IWDSC 2021
2021 ICCV OCTOBER 11-17 VIRTUAL

Pedestrian Tracking through Coordinated Mining of Multiple Moving Cameras

Yanting Zhang, Qingxiang Wang

Donghua University, Shanghai, 201620

ytzhang@dhu.edu.cn

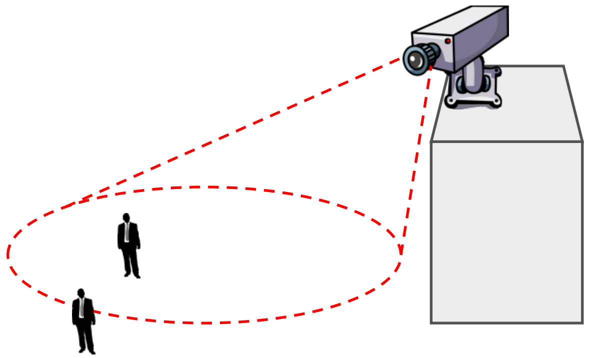
Outline

- 1 Problem Statement
- 2 Dataset
- 3 Method
- 4 Experimental Results
- 5 Conclusion

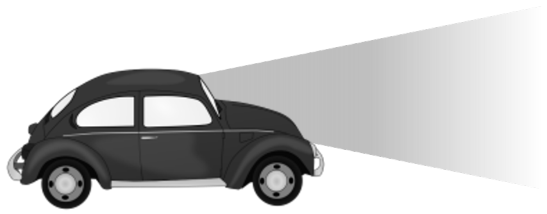
Outline

- **1 Problem Statement**
- 2 Dataset
- 3 Method
- 4 Experimental Results
- 5 Conclusion

