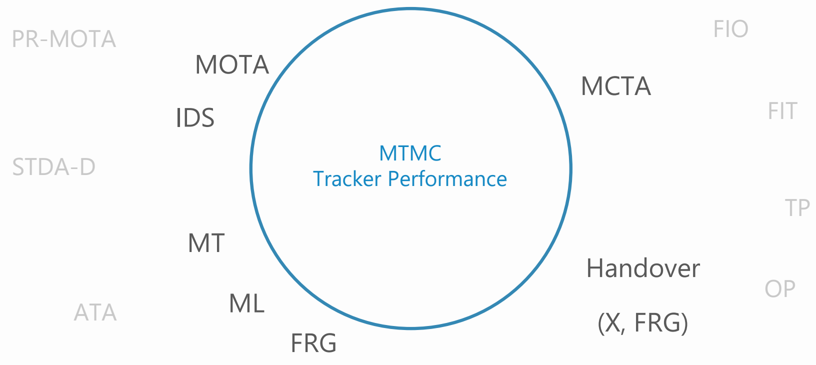
- Truth-to-Result Matching

$$\mu : T \to C$$



- Scoring Function

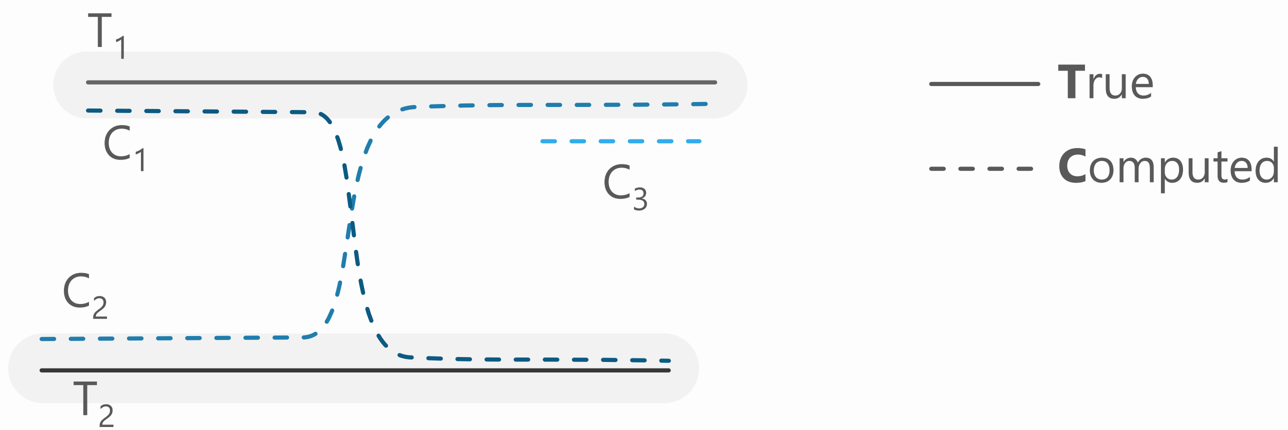
$$f(T, C, \mu)$$



Identity Measures

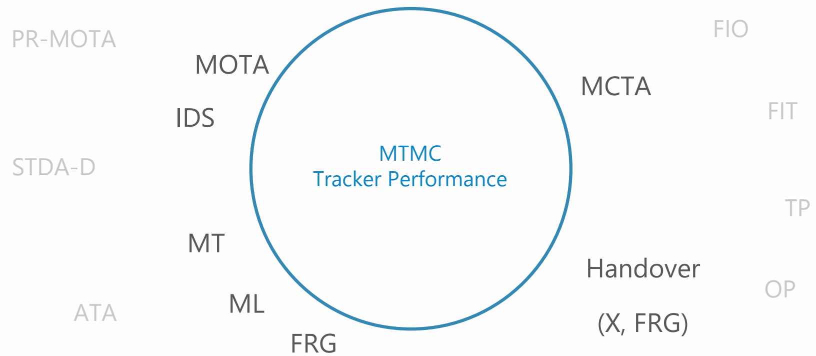
- Truth-to-Result Matching

$$\mu : T \to C$$



- Scoring Function

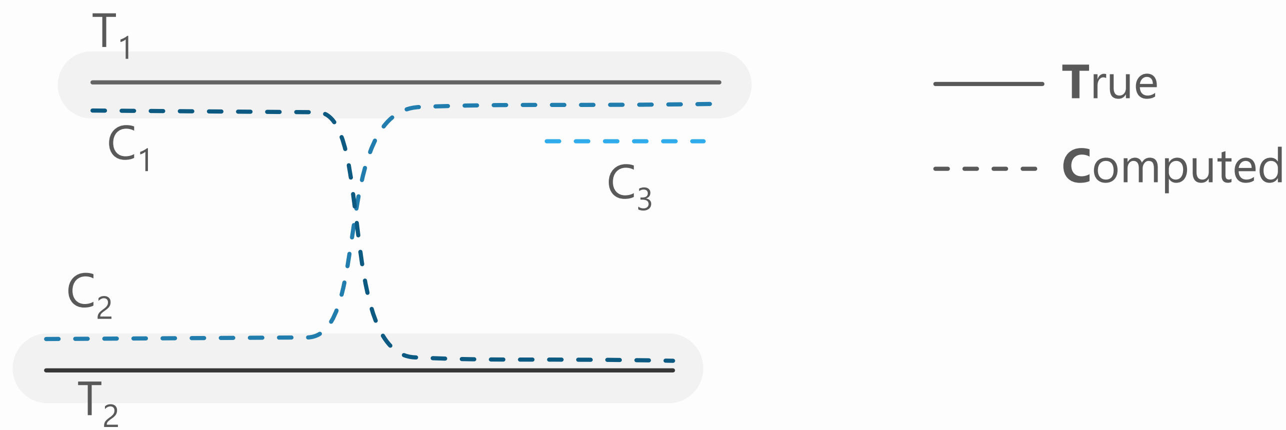
$$f(T, C, \mu)$$



Identity Measures

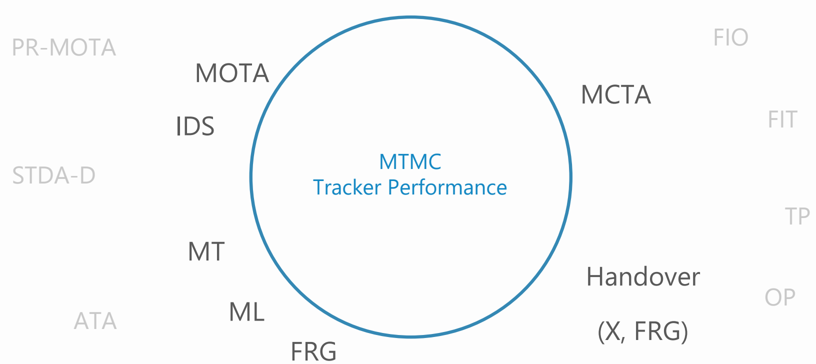
- Truth-to-Result Matching

$$\mu : T \to C$$



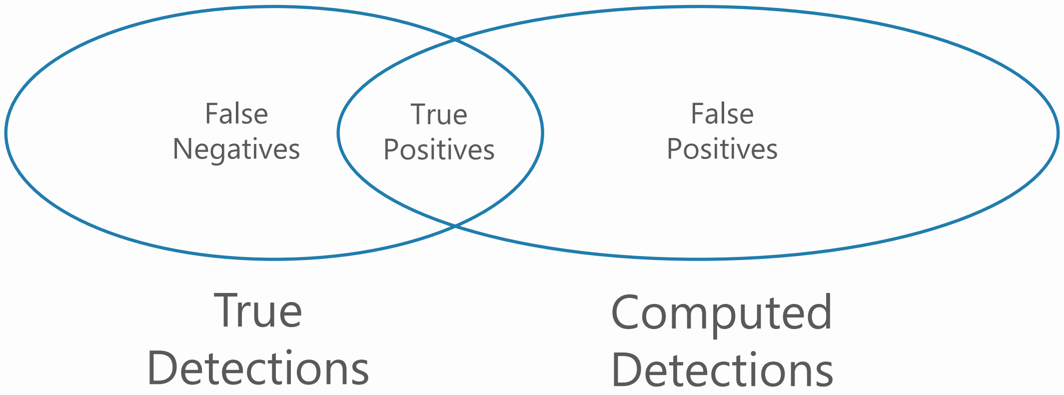
- Scoring Function

$$f(T, C, \mu)$$



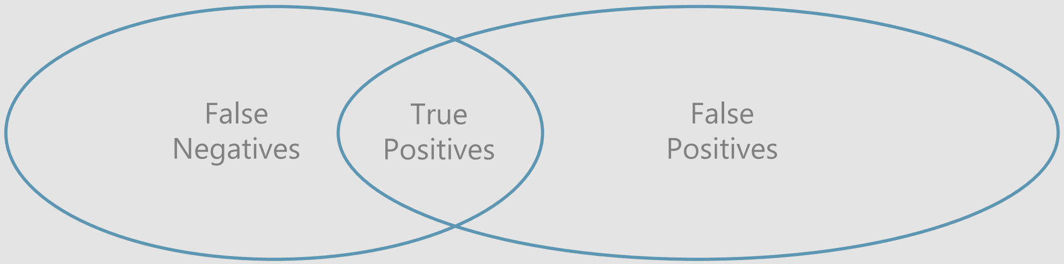
Identity Measures

- Proposed Scores
  - ID Precision $P = \frac{TP}{TP+FP} = \frac{TP}{C}$ $[0, 1]$
  - ID Recall $R = \frac{TP}{TP+FN} = \frac{TP}{T}$ $[0, 1]$
  - F_1-score $F_1 = 2\frac{PR}{P+R} = \frac{TP}{\frac{T+C}{2}}$ $[0, 1]$



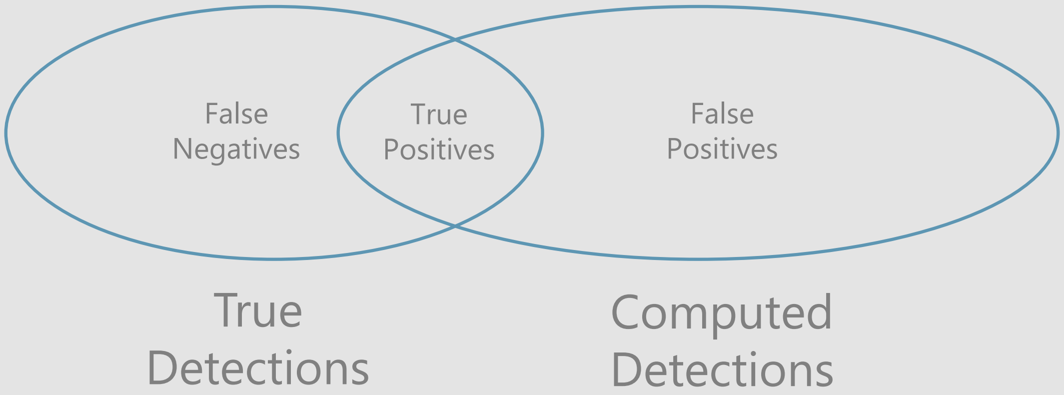
Identity Measures

- Proposed Scores
  - IDP: Fraction of computed detections that are correctly identified.
  - ID Recall $R = \frac{TP}{TP+FN} = \frac{TP}{T}$ $[0, 1]$
  - F_1-score $F_1 = 2\frac{PR}{P+R} = \frac{TP}{\frac{T+C}{2}}$ $[0, 1]$



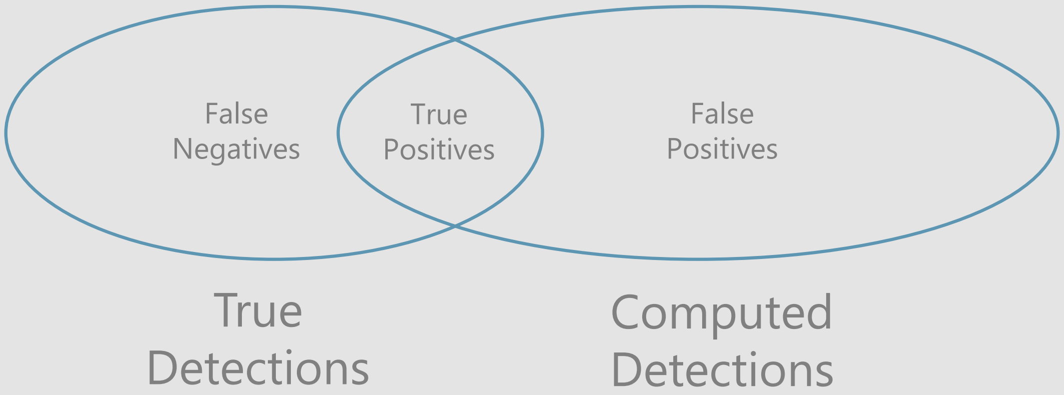
Identity Measures

- Proposed Scores
  - IDP: Fraction of computed detections that are correctly identified.
  - IDR: Fraction of ground-truth detections that are correctly identified.
  - F_1-score $\quad F_1 = 2\frac{PR}{P+R} = \frac{TP}{\frac{T+C}{2}} \quad [0, 1]$



Identity Measures

- Proposed Scores
  - IDP: Fraction of computed detections that are correctly identified.
  - IDR: Fraction of ground-truth detections that are correctly identified.
  - IDF_1: Ratio of correctly identified detections over the average number of ground-truth and computed detections.

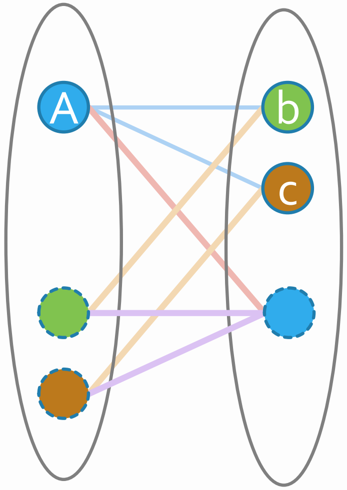


Identity Measures

- Properties
  - True and computed identities are mapped 1-1
  - Notion of identity switch is not present in ID measures
  - The truth-to-result matching is the most favorable to the algorithm
  - Applicable to single- and multi-camera tracking

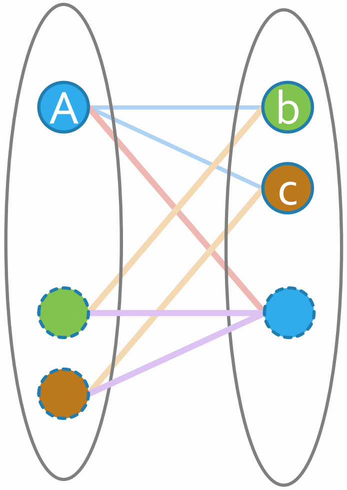
Identity Measures

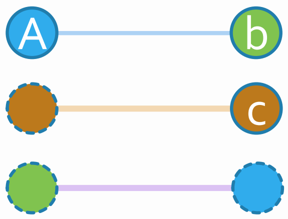
- Properties
  - True and computed identities are mapped 1-1
  - Notion of identity switch is not present in ID measures
  - The truth-to-result matching is the most favorable to the algorithm
  - Applicable to single- and multi-camera tracking



Identity Measures

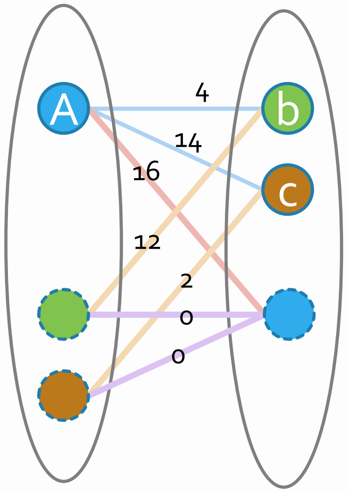
- Properties
  - True and computed identities are mapped 1-1
  - Notion of identity switch is not present in ID measures
  - The truth-to-result matching is the most favorable to the algorithm
  - Applicable to single- and multi-camera tracking

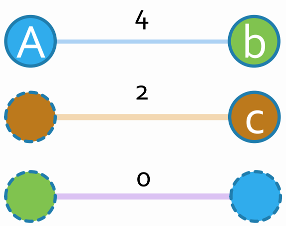




Identity Measures

- Properties
  - True and computed identities are mapped 1-1
  - Notion of identity switch is not present in ID measures
  - The truth-to-result matching is the most favorable to the algorithm
  - Applicable to single- and multi-camera tracking

