多目标跟踪

20200625

Related to MOT

Related Work

Related to MOT

方向	研究内容	与多目标追踪区别	研究趋势	数据集
单目标追踪	第一帧给一个bounding box，然后进行跟踪	只有单个目标，目标类别不做限定，难以区分相似的类内对象	从“基于Detection的Tracking”思维中摆脱，采用多分支多通道拟合目标的位置、姿态等信息	VOTChallenge
目标检测	输入一张图片，输出目标的类别信息和位置信息	是多目标追踪问题的输入	从One stage、Two stage的Anchor based到Anchor free	VOC、COCO
基于视频的检测	输入为视频序列	没有数据关联的过程（没有轨迹ID）	两种流派 1.基于检测和跟踪的算法 2.基于光流等动态信息	ImageNet VID、YTO
Re-ID	图像检索的子问题。根据与被查询图片的相似度，降序排列	作为目标跟踪中一种非常有效的外观特征。缺少时空信息和运动信息	表征学习、度量学习、局部特征、视频序列、GAN造图、无监督、半监督。最大难点：domain的变化	Market1501、~~DukeMTMC~~、CHUK03
MTMCT	多目标多相机追踪	跨摄像头、多视频	基于Re-ID延申的方向	~~DukeMTMC~~
姿态追踪	人体姿态估计+人体姿态追踪	可以引入MOT中，人少时精度更高，但人多场景性能不好		PoseTrack

1、Unified optimization framework for localization and tracking of multiple targets with multiple cameras（2018）

- Trajectory（轨迹）：一条轨迹对应这一个目标在一个时间段内的位置序列
- Tracklet（轨迹段）：形成Trajectory过程中的轨迹片段。完整的Trajectory是由属于同一物理目标的Tracklets构成的。
- ID switch（ID切换）：又称ID sw.。对于同一个目标，由于跟踪算法误判，导致其ID发生切换的次数称为ID sw.。跟踪算法中理想的ID switch应该为0。

Tracking-by-detection

- problems：

  data association problem：linking observations from an object.

  trajectory estimation problem：predict states (location, velocity, etc.) of each object.

- methods:

  the separate approach： the tracking problem via data association becomes the main research issue.

  the unified approach：these two problems are formulated in a unified optimization framework having trajectory assignment variables as well as trajectory location variables.

papers:

1、Brau, E., Guan, J., Simek, K., Del Pero, L., Dawson, C., Barnard, K., 2013. Bayesian 3D tracking from monocular video. Proceedings of the International Conference on Computer Vision (ICCV).

