Features for Multi-Target Multi-Camera Tracking and Re-Identification

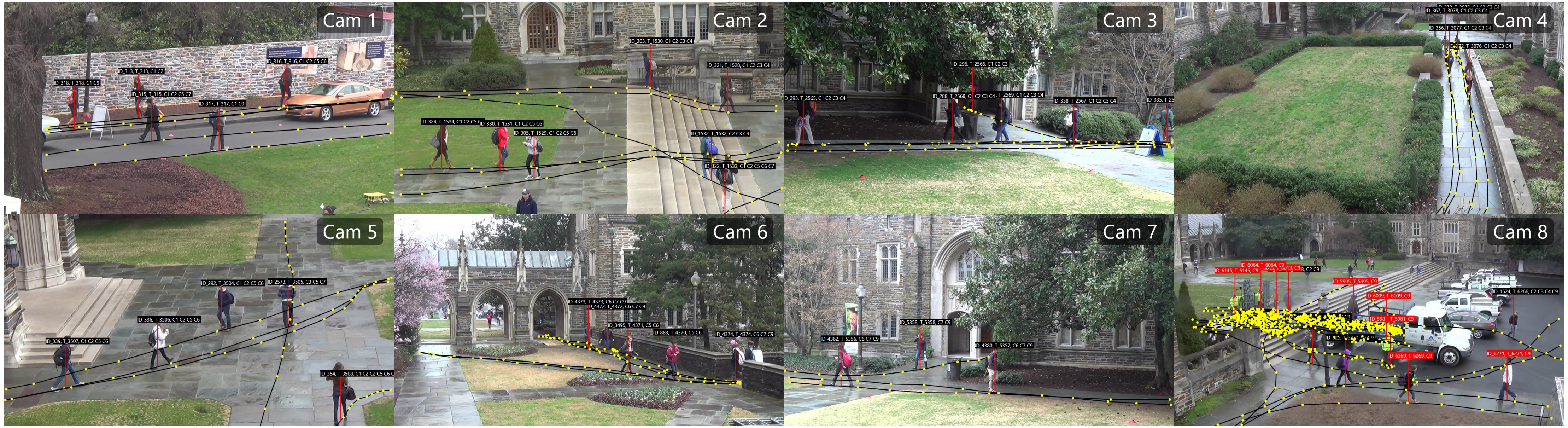
Ergys Ristani, Carlo Tomasi

Duke University, Durham, NC, USA

CVPR 2018 Spotlight

MTMCT & Re-ID

- DukeMTMC
- DukeMTMC-reID



Difficulties of MTMCT

- Cameras are placed far apart, no overlap FOVs
- Occlusion
- Large change of view point
- Different illumination
- Unknown people
- …

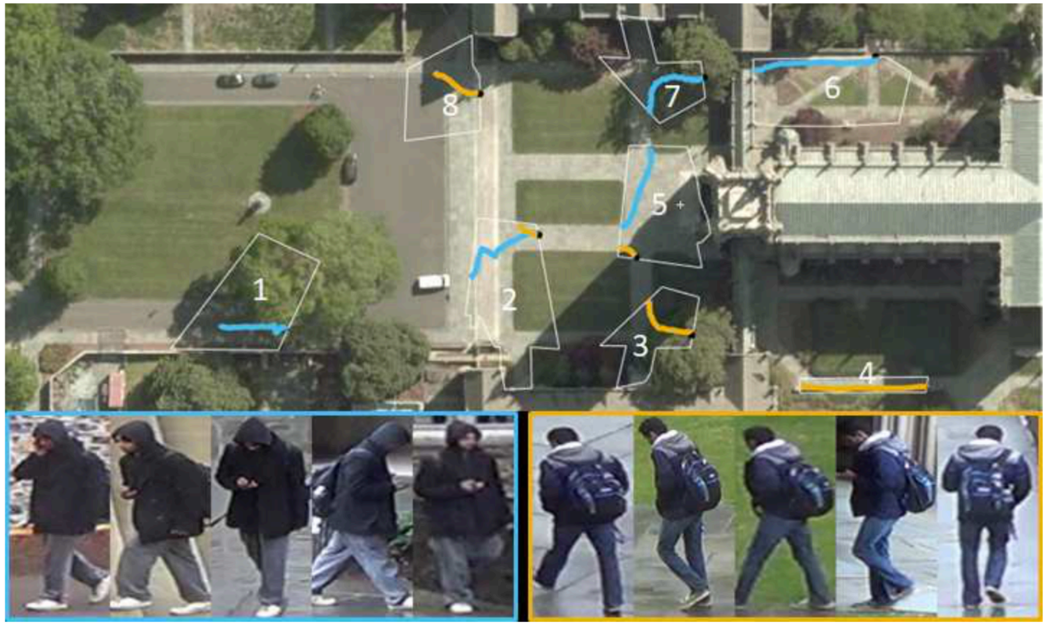


Figure 1. Two example multi-camera results from our tracker on the DukeMTMC dataset.