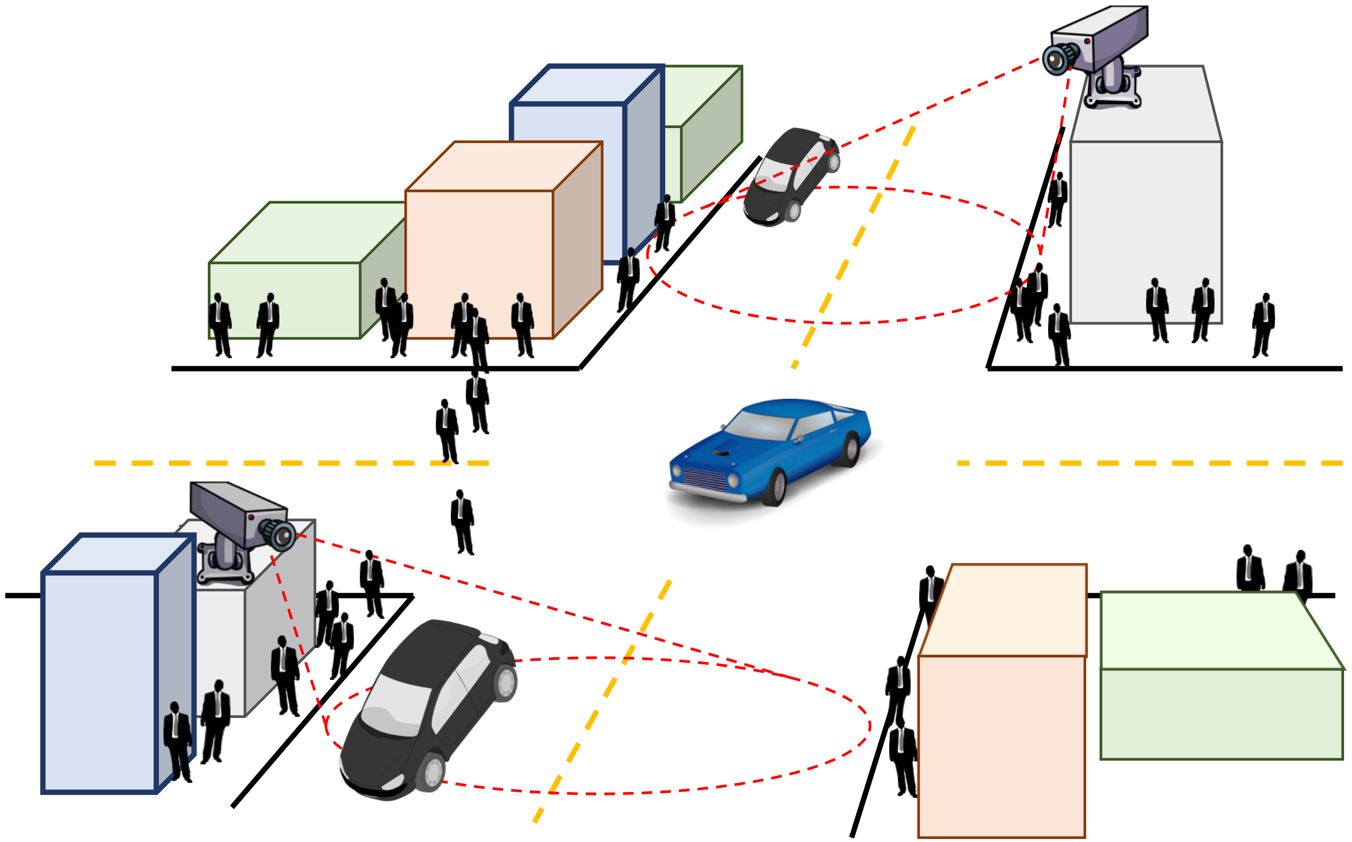
1 Problem Statements



Tracking under a single static camera

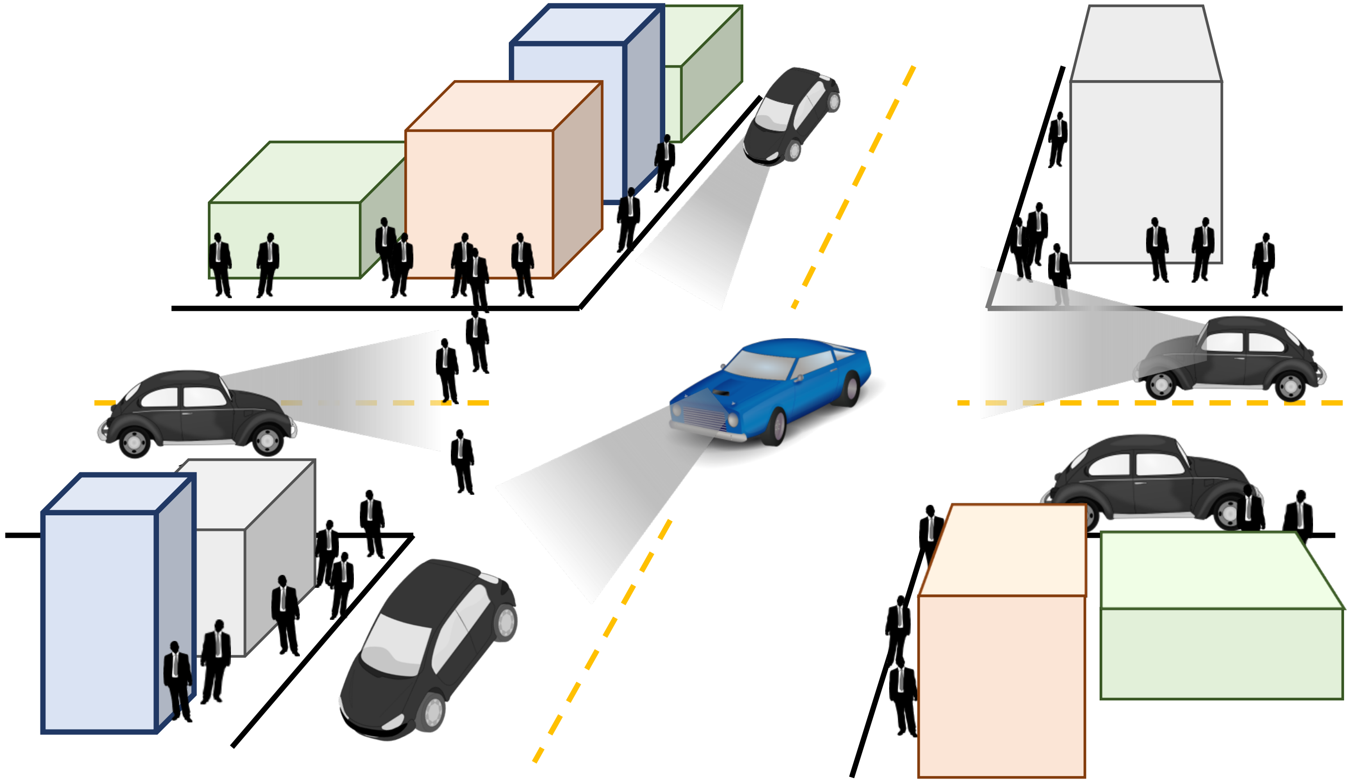


Tracking under a single moving camera



Tracking across multiple static cameras