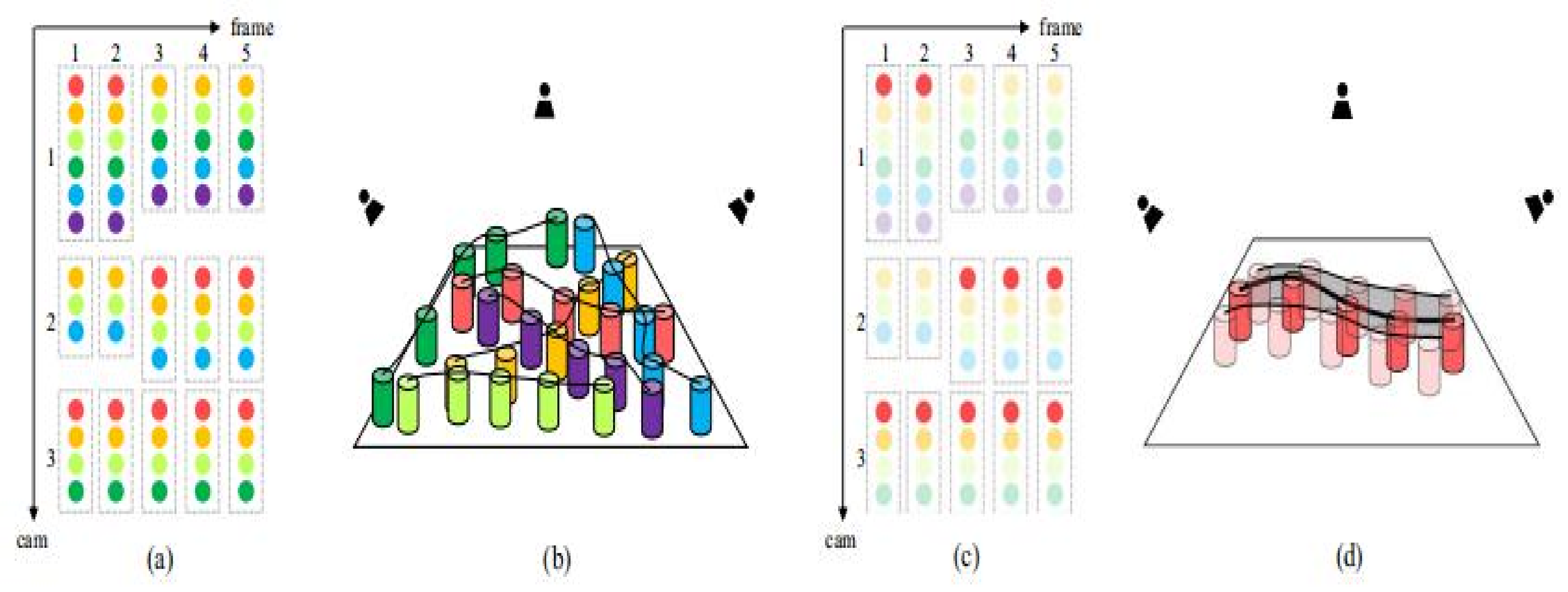
2、Andriyenko, A., Schindler, K., Roth, S., 2012. Discrete-continuous optimization for multi_x0002_target tracking. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

3、Ayazoglu, M., Li, B., Dicle, C., Sznaier, M., 2011. Dynamic subspace-based coordinated multi camera tracking. Proceedings of the International Conference on Computer Vision (ICCV).

4、Hofmann, M., Wolf, D., Rigoll, G., 2013. Hypergraphs for joint multi-view reconstruction and multi-object tracking. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

5、Byeon, M., Oh, S., Kim, K., Yoo, H.-J., Choi, J., 2015. Efficient spatio-temporal data as_x0002_sociation using multidimensional assignment for multi-camera multi-target tracking. Proceedings of British Machine Vision Conference (BMVC).

Mcmtt



Optimization variables:

1、an assignment (matrix) A

$$A = \mathbb{R}^{K\times F},\ \ [A]_{k,t} = i,\ \ i \in \mathbf{I}_{kt},\qquad \mathbf{I}_{kt} = \{0, 1, 2, \ldots, M_{kt}\}.$$

$$\begin{pmatrix} 0 & 2 & 1 & 1 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 2 & 3 & 1 & 0 \end{pmatrix}\text{:Starts at frame 2 and ends at frame 4}$$

$$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}\text{:False positive (trajectory length = 1)}$$

$$\begin{pmatrix} 1 & 2 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 2 & 2 & 3 & 3 & 2 \end{pmatrix}\text{:Missing detections at cam 2}$$

$$\begin{pmatrix} 1 & 2 & 0 & 1 & 2 \\ 0 & 0 & 0 & 3 & 2 \\ 1 & 2 & 0 & 3 & 2 \end{pmatrix}\text{:Missing detections at frame 3.}$$