Difference Between MTMCT and Re-ID

- **Definition**:
  - Re-ID: Rank distances to a query.
  - MTMCT: Classify a pair of images as being co-identical or not.

- **Different metric**:
  - Re-ID: Ranking performance.
  - MTMCT: Classification error rate.

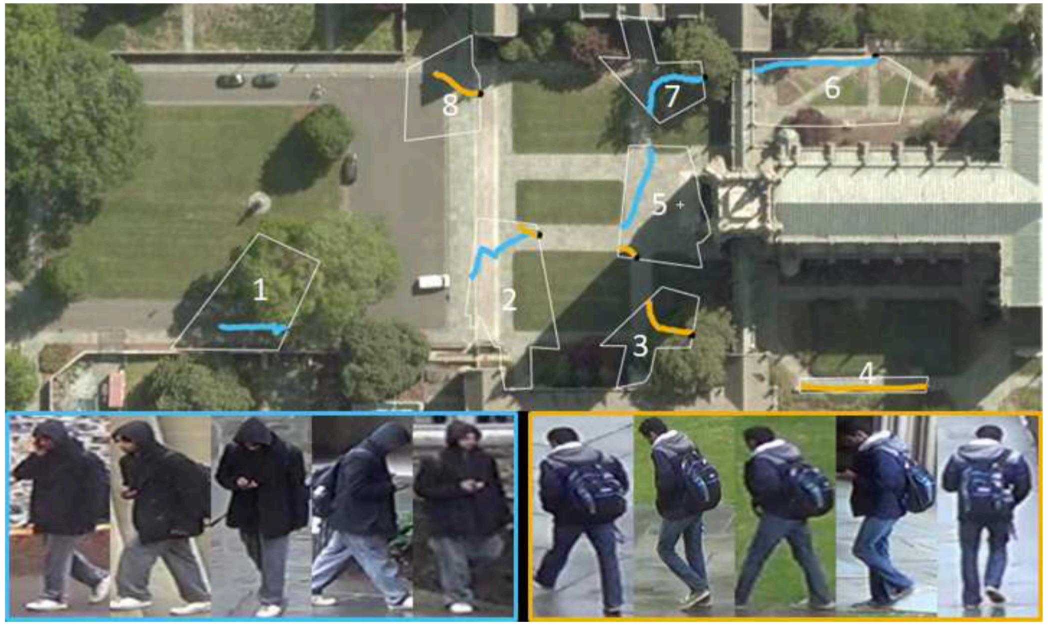


Figure 1. Two example multi-camera results from our tracker on the DukeMTMC dataset.

Difference Between MTMCT and Re-ID

- **Re-ID**: For any query, the largest distance between the query and the image with same ID is smaller than the smallest distance between the query and the image with different ID.
- **MTMCT**: The largest distance between any two images with same ID is smaller than the smallest distance between any two images with different ID.