1 Problem Statements

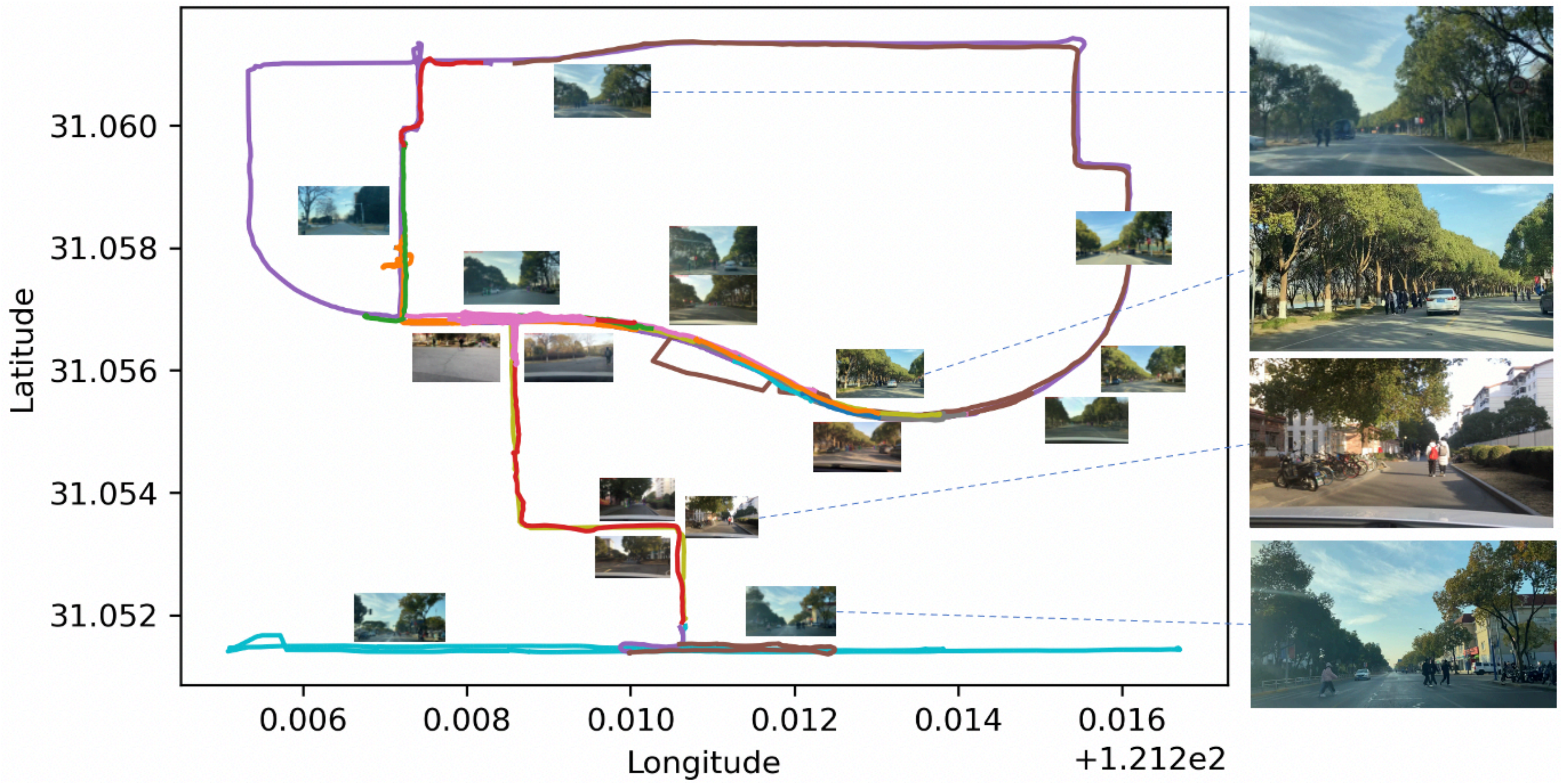


Tracking across multiple moving cameras

Outline

- 1 Problem Statement
- **2 Dataset**
- 3 Method
- 4 Experimental Results
- 5 Conclusion