2、trajectory hypothesis (vector) x

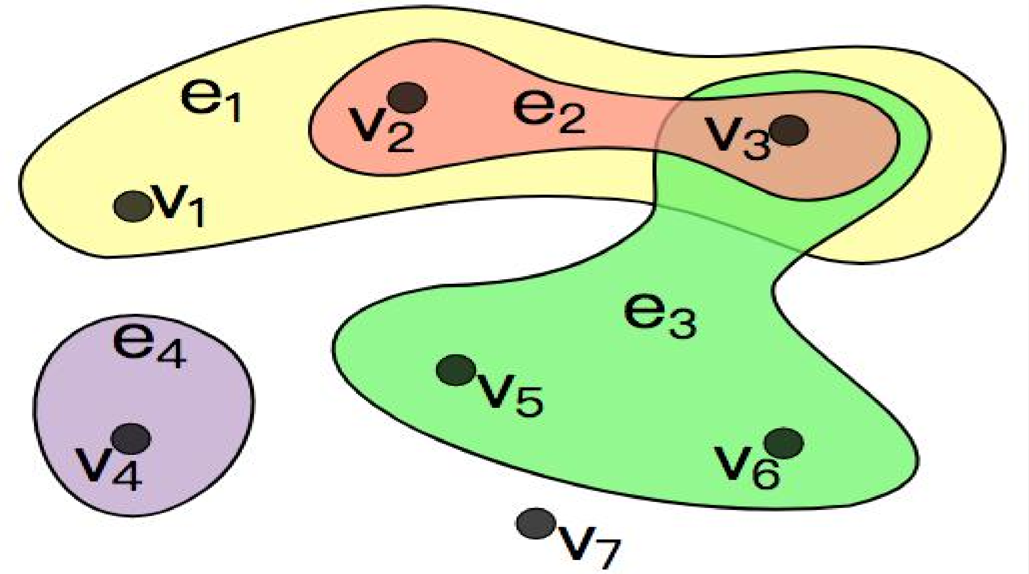
$$\mathrm{x} = (\mathrm{x}^s \ldots \mathrm{x}^e)^{\mathrm{T}} \in \mathbb{R}^{3(e-s+1)\times 1}\qquad \mathbf{x}^t = (x^t, y^t, z^t).$$

Optimization Formulation

$$G = (V, E) = (D_{11} \cup \ldots \cup D_{KF}, E),$$

$$\mathbf{d}_i^{k,t} \in \mathbf{D}_{kt}$$

- hypergraph:



An example of a hypergraph, with $X = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7\}$ and $E = \{e_1, e_2, e_3, e_4\} = \{\{v_1, v_2, v_3\}, \{v_2, v_3\}, \{v_3, v_5, v_6\}, \{v_4\}\}$.

Optimization Formulation

$$\min_{\mathbf{A},\mathcal{X}} \sum_{p=1}^{\mathrm{P}} c(A_p, \mathcal{X}) \tag{8}$$

subject to

$$[A_u]_{k,t} \neq [A_v]_{k,t}, \ \forall\, u \neq v, \quad \text{s.t. } [A_u]_{k,t}, [A_v]_{k,t} > 0, \tag{9}$$

$$\exists\, A_u \in \mathbf{A}, \quad \forall\, i \in \mathbf{I}_{kt} \setminus \{0\} \quad \text{s.t. } [A_u]_{k,t} = i, \quad k = 1, \ldots, \mathrm{K}, \ t = 1, \ldots, \mathrm{F}, \tag{10}$$

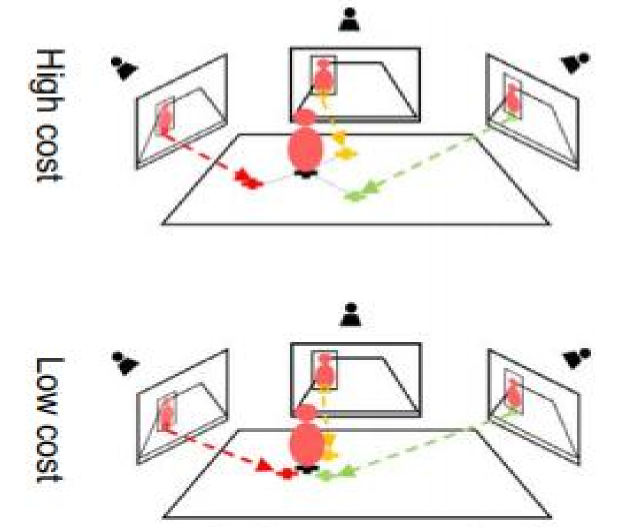
$$\mathbf{A} = \{A_1, A_2, \ldots, A_{\mathrm{P}}\}.$$

$$\left\{\mathbf{x}_{n_1}, \ldots, \mathbf{x}_{n_p}, \ldots, \mathbf{x}_{n_{\mathrm{P}}}\right\} \subset \mathcal{X},$$