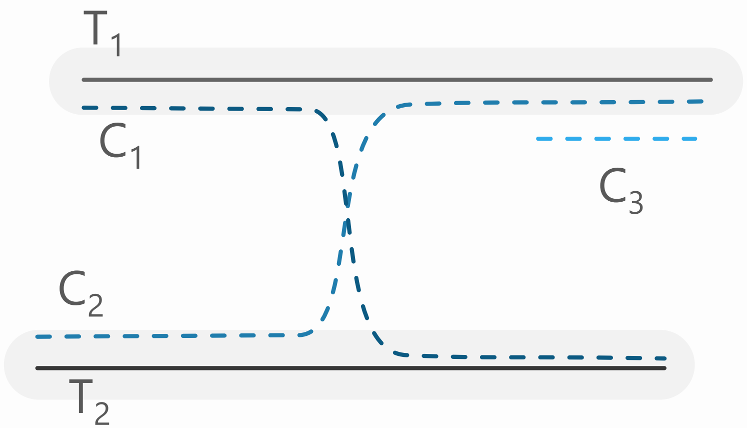




Min Cost Matching maximizes IDF_1

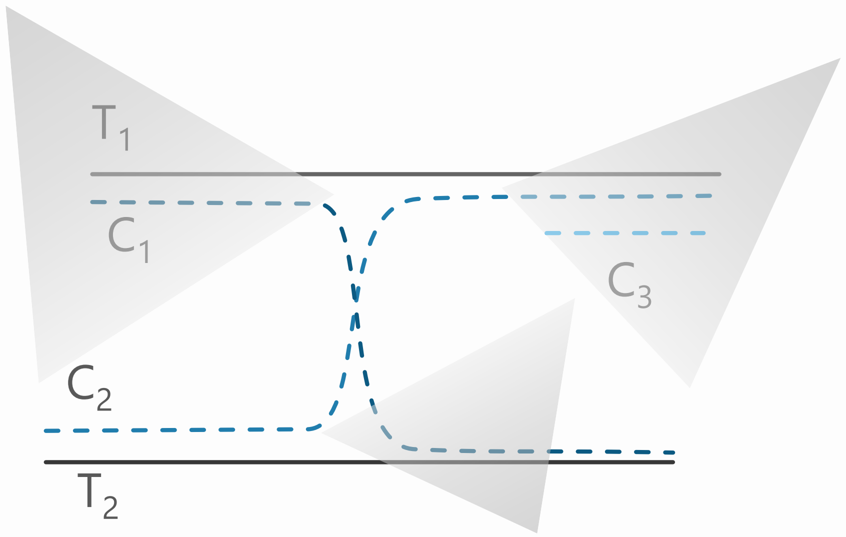
Identity Measures

- Properties
  - True and computed identities are mapped 1-1
  - Notion of identity switch is not present in ID measures
  - The truth-to-result matching is the most favorable to the algorithm
  - Applicable to single- and multi-camera tracking



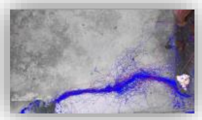
Identity Measures

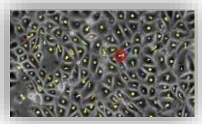
- Properties
  - True and computed identities are mapped 1-1
  - Notion of identity switch is not present in ID measures
  - The truth-to-result matching is the most favorable to the algorithm
  - Applicable to single- and multi-camera tracking



Scope and Limitations

- ID measures are useful in:
  - Surveillance
  - Sports
  - Retail
- ID measures not relevant when:
  - Tracking indistinguishable targets (ants, sheep)
- Not applicable to:
  - Tracking targets that merge/split (cells)





Practical Implications

Practical Implications

MOT 16 Challenge

Tracker	MOTA ↑
HCC	0.492
LMP	0.487
FWT	0.477
NLLMPa	0.475
MDPNN16	0.471
NOMT_16	0.464
JMC	0.462
QuadMOT16	0.441
oICF_16	0.432
MHT_DAM_16	0.429
LINF1_16	0.410
EAMTT_pub	0.388
OVBT	0.384
LTTSC-CRF	0.375
LP2D_16	0.357
TBD_16	0.337
CEM_16	0.331
DP_NMS_16	0.321
GMPHD_HDA	0.305
SMOT_16	0.297
JPDA_m_16	0.261

Practical Implications

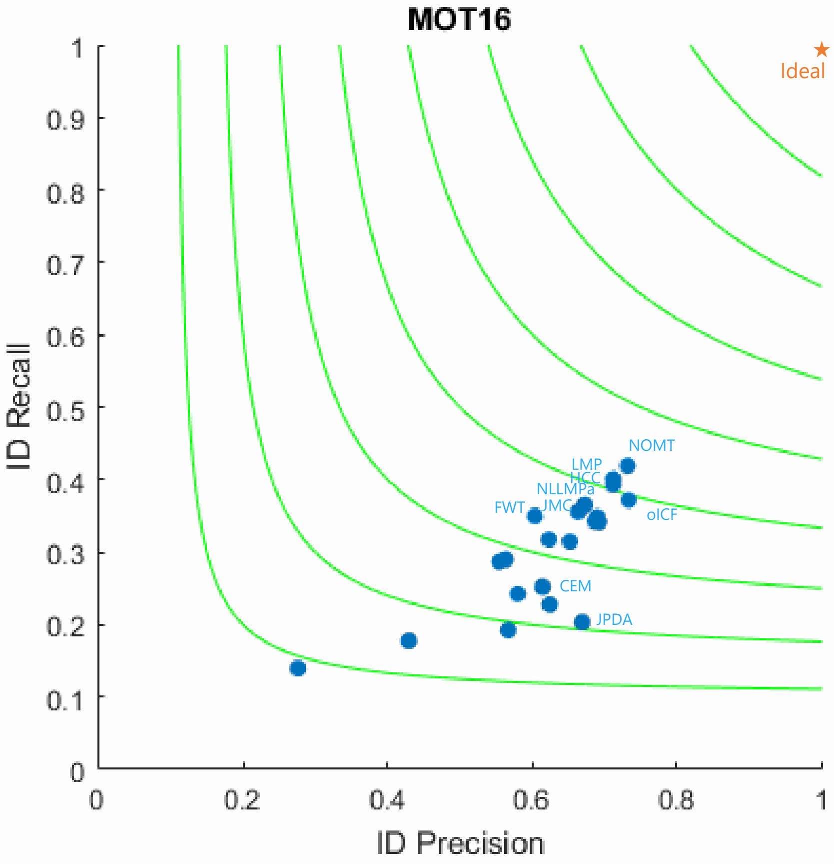
MOT 16 Challenge

Tracker	MOTA ↑
HCC	0.492
LMP	0.487
FWT	0.477
NLLMPa	0.475
MDPNN16	0.471
NOMT_16	0.464
JMC	0.462
QuadMOT16	0.441
oICF_16	0.432
MHT_DAM_16	0.429
LINF1_16	0.410
EAMTT_pub	0.388
OVBT	0.384
LTTSC-CRF	0.375
LP2D_16	0.357
TBD_16	0.337
CEM_16	0.331
DP_NMS_16	0.321
GMPHD_HDA	0.305
SMOT_16	0.297
JPDA_m_16	0.261

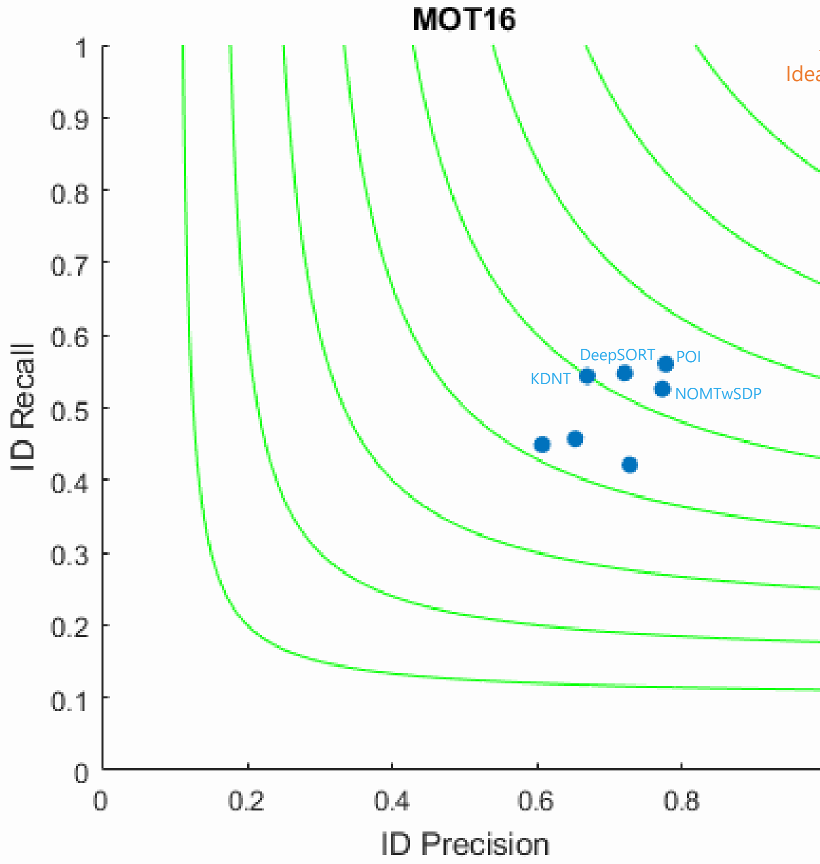
Tracker	IDF1 ↑	Rank Difference
NOMT_16	0.533	+5
LMP	0.512	0
HCC	0.506	-2
oICF_16	0.493	+5
NLLMPa	0.473	-1
JMC	0.463	+1
MDPNN16	0.462	-2
MHT_DAM_16	0.457	+2
LINF1_16	0.456	+2
FWT	0.442	-7
EAMTT_pub	0.424	+1
LTTSC-CRF	0.420	+2
QuadMOT16	0.382	-5
OVBT	0.378	-1
CEM_16	0.357	+2
LP2D_16	0.341	-1
GMPHD_HDA	0.333	+2
JPDA_m_16	0.311	+3
DP_NMS_16	0.287	-1
TBD_16	0.251	-4
SMOT_16	0.185	-1

ID P-R Curves

DPM detector

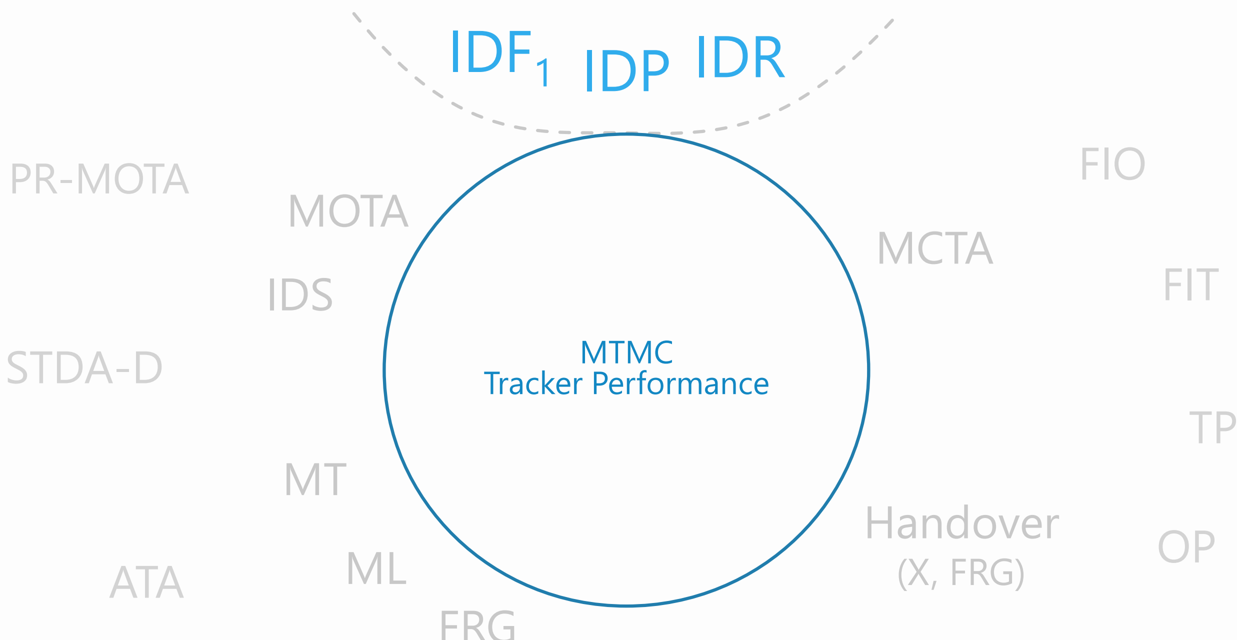


SOTA detector



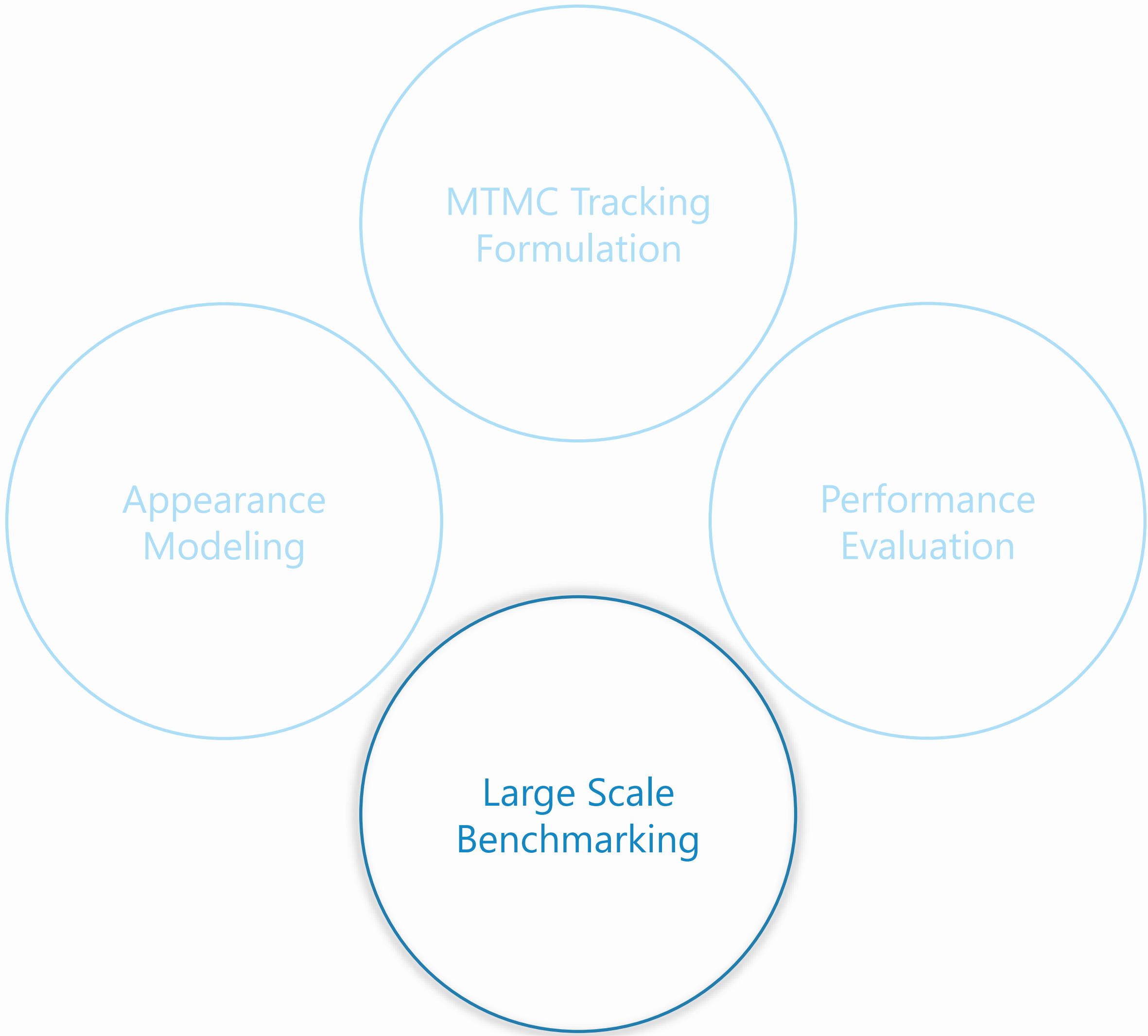
Summary

- New measures of performance appropriate for end-users
- Conceptually simple, focus on identity, fair