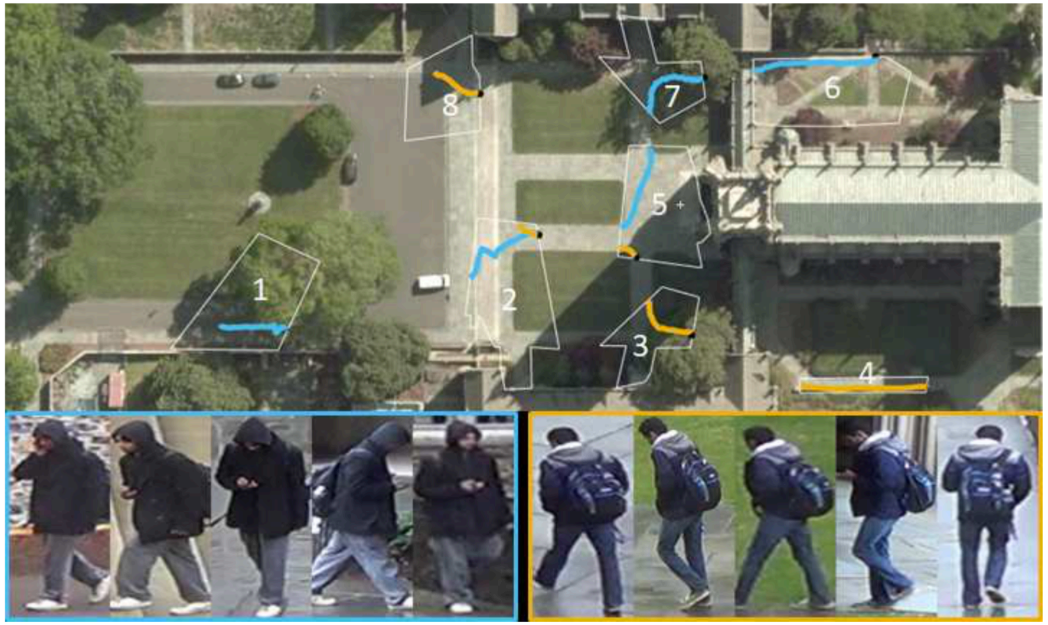
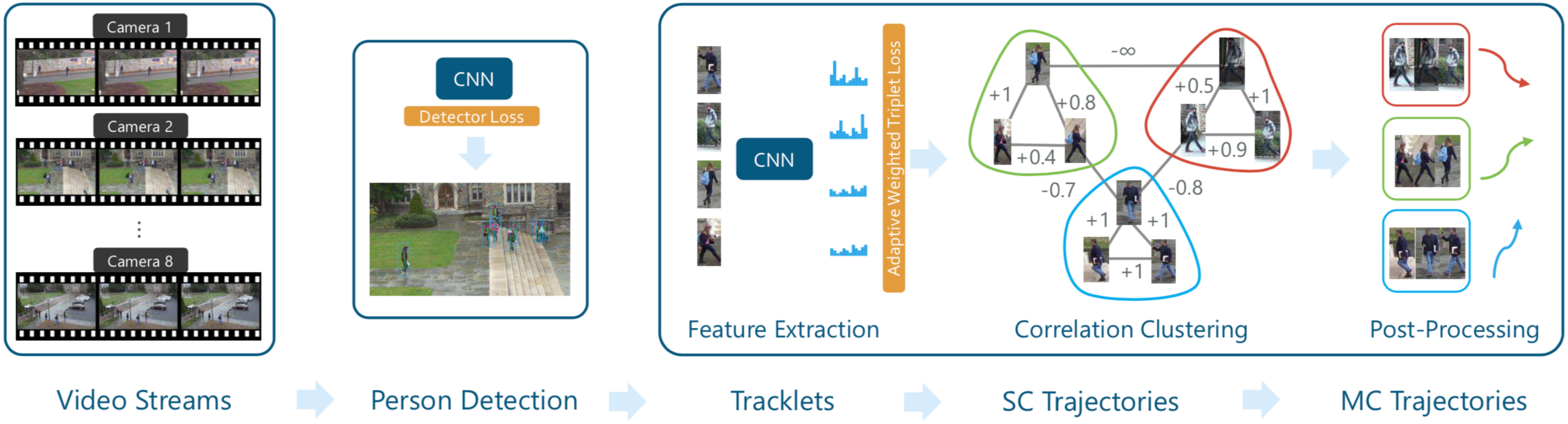


Figure 1. Two example multi-camera results from our tracker on the DukeMTMC dataset.

Proposed Method

- Learning Appearance Feature
- MTMC Tracker



A. Learning Appearance Feature

- **General Triplet Loss:**
  - $L_3 = [m + \sum_{x_p \in P(a)} w_p d\left(x_a, x_p\right) - \sum_{x_n \in N(a)} w_n d(x_a, x_n)]_+$
  - where, m is the given inter-person separation margin, d is the distance, x_a, x_p, x_n are features of the anchor sample, the positive sample and the negative sample.
- **The positive/negative class imbalance** can be handled by reflecting it in the weight distribution.