Cost design

- cost

$$\tilde{c}(A, \mathbf{x}) = \lambda_{rec} \cdot c_{rec} + \lambda_{mot} \cdot c_{mot} + \lambda_{mid} \cdot c_{mid} + \lambda_{tse} \cdot c_{tse} + \lambda_{tfm} \cdot c_{tfm} + \lambda_{fpt} \cdot c_{fpt}, \quad (12)$$



Accuracy in 3D reconstruction c_{rec}

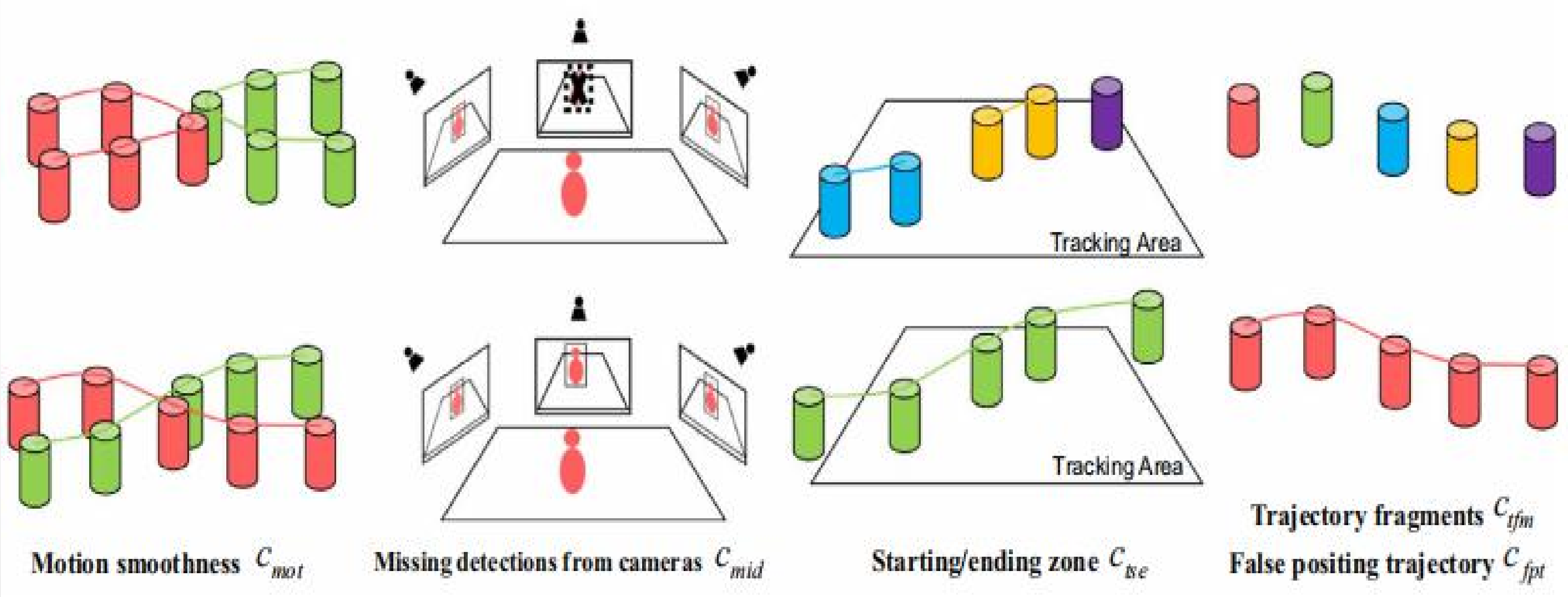
$$c_{rec}(A, \mathbf{x}) = \sum_{t=s}^{e} \sum_{k \in \mathbf{N}(\mathbf{x}^t)} \frac{\varepsilon_{rec}(A^t, \mathbf{x}^t, k)^2}{|\mathbf{N}(\mathbf{x}^t)|}, \quad (13)$$