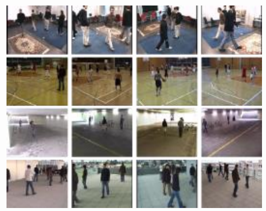


[2] Performance Measures and a Data Set for Multi-Target, Multi-Camera Tracking. Ristani et al. ECCV 2016 BMTT Workshop

My Research


MTMC Tracking Formulation
Appearance Modeling
Performance Evaluation
Large Scale Benchmarking

Previous MTMCT Data Sets

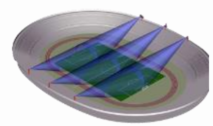


EPFL

NLPR 1, 2

NLPR 3

APIDIS

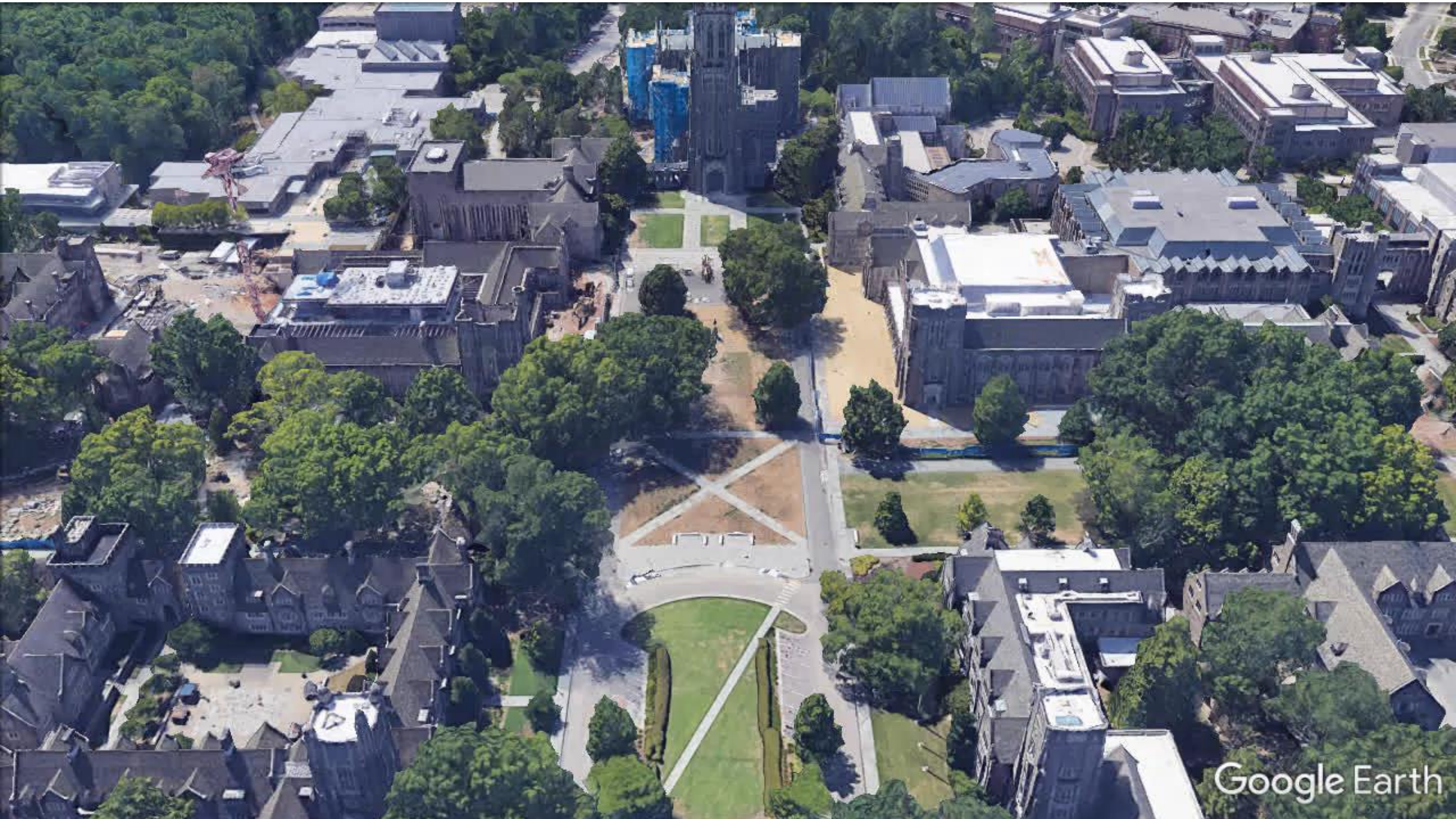


ISSIA

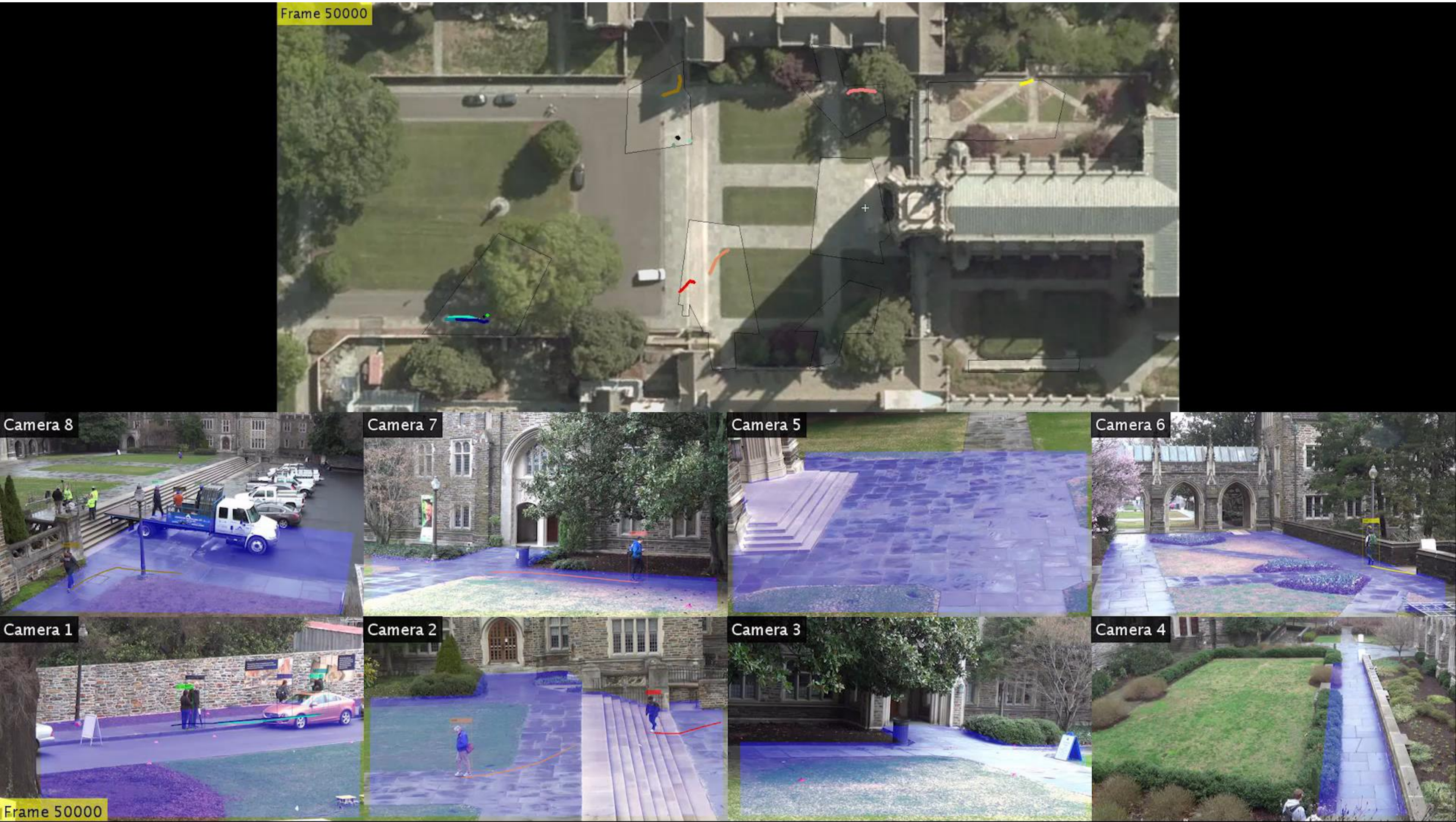
PETS2009

Dataset	IDs	Duration	Cams	Actors	Overlap	Blind Spots	Calib.	Resolution	FPS	Scene	Year
Laboratory	3	2.5 min	4	Yes	Yes	No	Yes	320x240	25	Indoor	2008
Campus	4	5.5 min	3	Yes	Yes	No	Yes	320x240	25	Outdoor	2008
Terrace	7	3.5 min	4	Yes	Yes	No	Yes	320x240	25	Outdoor	2008
Passageway	4	20 min	4	Yes	Yes	No	Yes	320x240	25	Mixed	2011
Issia Soccer	25	2 min	6	No	Yes	No	Yes	1920x1080	25	Outdoor	2009
Apidis Basket.	12	1 min	7	No	Yes	No	Yes	1600x1200	22	Indoor	2008
PETS2009	30	1 min	8	Yes	Yes	No	Yes	768x576	7	Outdoor	2009
NLPR MCT 1	235	20 min	3	No	No	Yes	No	320x240	20	Mixed	2015
NLPR MCT 2	255	20 min	3	No	No	Yes	No	320x240	20	Mixed	2015
NLPR MCT 3	14	4 min	4	Yes	Yes	Yes	No	320x240	25	Indoor	2015
NLPR MCT 4	49	25min	5	Yes	Yes	Yes	No	320x240	25	Mixed	2015
Dana36	24	N/A	36	Yes	Yes	Yes	No	2048x1536	N/A	Mixed	2012
USC Campus	146	25 min	3	No	No	Yes	No	852x480	30	Outdoor	2010
CamNeT	50	30 min	8	Yes	Yes	Yes	No	640x480	25	Mixed	2015

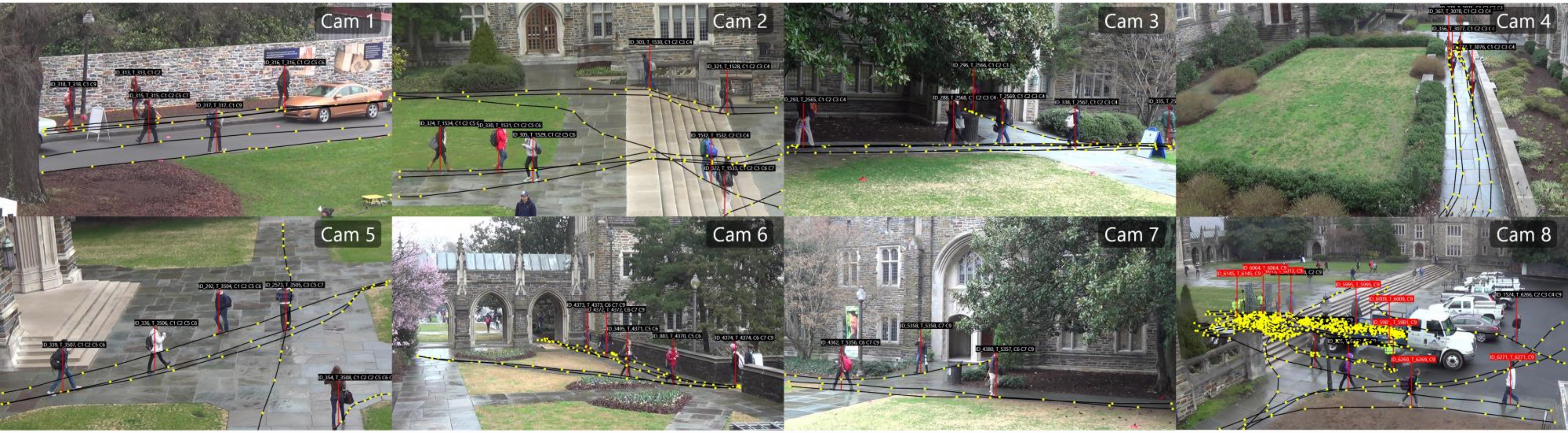
DukeMTMC Dataset



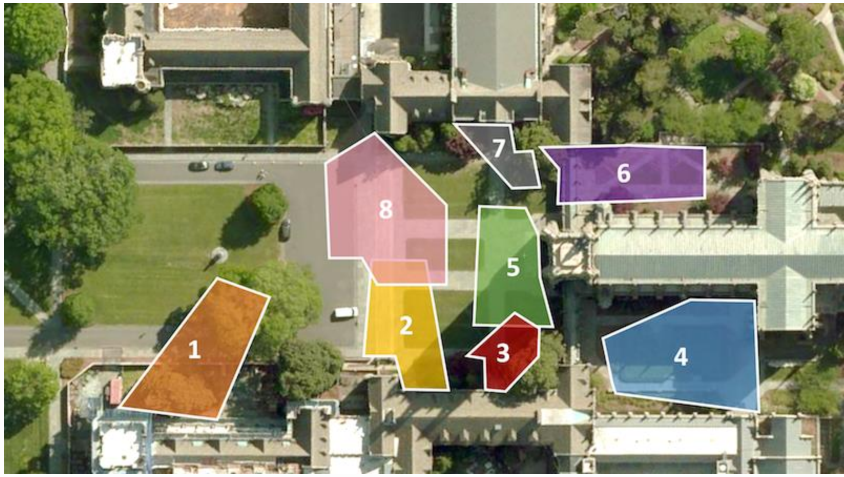
DukeMTMC Dataset



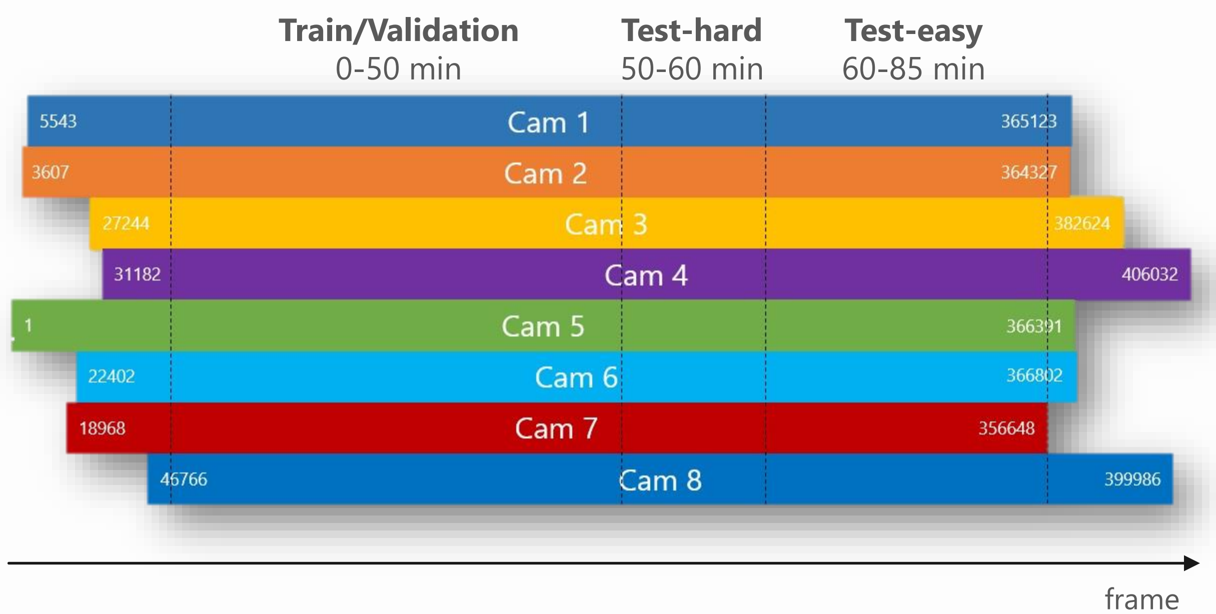
DukeMTMC Highlights



- 8 static cameras x 85 minutes of 1080p 60 fps video
- More than 2,000,000 manually annotated frames
- More than 2,000 identities
- Annotation by 5 people over 1 year
- More identities than all existing datasets combined
- Unconstrained paths, diverse appearance