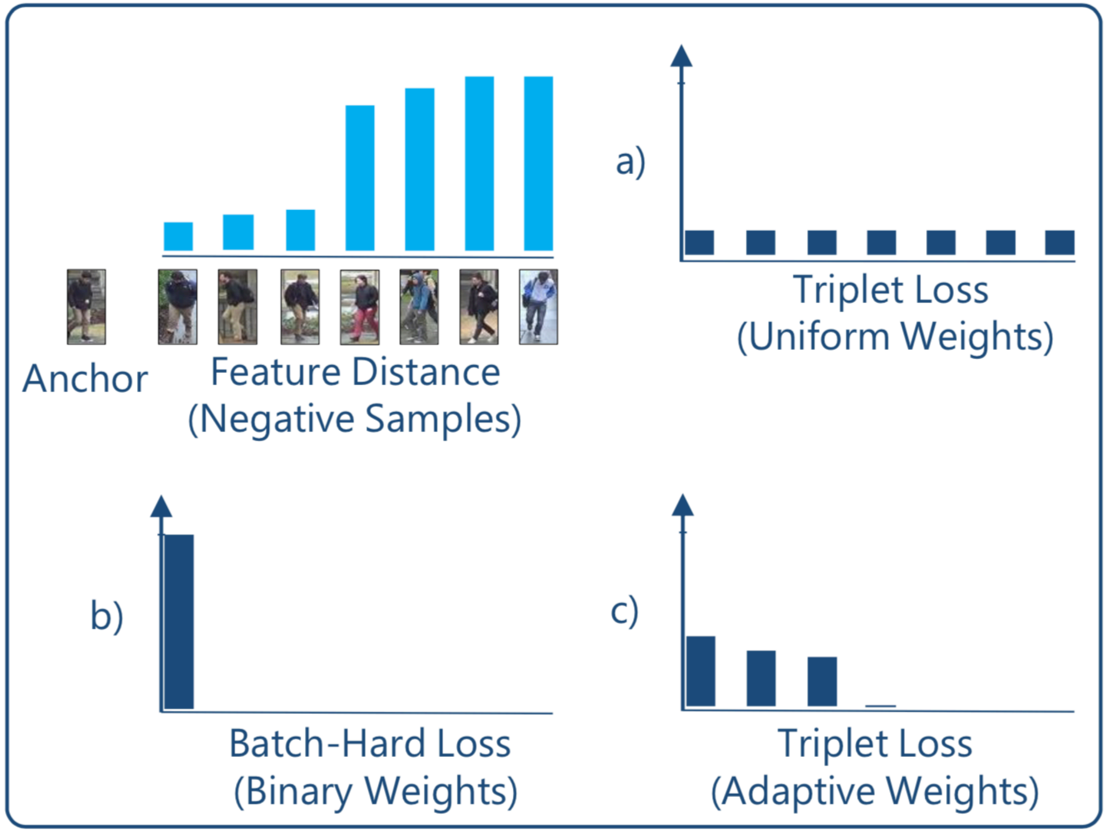
A-0. Hermans Batch-Hard Triplet Loss

- **The loss consider only the most difficult positive and negative sample.**
  - $w_p = \left[x_p == \arg\max_{x \in P(a)} d(x_a, x) \right]$
  - $w_n = \left[x_n == \arg\min_{x \in N(a)} d(x_n, x) \right]$
- **The advantage and disadvantage of the uniformed weighted loss:**
  - +: Robust to outliers
  - -: Wash out contributions of hard samples
- How to maintain both the performance of the batch-hard loss and robusts to outliers?

- A. Hermans, L. Beyer, and B. Leibe. In defense of the triplet loss for person re-identification. arXiv preprint, 2017.

A-1. Adaptive Weights

- Using softmax/min weight distributions:
- $w_p = \frac{e^{d(x_a, x_p)}}{\sum_{x \in P(a)} e^{d(x_a, x)}}$
- $w_n = \frac{e^{-d(x_a, x_n)}}{\sum_{x \in N(a)} e^{-d(x_a, x)}}$