$$\varepsilon_{rec}(A^t, \mathbf{x}^t, k) = \begin{cases} dist(\Phi^k(\mathbf{d}_i^{k,t}), \mathbf{x}^t), & \text{if } [A^t]_k = i, \ i > 0, \\ \mathrm{r}, & \text{if } [A^t]_k = 0, \ k \in \mathbf{N}(\mathbf{x}^t), \\ 0, & otherwise, \end{cases} \quad (14)$$

$$\|\mathbf{x} - \Phi^k(\mathbf{d})\| = \|(x', y', z')^{\mathrm{T}} - (\mathrm{a}z' + \mathrm{c}, \ \mathrm{b}z' + \mathrm{d}, \ z')^{\mathrm{T}}\|$$

$$\frac{x - \mathrm{c}}{\mathrm{a}} = \frac{y - \mathrm{d}}{\mathrm{b}} = z,$$

Cost design



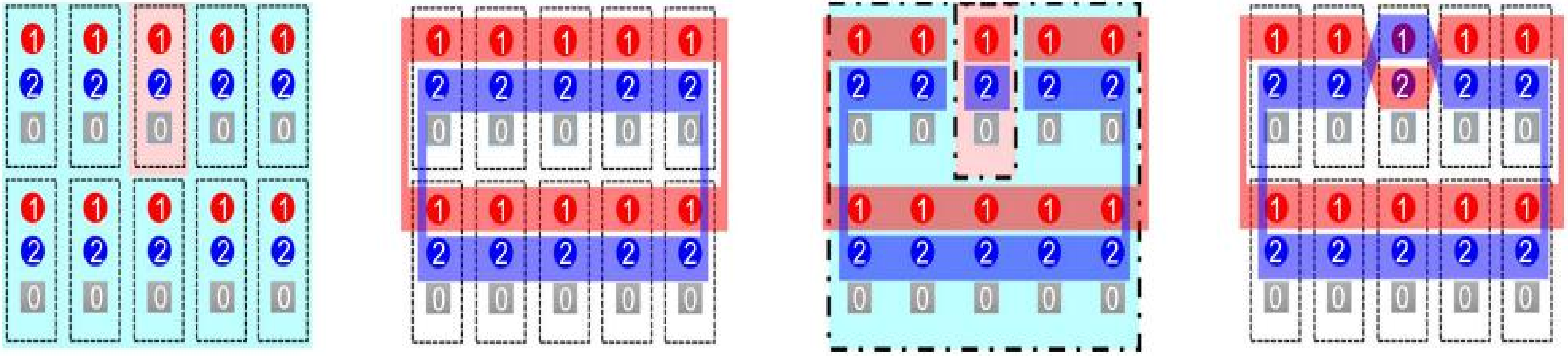
$$c_{mot}(A, \mathbf{x}) = \alpha_m \cdot \varepsilon_d + (1 - \alpha_m) \cdot \varepsilon_c,$$

$$c_{mid}(A, \mathbf{x}) = \sum_{t=s}^{e} (|\mathbf{N}(\mathbf{x}^t)| - d(A^t)),$$

$$c_{tse}(A, \mathbf{x}) = d(A^s) \cdot e(\mathbf{x}^s) + d(A^e) \cdot e(\mathbf{x}^e)$$

$$c_{tfm}(A) = d(A^s) + d(A^e).$$

Example of a splitting/re-merging



(a) $M^{\mathrm{I}}, M^{\mathrm{J}}$ (b) $\mathbf{A}^{l-1}$ (c) $\mathbf{A}^{\mathrm{I}}, \mathbf{A}^{\mathrm{J}}$ (d) $\mathbf{A}^{l}$

$$M^{\mathrm{I}} = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}, \quad M^{\mathrm{J}} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

$$\widetilde{\mathbf{A}}^{(l-1)} = \{\widetilde{A}_p\}, \quad p = 1, \ldots, \mathrm{P},$$

$$\widetilde{\mathbf{A}}^{\mathrm{I}} = \{M^{\mathrm{I}} \otimes \widetilde{A}_p\} \setminus \{O_{\mathrm{K}\times\mathrm{F}}\},$$

$$\widetilde{\mathbf{A}}^{\mathrm{J}} = \{M^{\mathrm{J}} \otimes \widetilde{A}_p\} \setminus \{O_{\mathrm{K}\times\mathrm{F}}\},$$