2 Dataset

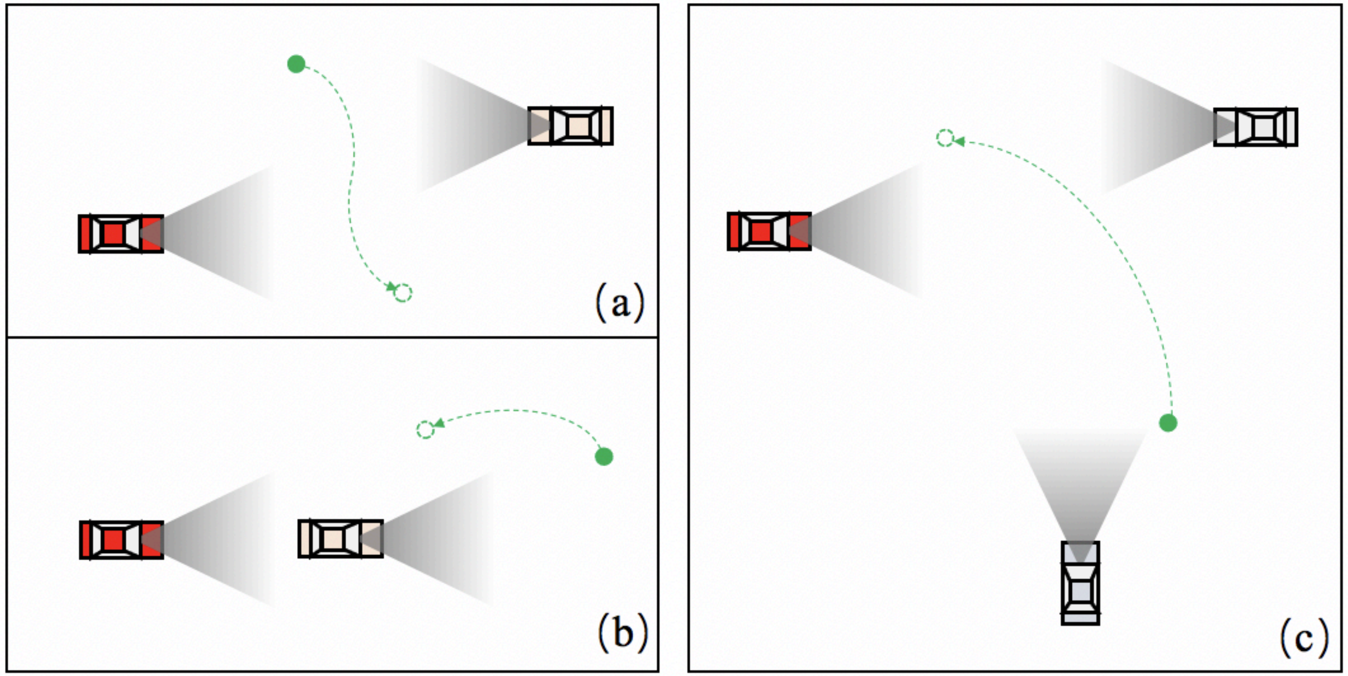


DHU-MTMMC for multi-target multi-moving camera tracking

2 Dataset

Table 1. Configurations of the devices

Device	Type	Resolution	fps
1	Iphone 6S	1920×1080	30
2	Iphone 11	1920×1080	30
3	Iphone 8	1920×1080	30
4	Oppo Reno3	1920×1080	30



Different driving cases during the data collection.
Some possible exampled pedestrian movements in green color.