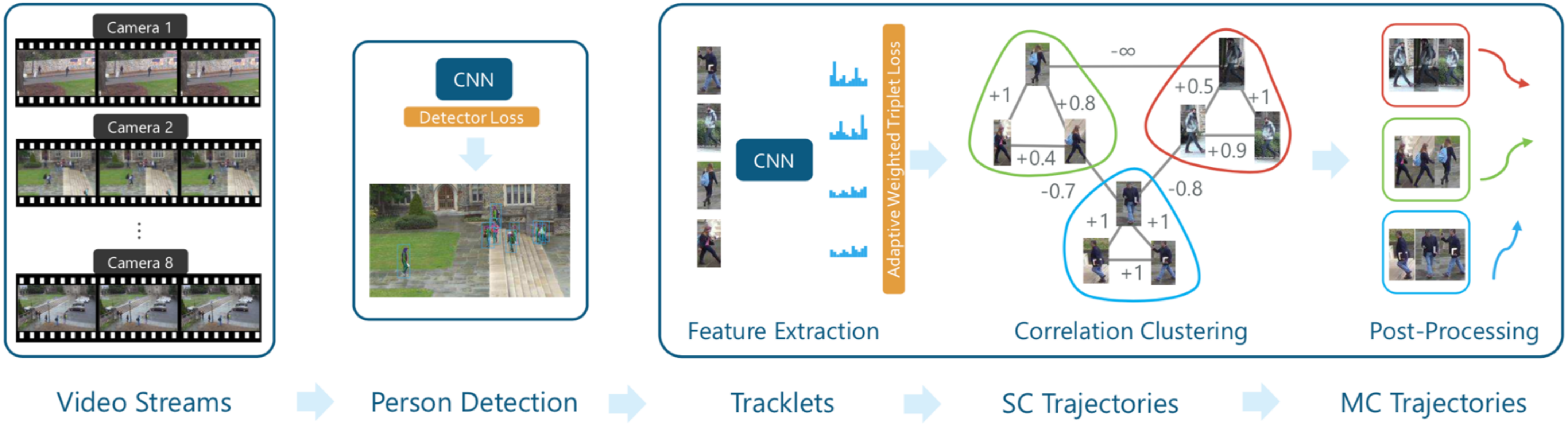
A-2. Hard-Identity Mining Scheme

- **Batch Construction**: PK Batches (proposed in Hermans paper)
  - In each batch there are K sample images for each of P identities.
- **General scheme**: During a training epoch, each identity is selected in its batch in turn, and the remaining P-1 identities are sampled at random.
- **Disadvantage of the general scheme**: Random sampling rarely picks the hardest negatives when the size of the training set increases.

A-2. Hard-Identity Mining Scheme

- **The hard identity pool** consists of the H most difficult identities given the anchor.
- **The random identity pool** consists of the remaining identities.
- Scheme: In a PK batch, we sample the remaining P-1 identities from the hard or random identity pool with equal probability.

B. MTMC Tracker



- **Detection**: OpenPose (CVPR 2017) Person detector.
- **Post-Processing**: Interpolation (add detections to fill gaps) and pruning (remove trajectories with low confidence).

B-1. Correlations between Detections

- **Appearance Correlation**:
  - Feature extraction: ResNet50 model pre-trained on ImageNet and follow its pool5 layer by two dense layers. (128-dimensional vector)
  - Appearance correlation between detections x_i and x_j: $w_{ij} = \frac{t_a - d(x_i, x_j)}{t_a}$, where $t_a = \frac{1}{2}(\mu_p + \mu_n)$, and μ_p, μ_n are the means of the positive and negative distances of all training pairs.
- **Motion Correlation**:
  - Linear motion model for prediction.
  - Forward-backward prediction error: $e_m = e_f + e_b$.
  - Motion correlation: $w_m = \alpha(t_m - e_m)$.

B-2. Optimization

- **Correlation matrix**: $W = (W_a + W_m) \odot D$, where $d_{ij} \in D, d_{ij} = e^{-\beta \Delta t_{ij}}$.
  - Decay correlations as the time lag between observations increases.
- **Correlation clustering** in a weighted graph $G = (V, E, W)$:
  - $X^* = arg \max_{\{x_{ij}\}} \sum_{(i,j) \in E} w_{ij} x_{ij}$
  - Subject to: $x_{ij} + x_{jk} \leq 1 + x_{ik}, \forall i, j, k \in V$.

B-3. Multi-Level Reasoning

- **The first level**: compute one-second long tracklets.
- **The second level**: associate tracklets into single-camera trajectories.
- **The third level**: associate single-camera trajectories into multi-camera identities.

B-4. Data Augmentation

- **For detector localization errors**: crops and horizontal flips.
- **For illumination invariance**: contrast normalization.
- **For resolution invariance**: Gaussian blur.
- **For additional viewpoint/pose invariance**: perspective transformation.
- **For occlusion**: additionally hide small rectangular image patches.