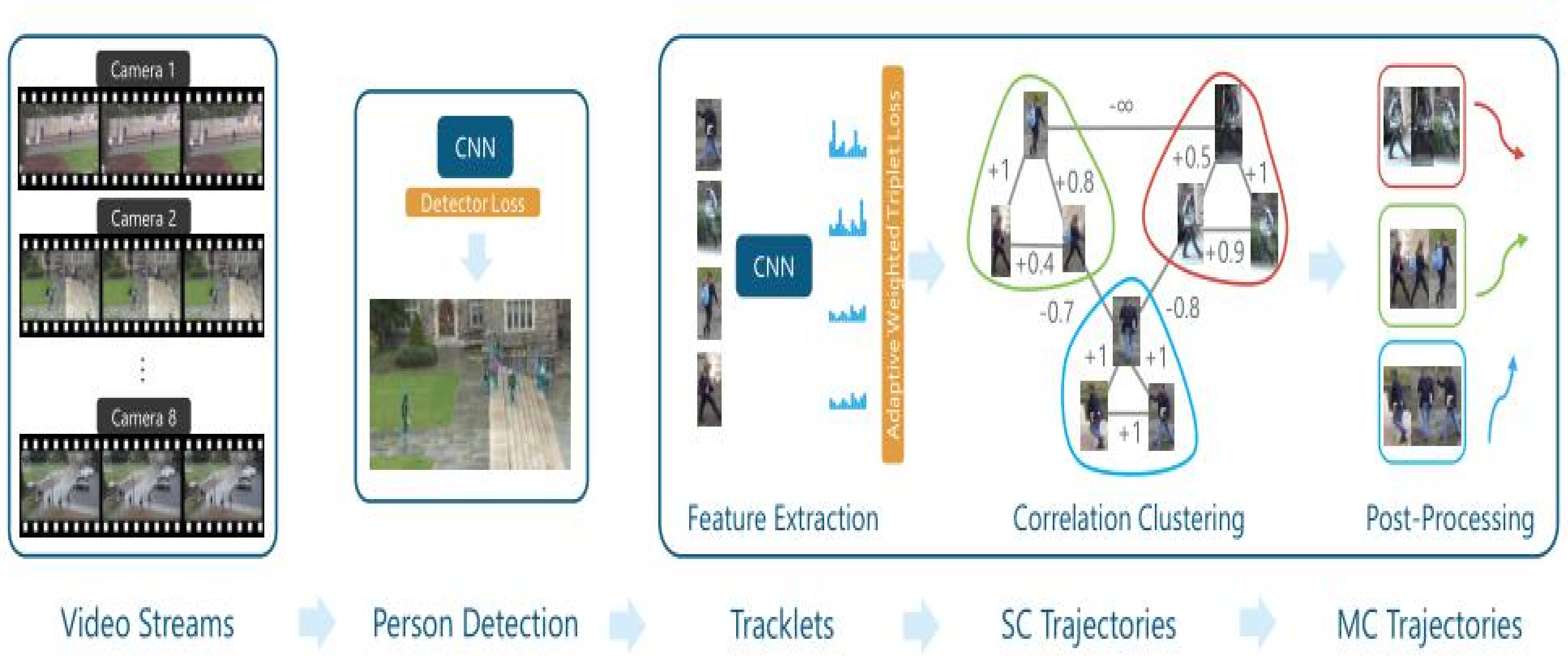
$$|\widetilde{\mathbf{A}}^{\mathrm{I}}| = \mathrm{P}^{\mathrm{I}}, \quad |\widetilde{\mathbf{A}}^{\mathrm{J}}| = \mathrm{P}^{\mathrm{J}},$$