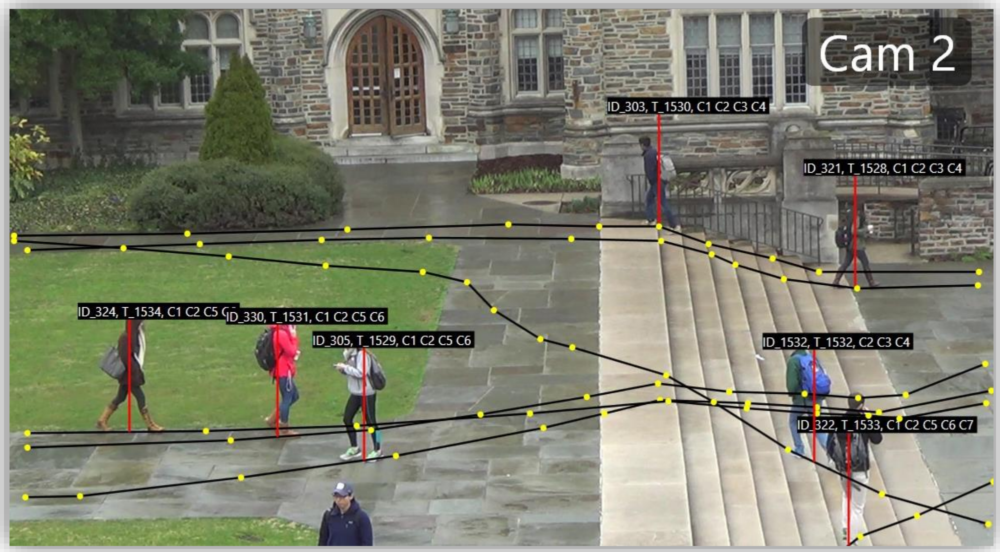


Train/Test Split



- 50 minutes of training data to enable deep learning in MTMCT

Annotation Protocol

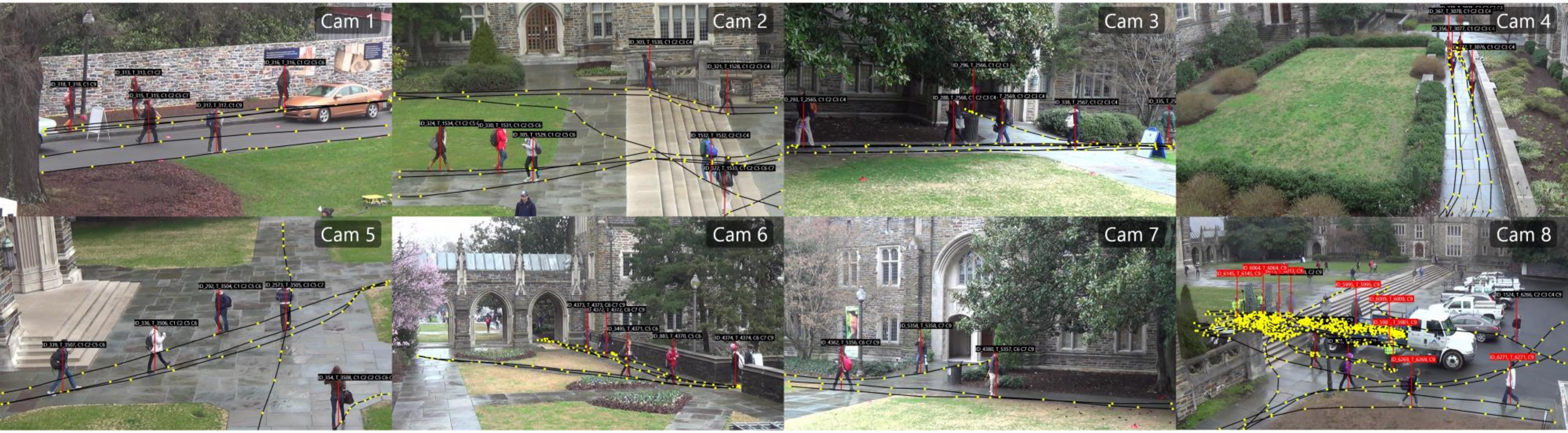


- Single camera annotation with point clicks on people's feet

Annotation Protocol

- Box height annotated once per trajectory per camera (then extrapolated)

Annotation Protocol



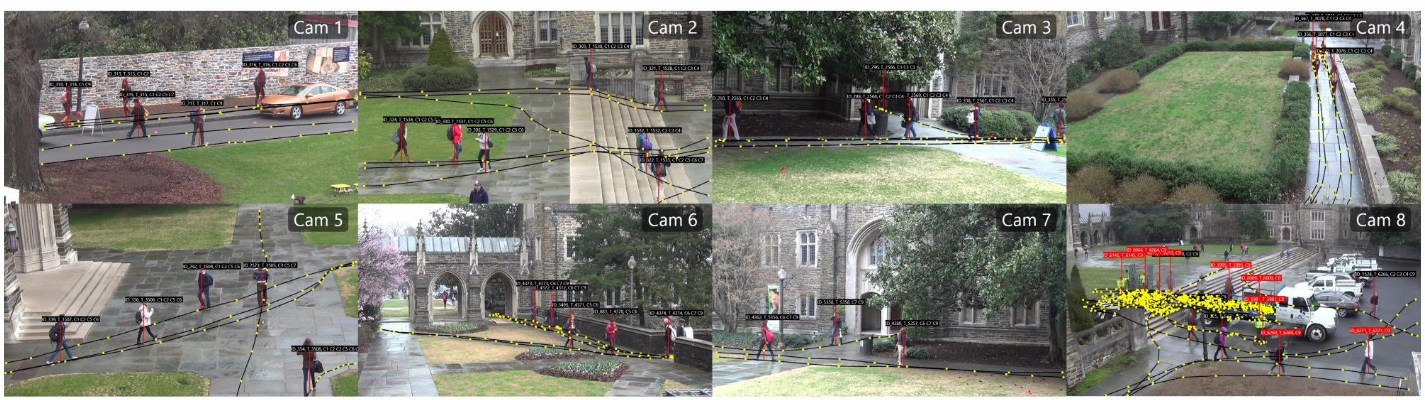
Query

Most Similar Trajectories

- Across camera annotation bootstrapped by person re-identification model

Dataset Impact

Dataset	IDs	Duration	Cams	Actors	Overlap	Blind Spots	Calib.	Resolution	FPS	Scene	Year
Laboratory	3	2.5 min	4	Yes	Yes	No	Yes	320x240	25	Indoor	2008
Campus	4	5.5 min	3	Yes	Yes	No	Yes	320x240	25	Outdoor	2008
Terrace	7	3.5 min	4	Yes	Yes	No	Yes	320x240	25	Outdoor	2008
Passageway	4	20 min	4	Yes	Yes	No	Yes	320x240	25	Mixed	2011
Issia Soccer	25	2 min	6	No	Yes	No	Yes	1920x1080	25	Outdoor	2009
Apidis Basket.	12	1 min	7	No	Yes	No	Yes	1600x1200	22	Indoor	2008
PETS2009	30	1 min	8	Yes	Yes	No	Yes	768x576	7	Outdoor	2009
NLPR MCT 1	235	20 min	3	No	No	Yes	No	320x240	20	Mixed	2015
NLPR MCT 2	255	20 min	3	No	No	Yes	No	320x240	20	Mixed	2015
NLPR MCT 3	14	4 min	4	Yes	Yes	Yes	No	320x240	25	Indoor	2015
NLPR MCT 4	49	25min	5	Yes	Yes	Yes	No	320x240	25	Mixed	2015
Dana36	24	N/A	36	Yes	Yes	Yes	No	2048x1536	N/A	Mixed	2012
USC Campus	146	25 min	3	No	No	Yes	No	852x480	30	Outdoor	2010
CamNeT	50	30 min	8	Yes	Yes	Yes	No	640x480	25	Mixed	2015
DukeMTMC	2834	85 min	8	No	Yes	Yes	Yes	1920x1080	60	Outdoor	2016



DukeMTMC@MOTChallenge

Multiple Object Tracking Benchmark

home | data | results | vis | QVA | submit | FAQ | people | login | sign up

Easy Test Set

Single Camera (all)

Tracker	IDF1	IDP	IDR	MOTA	MOTP	FAF	MT	ML	FP	FN	ID Sw.	Frag
MTMC_CDSC 1.	77.0	87.6	68.6	70.9	75.8	0.05	740	110	38,655	268,398	693	4,717
Y. Tesfaye, E. Zemene, A. Prati, M. Pelillo, M. Shah. Multi-Target Tracking in Multiple Non-Overlapping Cameras using Constrained Dominant Sets. In CoRR, 2017.												
PT_BIPCC 2.	71.2	84.8	61.4	59.3	78.7	0.09	666	234	68,634	361,589	290	783
Anonymous submission												
lx_b 3.	70.3	88.1	58.5	61.3	78.7	0.04	640	247	26,845	382,524	246	788
Anonymous submission												
BIPCC 4.	70.1	83.6	60.4	59.4	78.7	0.09	665	234	68,147	361,672	300	801
E. Ristani, F. Solera, R. Zou, R. Cucchiara, C. Tomasi. Performance Measures and a Data Set for Multi-Target, Multi-Camera Tracking. In ECCV workshop on Benchmarking Multi-Target Tracking, 2016.												
dirBIPCC 5.	70.0	83.2	60.4	59.0	78.7	0.10	665	234	71,381	361,673	298	799
Anonymous submission												

Multi-Camera

Tracker	IDF1	IDP	IDR
MTMC_CDSC 1.	60.0	68.3	53.5
Y. Tesfaye, E. Zemene, A. Prati, M. Pelillo, M. Shah. Multi-Target Tracking in Multiple Non-Overlapping Cameras using Constrained Dominant Sets. In CoRR, 2017.			
lx_b 2.	58.0	72.6	48.2
Anonymous submission			
BIPCC 3.	56.2	67.0	48.4
E. Ristani, F. Solera, R. Zou, R. Cucchiara, C. Tomasi. Performance Measures and a Data Set for Multi-Target, Multi-Camera Tracking. In ECCV workshop on Benchmarking Multi-Target Tracking, 2016.			
dirBIPCC 4.	52.1	62.0	45.0
Anonymous submission			
PT_BIPCC 5.	34.9	41.6	30.1
Anonymous submission			

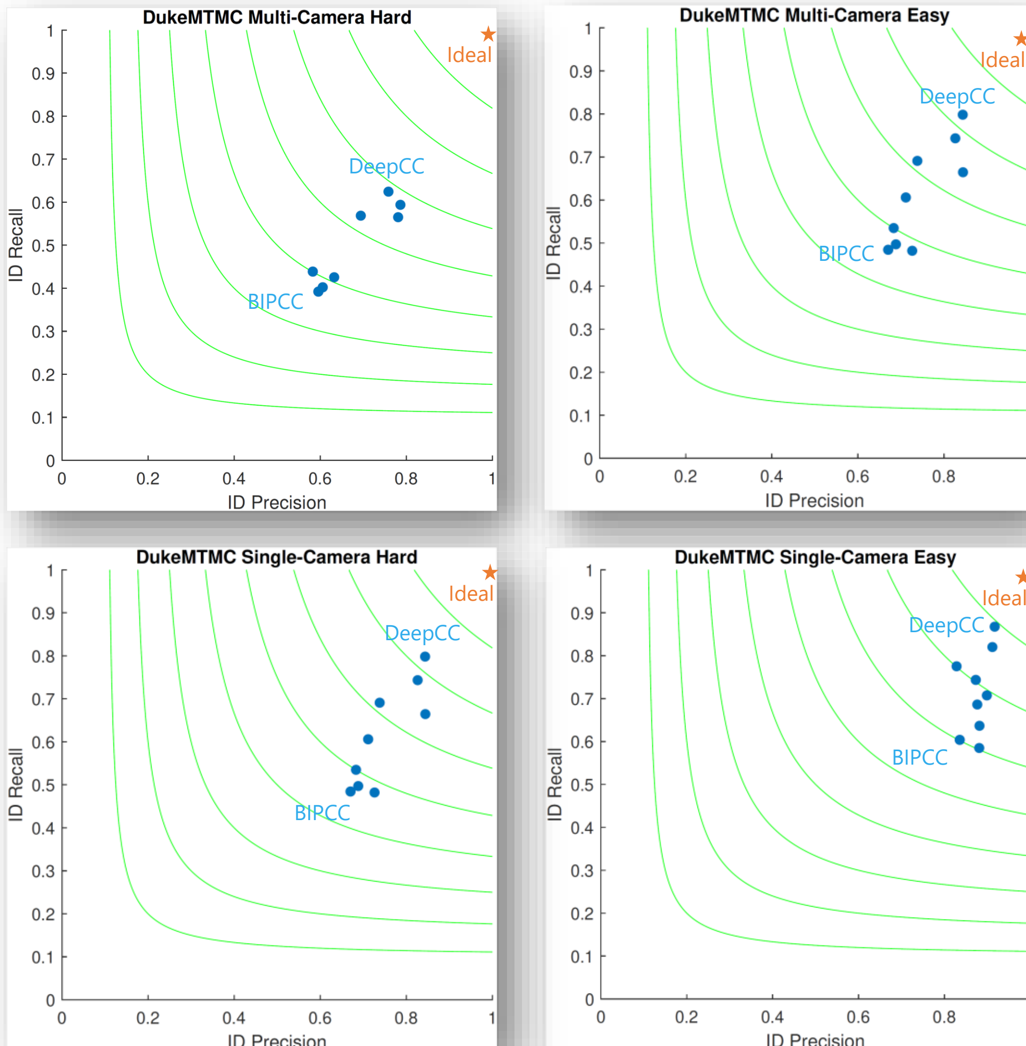
Recent Progress

Test-Hard

Test-Easy

Multi-Camera

Single-Camera



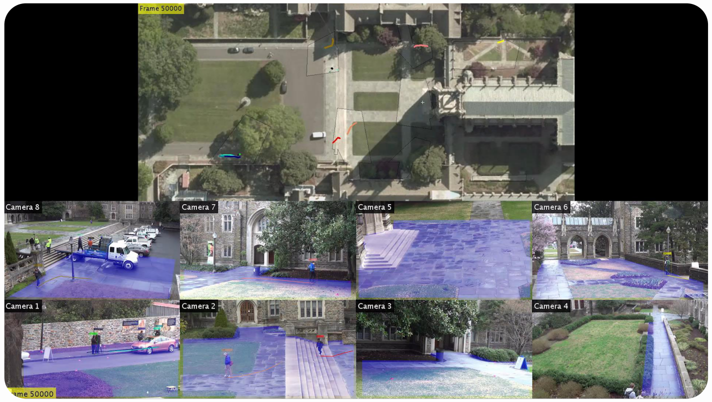
Summary

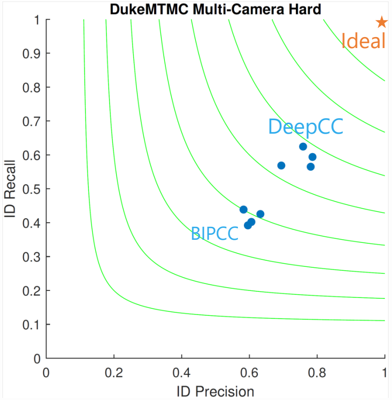
- New dataset which has enabled fast progress in ReID and MTMCT

	Euclidean	
	mAP	rank-1
BoW+kissme	12.17	25.13
LOMO+XQDA	17.04	30.75
Baseline	44.99	65.22
DPFL	48.90	70.10
PAN	51.51	71.59
HardTriplet	56.65	74.91
SVDNet	56.80	76.70
Ours	**58.74**	**77.69**
DPFL (2-stream)	60.60	79.20
Ours (2-stream)	**63.40**	**79.80**