Benchmarks

- DukeMTMC:
  - 2.8k identities
  - 8 cameras
  - 1080p 60fps 85min
- DukeMTMC-reID:
  - Subset of the DukeMTMC
  - 1404 identities (appearing in more than two cameras) + 408 identities (appearing in one camera, for distraction)
- Market-1501:
  - 1501 identities by 6 cameras

Experiments

	IDF1	IDP	IDR
BIPCC (DPM + HSV) [57]	54.98	62.67	48.97
DeepCC (OpenPose + HSV)	58.24	60.60	56.06
DeepCC (DPM + ResNet)	65.68	74.87	58.50
DeepCC (OpenPose + ResNet)	**80.26**	**83.50**	**77.25**

Table 2. Impact of improving detector and features on multi-camera performance for the validation sequence.

- Single-camera IDF1 increases from 75.0 to 85.5 by changing the detector.
- Good features are crucial for MTMC tracking.
- The good detector is most useful for improving single-camera performance.

Experiments

	Multi-Camera Easy			Multi-Camera Hard			Single-Camera Easy				Single-Camera Hard			
	IDF1	IDP	IDR	IDF1	IDP	IDR	IDF1	IDP	IDR	MOTA	IDF1	IDP	IDR	MOTA
BIPCC [57]	56.2	67.0	48.4	47.3	59.6	39.2	70.1	83.6	60.4	59.4	64.5	81.2	53.5	54.6
lx_b [45]	58.0	72.6	48.2	48.3	60.6	40.2	70.3	88.1	58.5	61.3	64.2	80.4	53.4	53.6
PT_BIPCC [49]	-	-	-	-	-	-	71.2	84.8	61.4	59.3	65.0	81.8	54.0	54.4
MTMC_CDSC [68]	60.0	68.3	53.5	50.9	63.2	42.6	77.0	87.6	68.6	70.9	65.5	81.4	54.7	59.6
MYTRACKER [72]	64.8	70.8	59.8	47.3	55.6	41.2	80.0	87.5	73.8	77.7	63.4	74.5	55.2	59.0
MTMC_ReID [79]†	78.3	82.6	74.3	67.7	**78.6**	59.4	86.3	91.2	82.0	83.6	77.6	**90.1**	68.1	69.6
DeepCC	**82.0**	**84.3**	**79.8**	**68.5**	75.8	**62.4**	**89.2**	**91.7**	**86.7**	**87.5**	**79.0**	87.4	**72.0**	**70.0**

Experiments

	Euclidean		SqEuclidean	
	mAP	rank-1	mAP	rank-1
BoW+KISSME [81]	12.17	25.13	-	-
LOMO+XQDA [46]	17.04	30.75	-	-
Baseline [82]	44.99	65.22	-	-
PAN [83]	51.51	71.59	-	-
SVDNet [64]	56.80	76.70	-	-
TriHard [35]	54.60	73.24	0.28	0.89
AWTL	**54.97**	**74.23**	**52.37**	**71.45**
TriHard (+Aug)	56.65	74.91	0.48	1.25
AWTL (+Aug)	**57.28**	**75.31**	**55.94**	**75.04**
TriHard (+Aug+HNM)	54.90	74.23	0.30	0.94
AWTL (+Aug+HNM)	**58.74**	**77.69**	**57.84**	**76.21**
DPFL (1-stream) [22]	48.90	70.10	-	-
DPFL (2-stream) [22]	60.60	79.20	-	-
AWTL (2-stream)	**63.40**	**79.80**	**63.27**	**79.08**