PSN University（3648，2752）

Dataset	Method	GT	MT	PT	Rcll	Prcn	MOTA	MOTP	IDS	FM
	Proposed	10	**100%**	**0%**	**96.6%**	94.4%	**90.1%**	**85.6%**	10	**6**
PSN-Univ.	Byeon et al. (2015)	10	90%	10%	93.4%	97.4%	89.5%	82.0%	17	16
standing	Hofmann et al. (2013)	10	90%	10%	93.1%	97.2%	89.7%	81.7%	**9**	8
	Baseline	10	10%	90%	73.4%	**98.5%**	67.0%	85.0%	65	68
	Detection (Head)	–	–	–	79.4%	95.1%	–	–	–	–
	Proposed	10	**100%**	**0%**	**95.1%**	93.9%	**88.2%**	**89.1%**	**12**	**4**
PSN-Univ.	Byeon et al. (2015)	10	**100%**	**0%**	93.8%	92.1%	85.0%	87.0%	14	14
siting	Hofmann et al. (2013)	10	100%	0%	92.9%	92.4%	84.5%	88.0%	12	10
	Baseline	10	60%	40%	76.8%	**95.4%**	68.4%	**89.1%**	85	85
	Detection (Head)	–	–	–	82.8%	93.1%	–	–	–	–
	Proposed	10	**100%**	**0%**	**93.0%**	84.9%	75.6%	85.4%	21	**10**
PSN-Univ.	Byeon et al. (2015)	10	90%	10%	90.2%	88.4%	**77.2%**	84.8%	29	27
sit.&stand.	Hofmann et al. (2013)	10	80%	20%	89.1%	88.8%	77.1%	86.4%	**18**	19
	Baseline	10	60%	40%	78.1%	**95.8%**	69.8%	**87.4%**	120	131
	Detection (Head)	–	–	–	75.1%	90.5%	–	–	–	–

Results

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



loss

$$L_3 = \left[m + \sum_{x_p \in P(a)} w_p d(x_a, x_p) - \sum_{x_n \in N(a)} w_n d(x_a, x_n) \right]_+$$

$$w_p = \left[x_p == \arg \max_{x \in P(a)} d(x_a, x) \right]$$

$$w_n = \left[x_n == \arg \min_{x \in N(a)} d(x_a, x) \right]$$

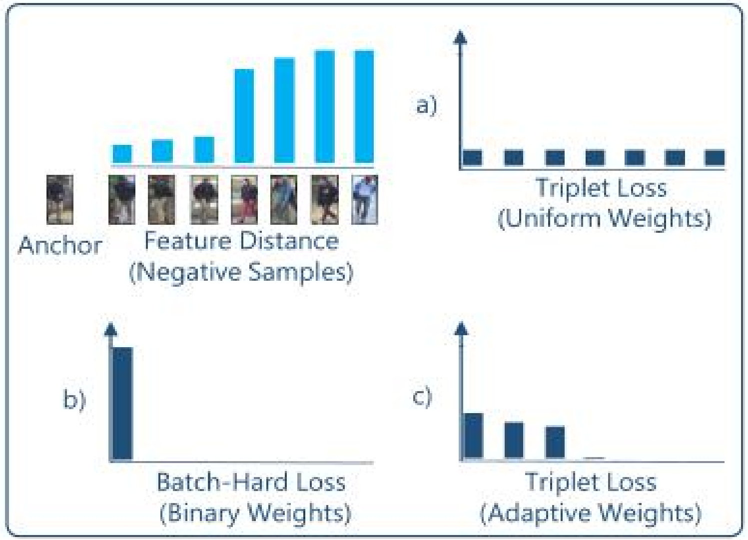


Figure 3. Triplet loss weighing schemes.

$$w_p = \frac{e^{d(x_a, x_p)}}{\sum_{x \in P(a)} e^{d(x_a, x)}}, \quad w_n = \frac{e^{-d(x_a, x_n)}}{\sum_{x \in N(a)} e^{-d(x_a, x)}}$$

sample

	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**

结束！