Re-ID results on DukeMTMC-ReID

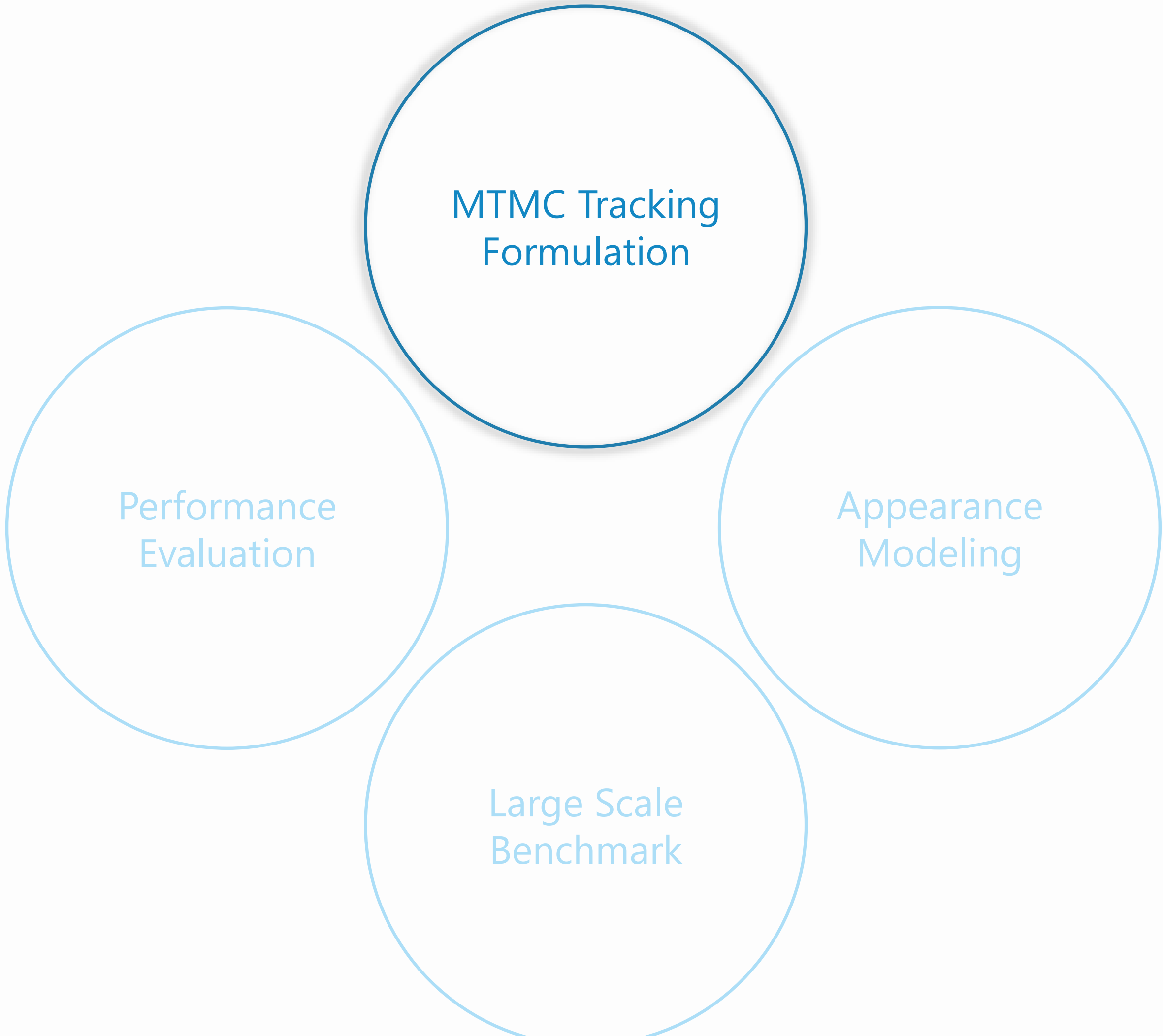
DukeMTMC





[2] Performance Measures and a Data Set for Multi-Target, Multi-Camera Tracking. Ristani et al. ECCV 2016 BMTT Workshop

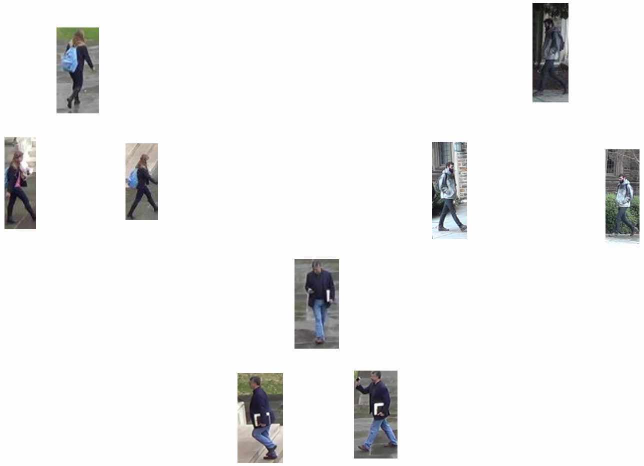
My Research


MTMC Tracking Formulation
Performance Evaluation
Appearance Modeling
Large Scale Benchmark

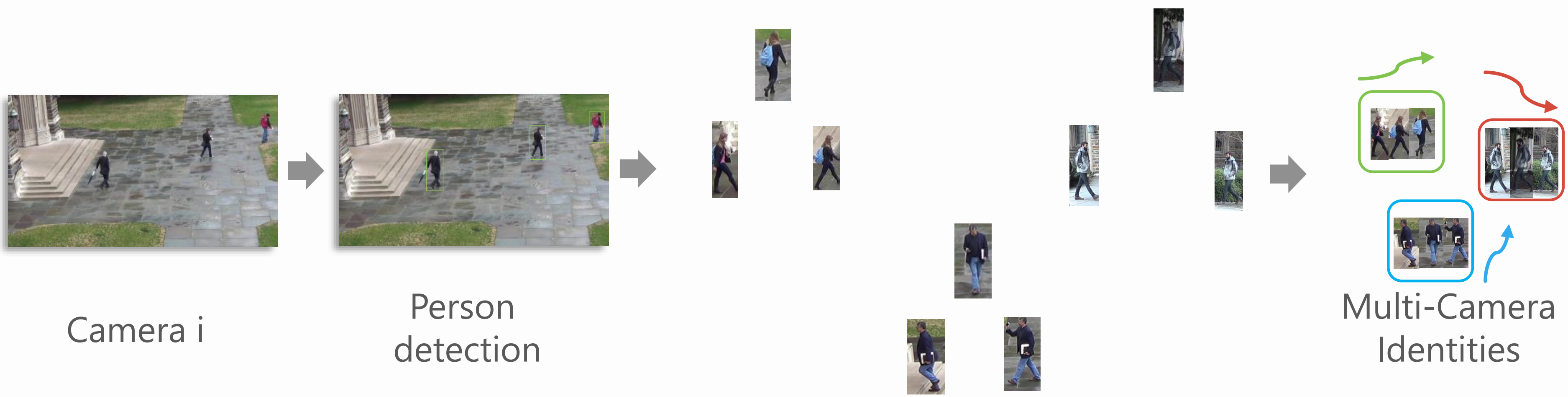
Problem Formulation

Camera i

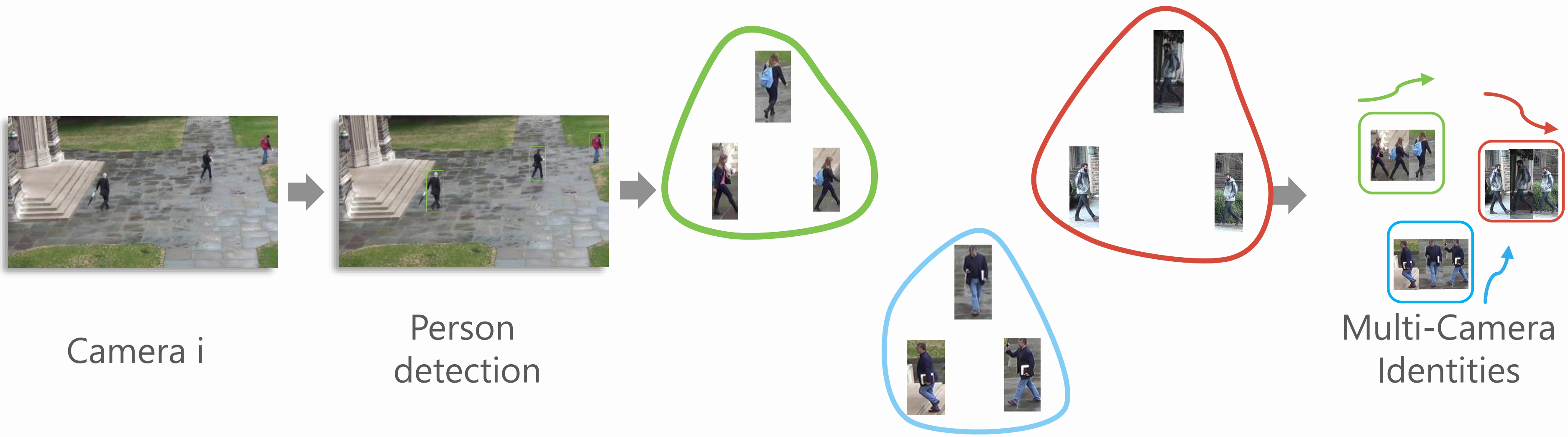
Person
detection



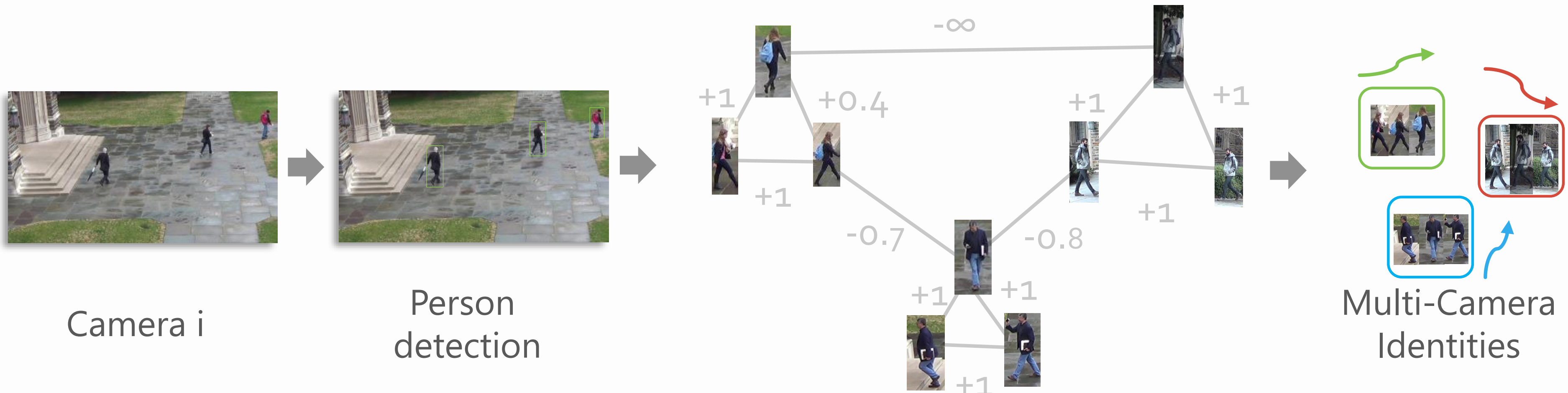
Problem Formulation



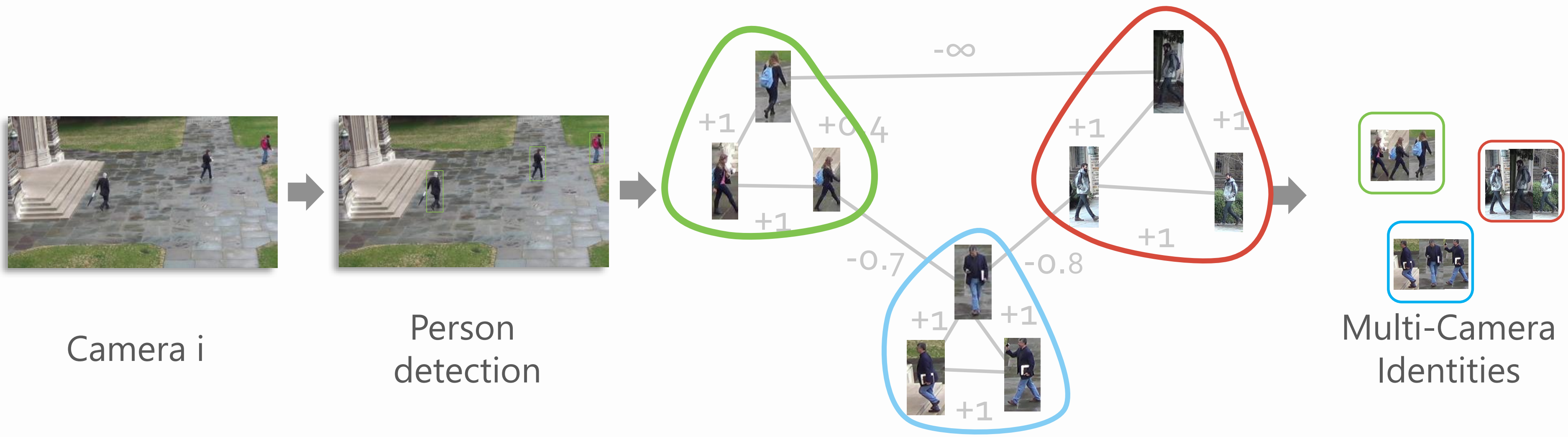
Problem Formulation



Problem Formulation



Problem Formulation



MTMC Tracking as Correlation Clustering

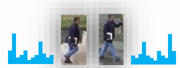
$$\arg\max_{X} \sum_{(u,v)\in E} w_{uv} x_{uv}$$

$$x_{uv} + x_{ut} \leq 1 + x_{vt} \qquad \forall (u,v),(v,t),(u,t) \in E$$

Computing Correlations

Indifference Threshold

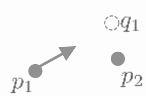
Appearance Correlation



$$w_{ij}^{A} = \frac{T^{A} - d(f_i, f_j)}{T^{A}}$$

Descriptor **d**istance

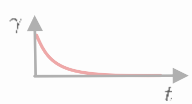
Motion Prediction Correlation



$$w_{ij}^{M} = \frac{T^{M} - c(p_i, p_j)}{T^{M}}$$

Forward/Backward Prediction **c**ost

Temporal Discount Factor



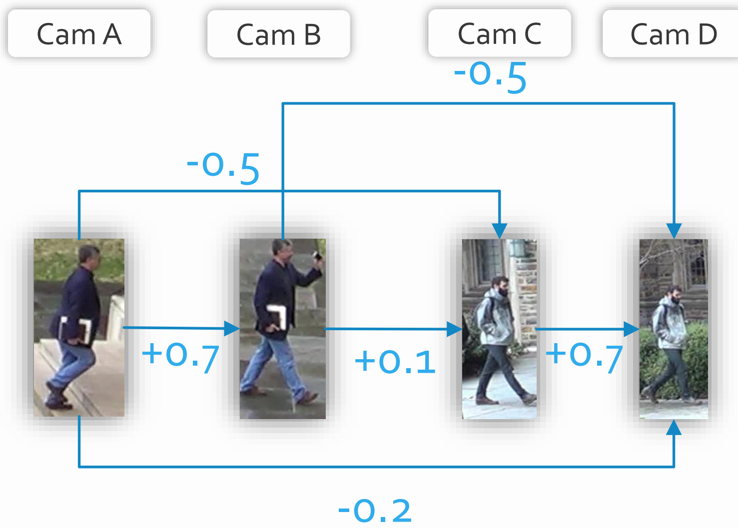
$$\gamma_{ij} = e^{\frac{-(t_i - t_j)^2}{\sigma}}$$