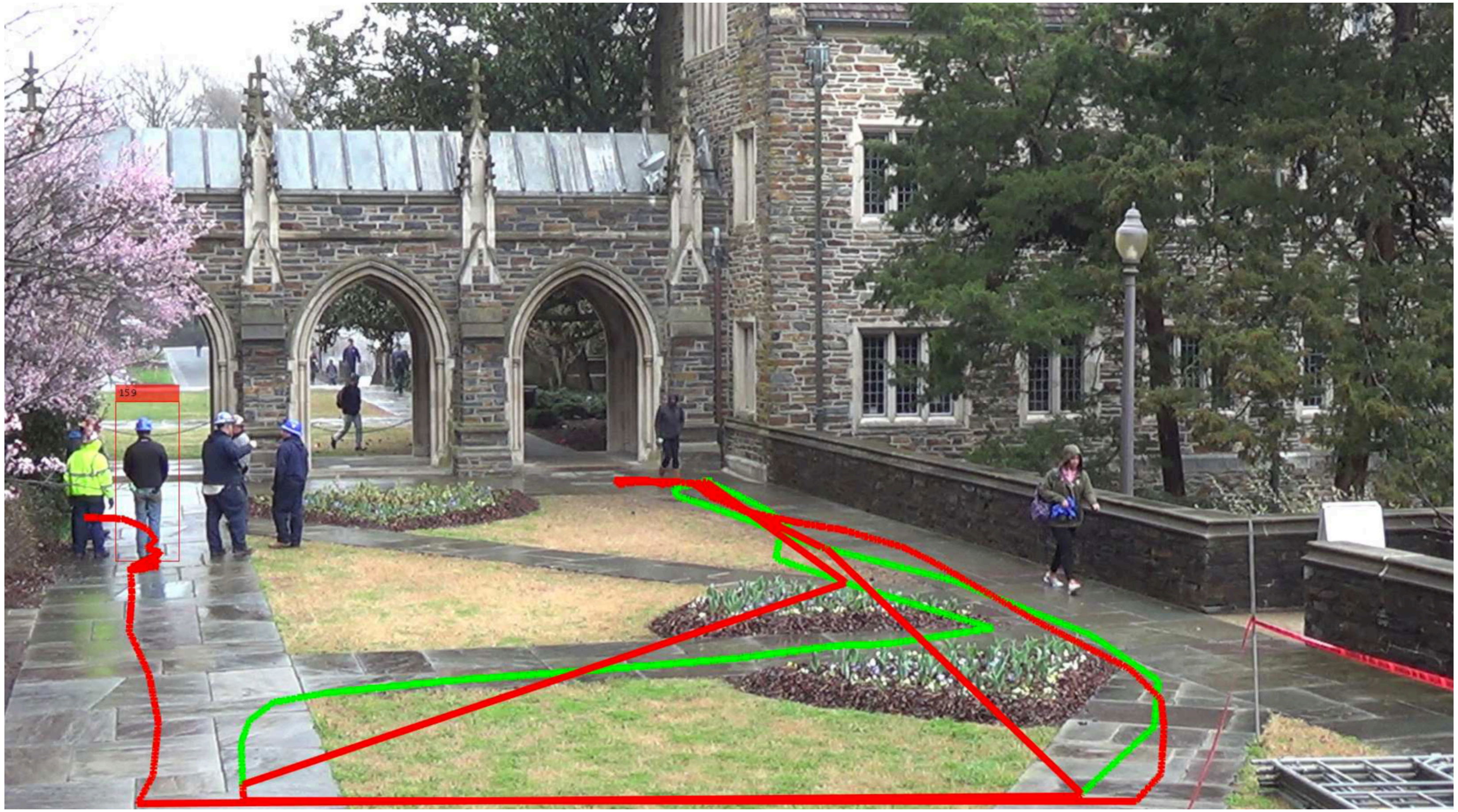
Table 4. Re-ID results on DukeMTMC-ReID

	Euclidean		SqEuclidean	
	mAP	rank-1	mAP	rank-1
DNS [76]	29.87	55.43	-	-
GatedSiamese [69]	39.55	65.88	-	-
PointSet [86]	44.27	70.72	-	-
SomaNet [5]	47.89	73.87	-	-
PAN [83]	63.35	82.81	-	-
TriHard [35]	66.63	82.99	64.47	82.01
AWTL	**68.03**	**84.20**	**65.95**	**82.16**
TriHard (+Aug)	69.57	85.14	68.92	84.12
AWTL (+Aug)	**70.83**	**86.11**	**69.64**	**84.71**
TriHard (+Aug+HNM)	71.13	86.40	0.16	0.36
AWTL (+Aug+HNM)	**71.76**	**86.94**	**70.19**	**85.39**
DPFL (1-stream) [22]	66.50	85.70	-	-
DPFL (2-stream) [22]	72.60	88.06	-	-
AWTL (2-stream)	**75.67**	**89.46**	**74.81**	**87.92**

Table 5. Re-ID results on Market-1501

- TriHard: Batch-hard triplet loss, 2017

Failure Cases – Similar Appearance



Conclusion

- Appearance feature learning is simple but useful.
- MTMC Framework is general.
- Some mistakes in the experiments.