Correlations decay to zero as **t**ime difference increases

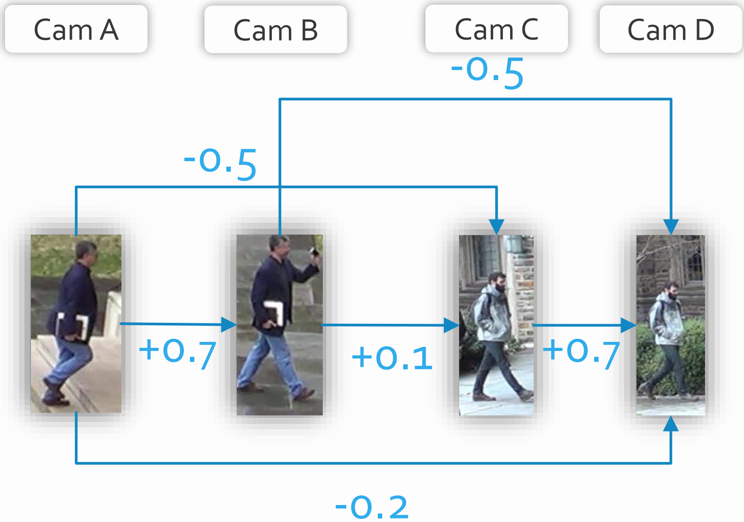
Final Correlation

$$w_{ij} = \gamma_{ij} \cdot (w_{ij}^{A} + w_{ij}^{M})$$

Correlation Clustering Advantage

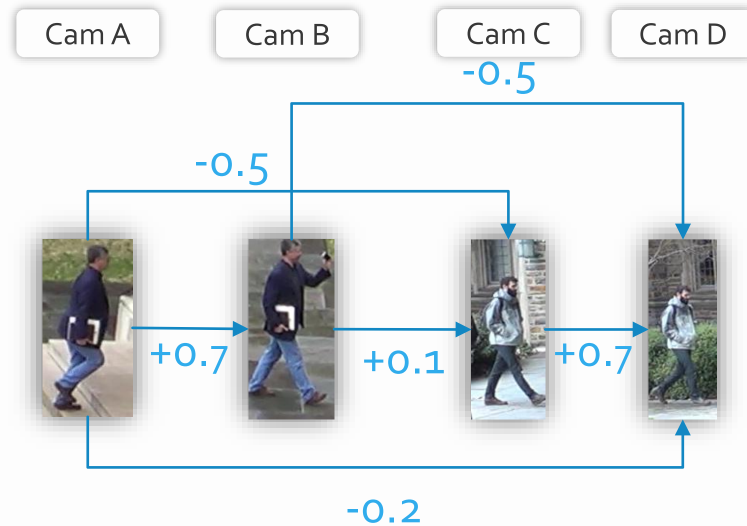


Correlation Clustering Advantage

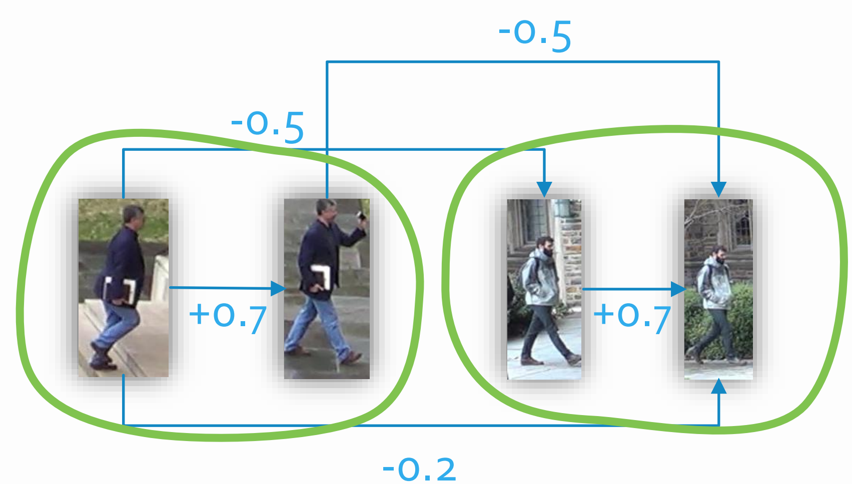


Consecutive edges solution

Correlation Clustering Advantage



Consecutive edges solution

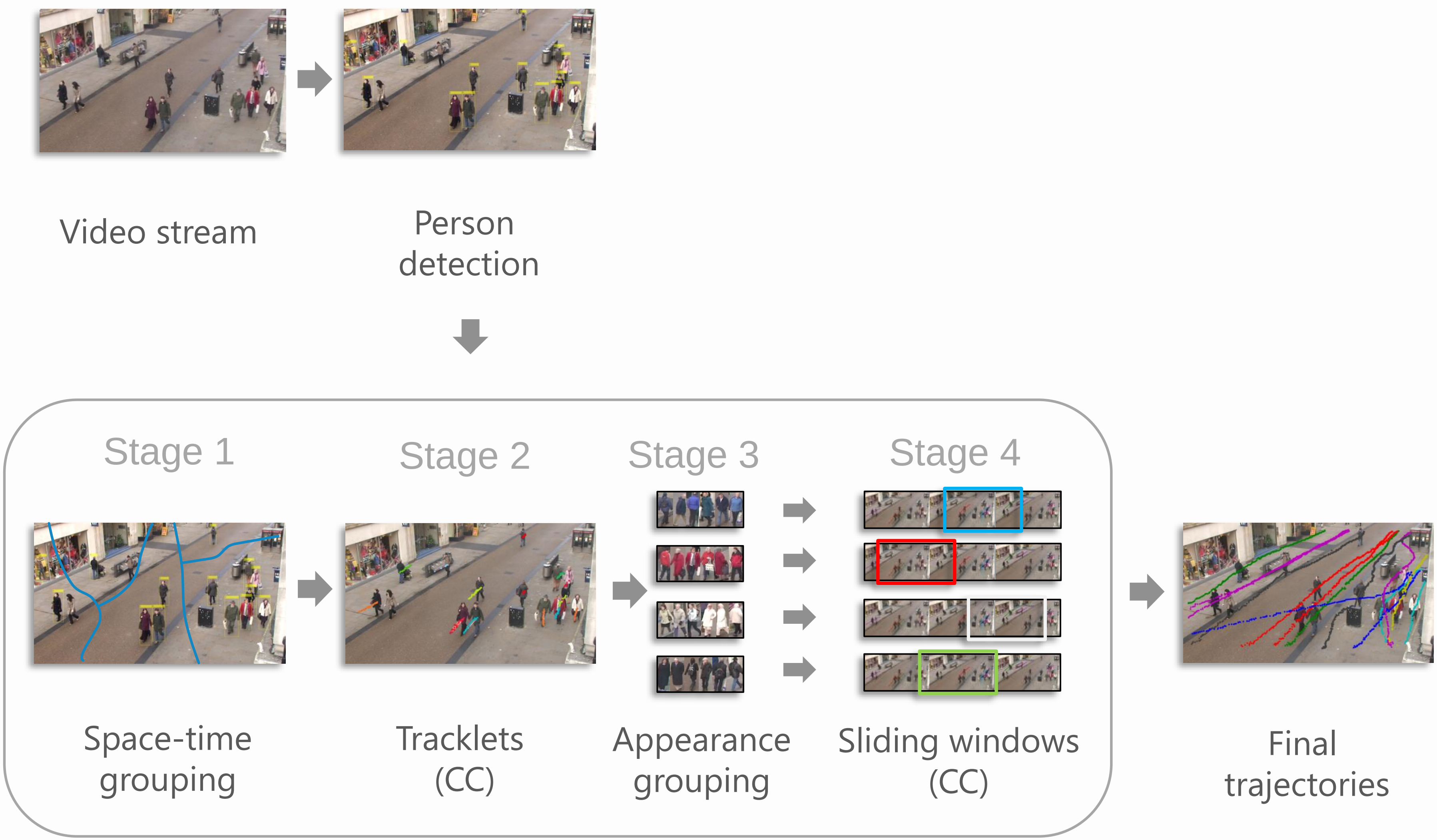


All-edges solution

Chronology of Formulations

- Appearance Modeling for Tracking in Multiple Non-Overlapping Cameras. Javed et al. CVPR 2005
  - Bipartite Matching
- Global Data Association for Multi-Object Tracking Using Network Flows. Zhang et al. CVPR 2008
- Multiple Object Tracking using K-Shortest Paths Optimization. Berclaz et al. TPAMI 2011
- Global Multi-Object Tracking Using Generalized Minimum Clique Graphs. Zamir et al. ECCV 2012
  - Solves one clique at a time
- Tracking Multiple People Online and in Real Time. Ristani and Tomasi. ACCV 2014
  - Correlation Clustering
- Consistent Re-Identification in a Camera Network. Das and Roy-Chowdhury. ECCV 2014
  - Identical objective
  - Additional constraint forbids associating identities that re-enter the same camera
- Subgraph decomposition for multi-target tracking. Tang et al. CVPR 2015
  - Adds unary terms
- Globally Optimal Generalized Maximum Multi Clique Problem for Multiple Object Tracking. Dehgan et al. CVPR 2015
  - Adds unary terms
  - Forbids association of co-occurrent detections
- Multi-Person Tracking by Multicuts and Deep Matching. Tang et al. BMTT 2016
  - Equivalent
- MTT in Multiple Non-Overlapping Cameras using Constrained Dominant Sets. Tesfaye et al. CoRR 2017
  - Solves one cluster at a time
- Multi-Person Tracking by Multicut and Deep Matching. Tang et al. CVPR 2017
  - Adds path consistency constraints
- Features for Multi-Target Multi-Camera Tracking and Re-Identification. Ristani and Tomasi. CVPR 2018
  - Correlation Clustering

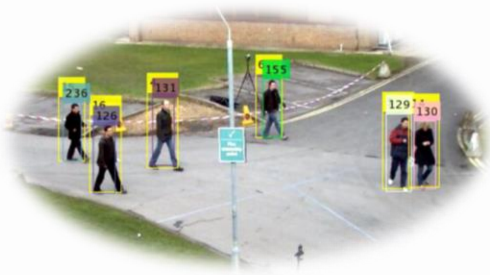
Single-Camera Tracker



[1] Tracking Multiple People Online and in Real Time. Ergys Ristani and Carlo Tomasi. ACCV 2014

Results

PETS2009-S2L1 View 1 sequence



	MOTA (%)	MOTP (%)	PREC. (%)	REC. (%)	ID sw
Berclaz	80.00	58.00	81.00	60.00	28
Shitrit	81.46	58.38	90.66	90.81	19
Andriyenko	81.84	73.93	96.28	85.13	15
Henriques	84.77	68.74	92.40	94.03	10
Zamir	90.30	69.02	93.64	**96.45**	8
Izadinia	90.7	76	96.8	95.2	N/A
Ours	**93.34**	**77.76**	**98.37**	95.71	**3**

Results

Town Center sequence

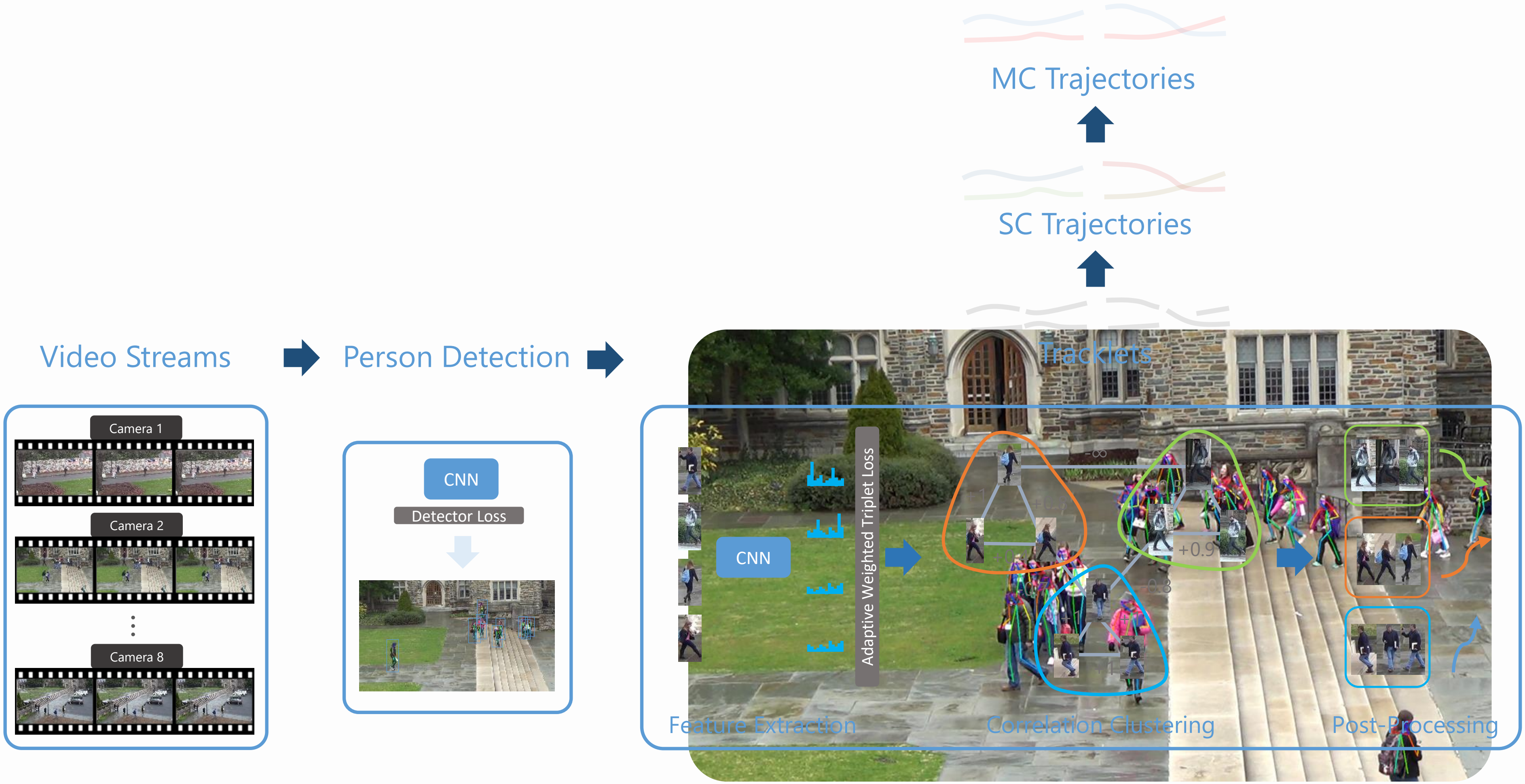
	MOTA (%)	MOTP (%)	PREC. (%)	REC. (%)	ID sw
Benfold	64.9	**80.4**	80.5	64.8	259
Zhang	65.7	71.5	71.5	66.1	114
Leal-Taixe	67.3	71.5	71.6	67.6	86
McLaughlin	74.15	72.41	90.40	83.27	N/A
Zamir	75.59	71.93	92.65	81.64	N/A
Izadinia	75.7	71.6	**93.6**	81.8	N/A
Ours	**78.8**	71.52	92.20	**86.51**	**68**

Results

Parking Lot sequence

	MOTA (%)	MOTP (%)	PREC. (%)	REC. (%)	ID sw
Zamir	92.27	**85.52**	95.70	96.64	**1**
Ours	**94.20**	85.38	**96.13**	**98.50**	**1**

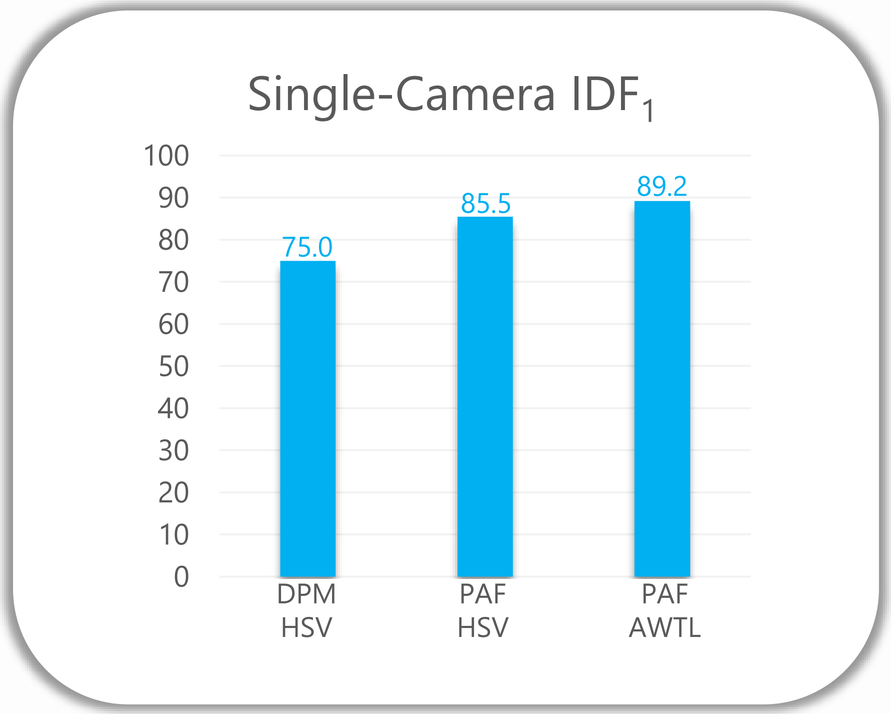
Pipeline: DeepCC



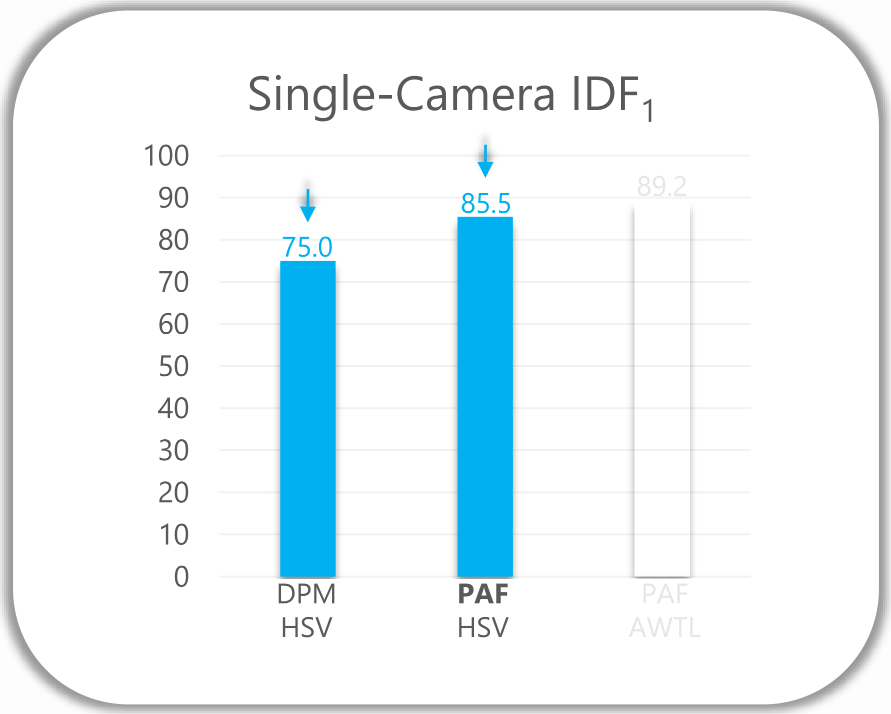
Detections from OpenPose. Cao et al. CVPR 2017

Pipeline extends BIPCC from *Ristani and Tomasi*. ACCV 2014

Ablation Experiments

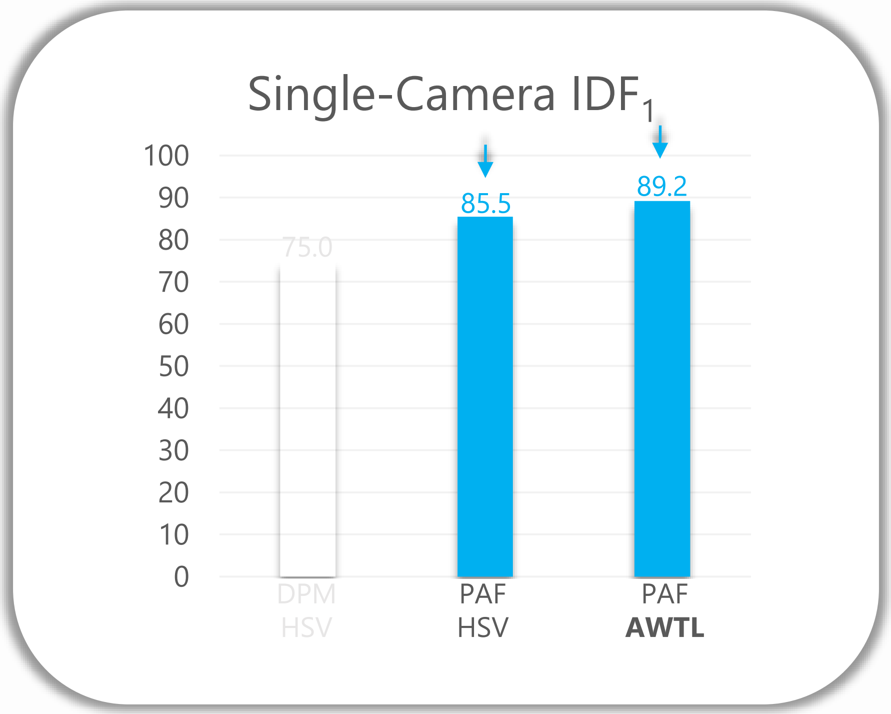


Ablation Experiments



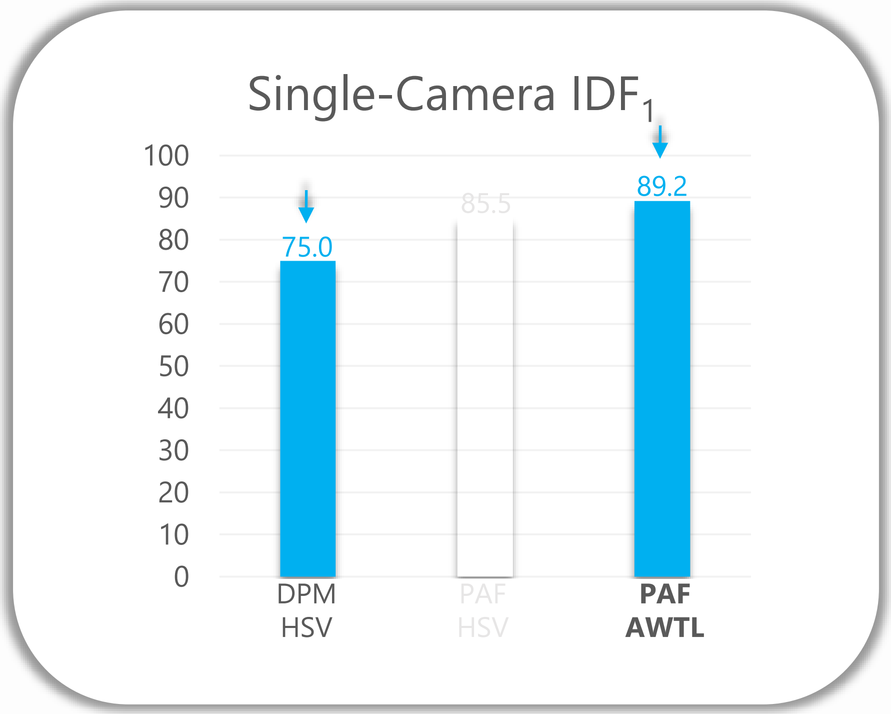
Improved detections help significantly in single-camera tracking

Ablation Experiments



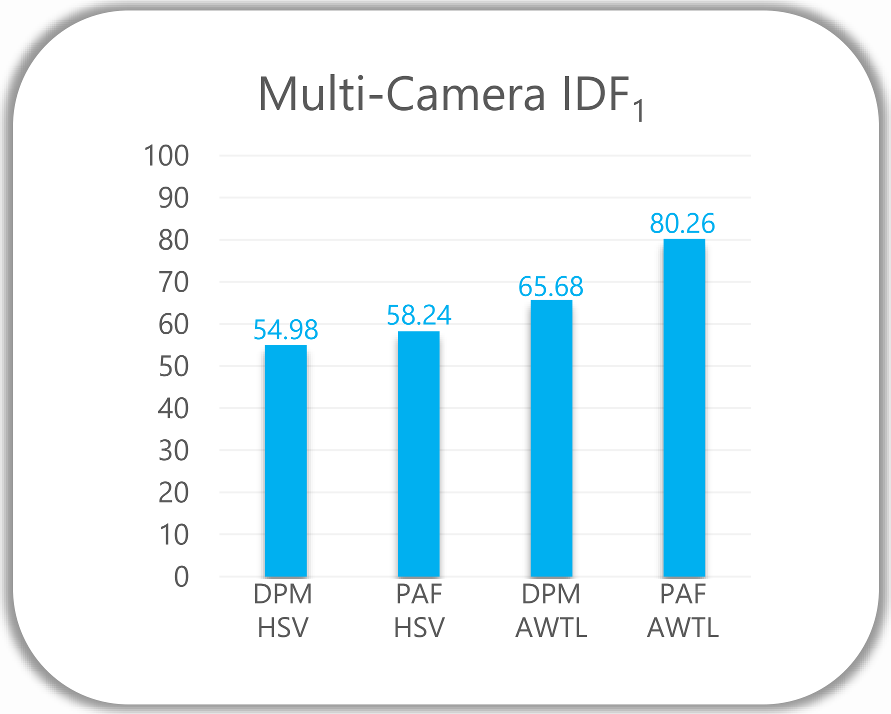
Significantly better features not as useful when detections are good

Ablation Experiments

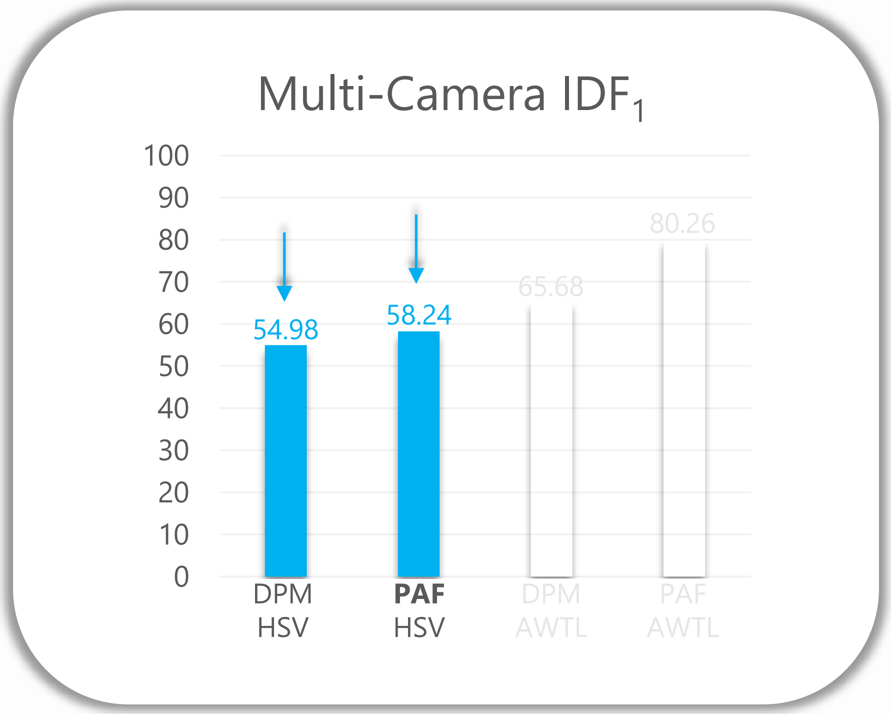


Best single-camera performance combines better detections and features

Ablation Experiments

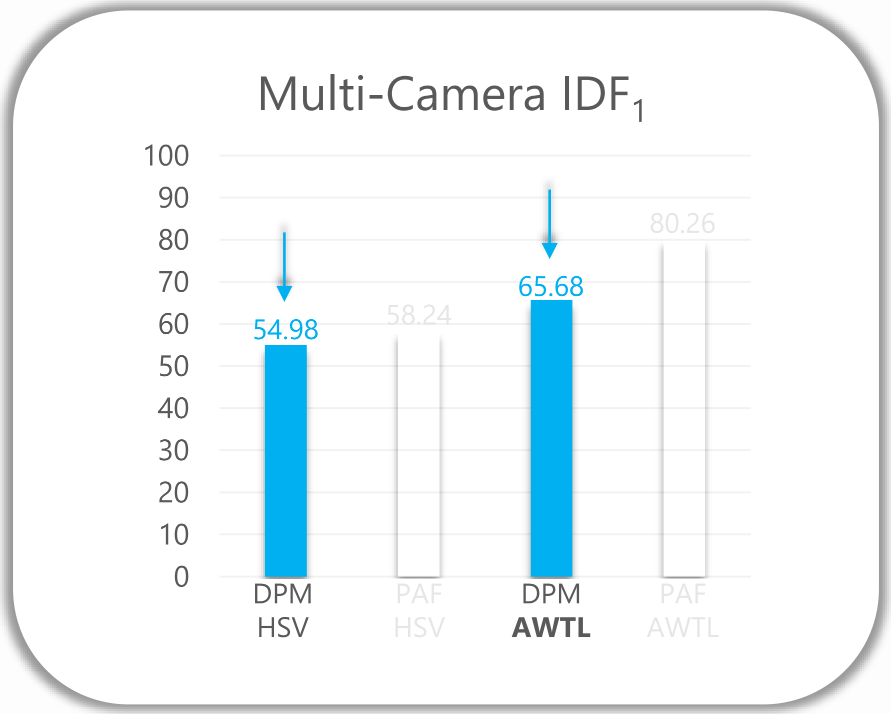


Ablation Experiments



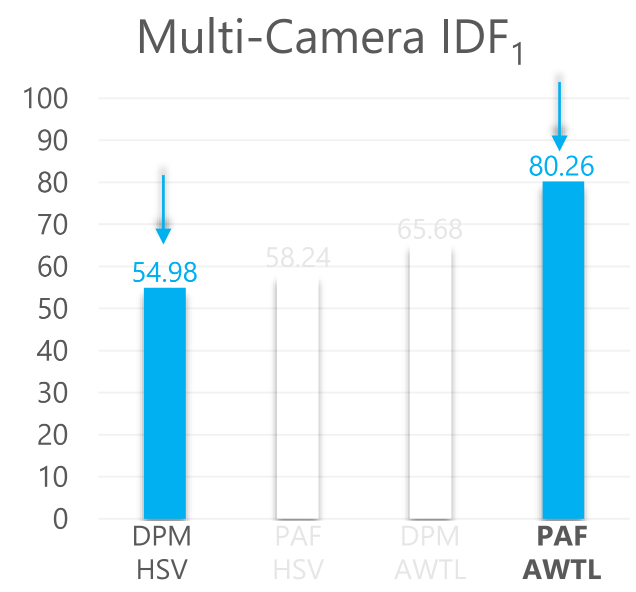
Improved detections help little across cameras when features are weak

Ablation Experiments



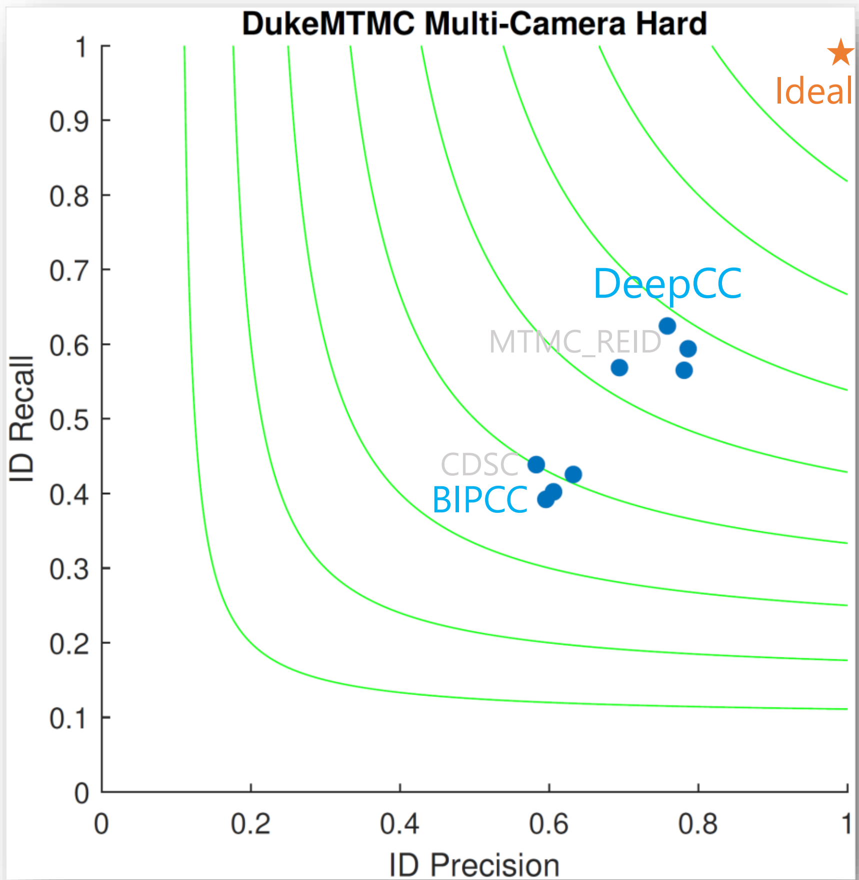
Good features more useful than good detector in MTMC tracking

Ablation Experiments

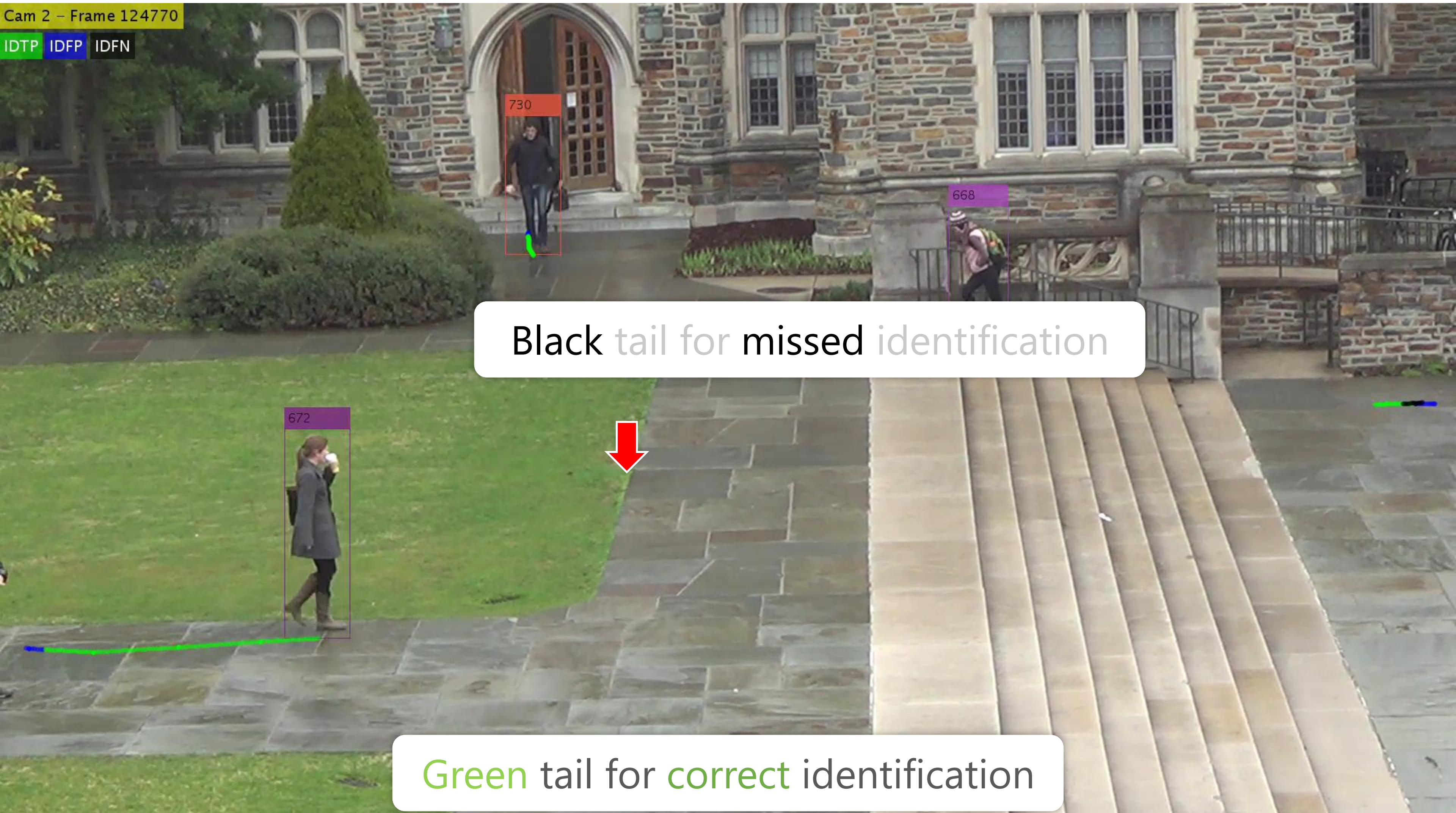


Better detections and features give best MTMC tracking performance

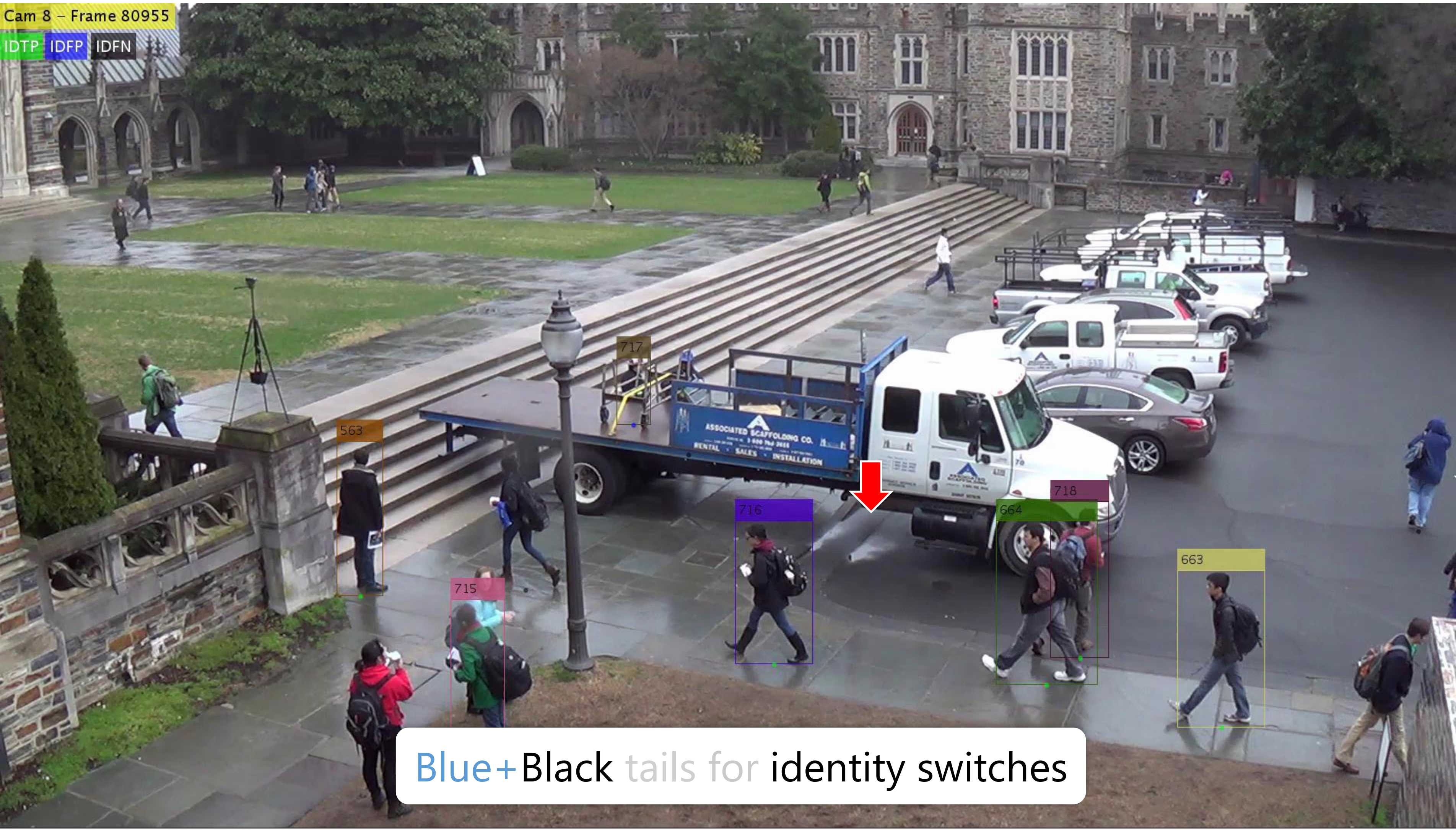
State Of The Art Comparison



Qualitative Results



Qualitative Results



Blue+Black tails for identity switches

Summary

- Correlation Clustering formulation improves over previous work
- State of the art DeepCC tracker

[1] Tracking Multiple People Online and in Real Time. Ristani and Tomasi. ACCV 2014
[2] Features for Multi-Target Multi-Camera Tracking and Re-Identification. Ristani and Tomasi. CVPR 2018

My Research

MTMC Tracking Formulation

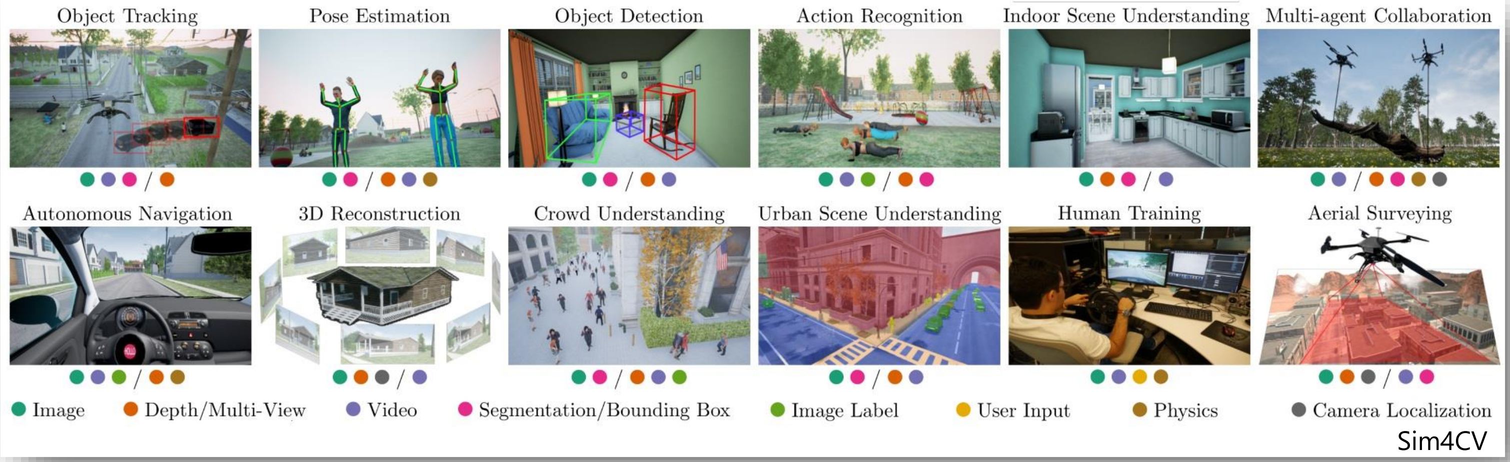
Appearance Modeling

Performance Evaluation

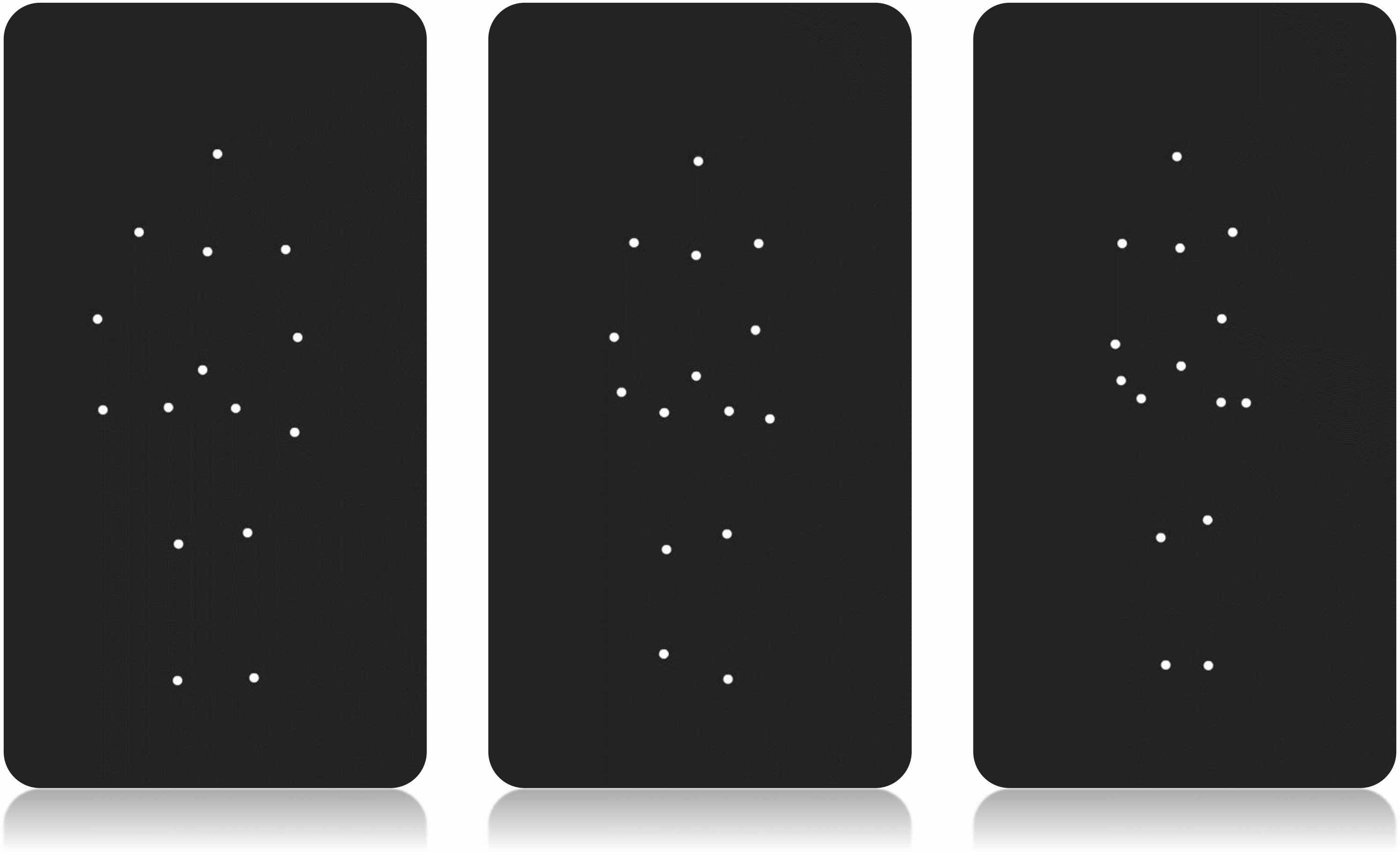
Large Scale Benchmark

Future Work

Photo-Realistic Data

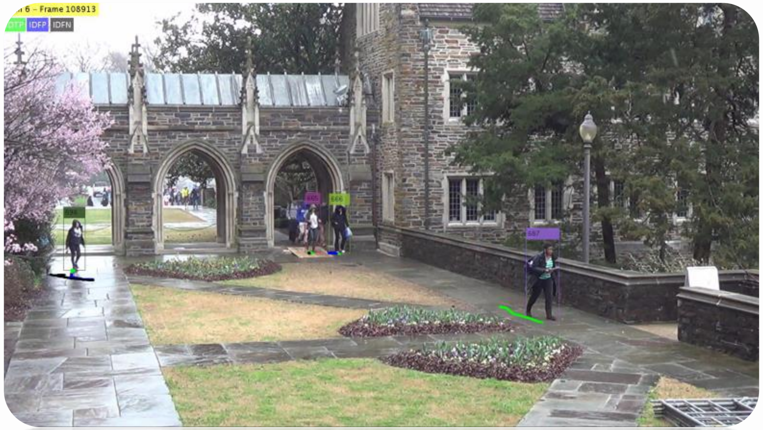


Motion-based ReID



Reasoning in 3D

Thank you!



Code/Data:

References:

[1] Tracking Multiple People Online and in Real Time
Ergys Ristani and Carlo Tomasi
ACCV 2014

[2] Performance Measures and Dataset for Multi-Target, Multi-Camera Tracking
Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, Carlo Tomasi
ECCV 2016 Workshop on Benchmarking Multi-Target Tracking

[3] Tracking Social Groups Within and Across Cameras
Francesco Solera, Simone Calderara, Ergys Ristani, Carlo Tomasi, Rita Cucchiara
IEEE Transactions on Circuits and Systems 2016

[4] Features for Multi-Target Multi-Camera Tracking and Re-Identification
Ergys Ristani and Carlo Tomasi
CVPR 2018