2 Dataset

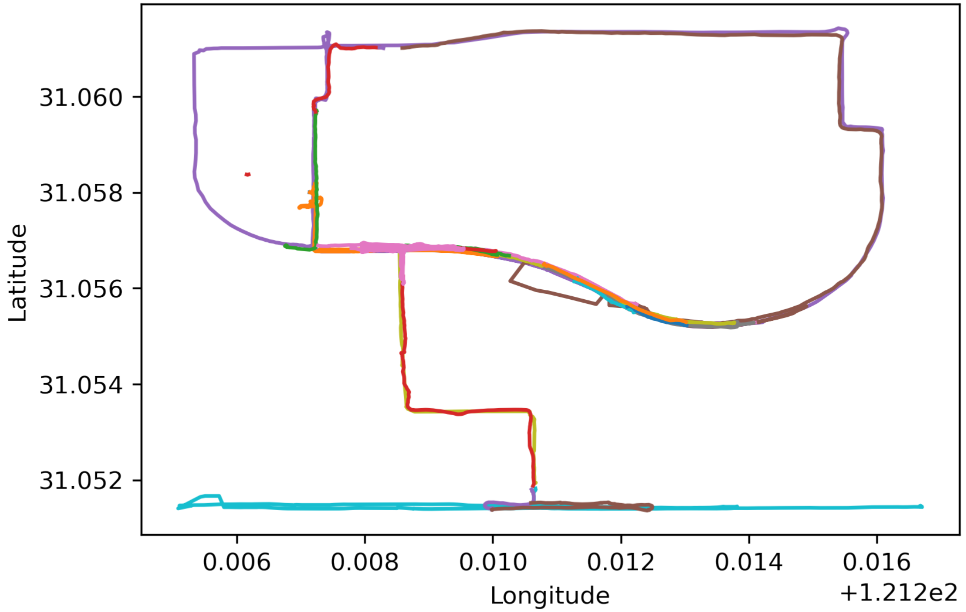


Table 2. Overview of the datasets

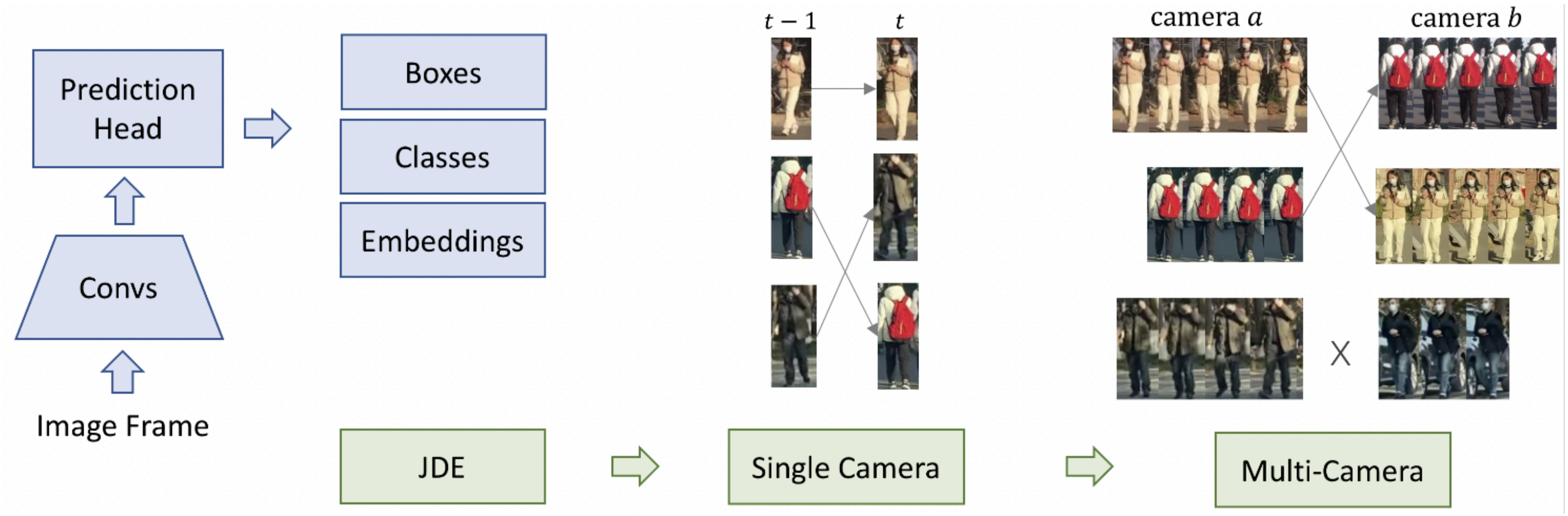
Sequence	Device	Length	Tracks	Boxes	Density
A-I	2	14s	4	171	1.9
A-II	1	52s	3	299	1.15
B-I	2	17s	23	837	7.27
B-II	1	21s	34	1041	9.91
C-I	2	9s	6	99	2.2
C-II	1	16s	16	880	11
D-I	2	84s	28	1262	3
D-II	1	86s	33	1598	3.7
E-I	2	30s	7	590	3.9
E-II	4	30s	2	148	0.98
E-III	3	25s	7	738	5.9
F-I	2	14s	5	186	2.65
F-II	4	12s	8	337	5.61
F-III	3	12s	4	185	3.08

Outline

- 1 Problem Statement
- 2 Dataset
- **3 Method**
- 4 Experimental Results
- 5 Conclusion

3 Method

MTMMC Tracking Workflow



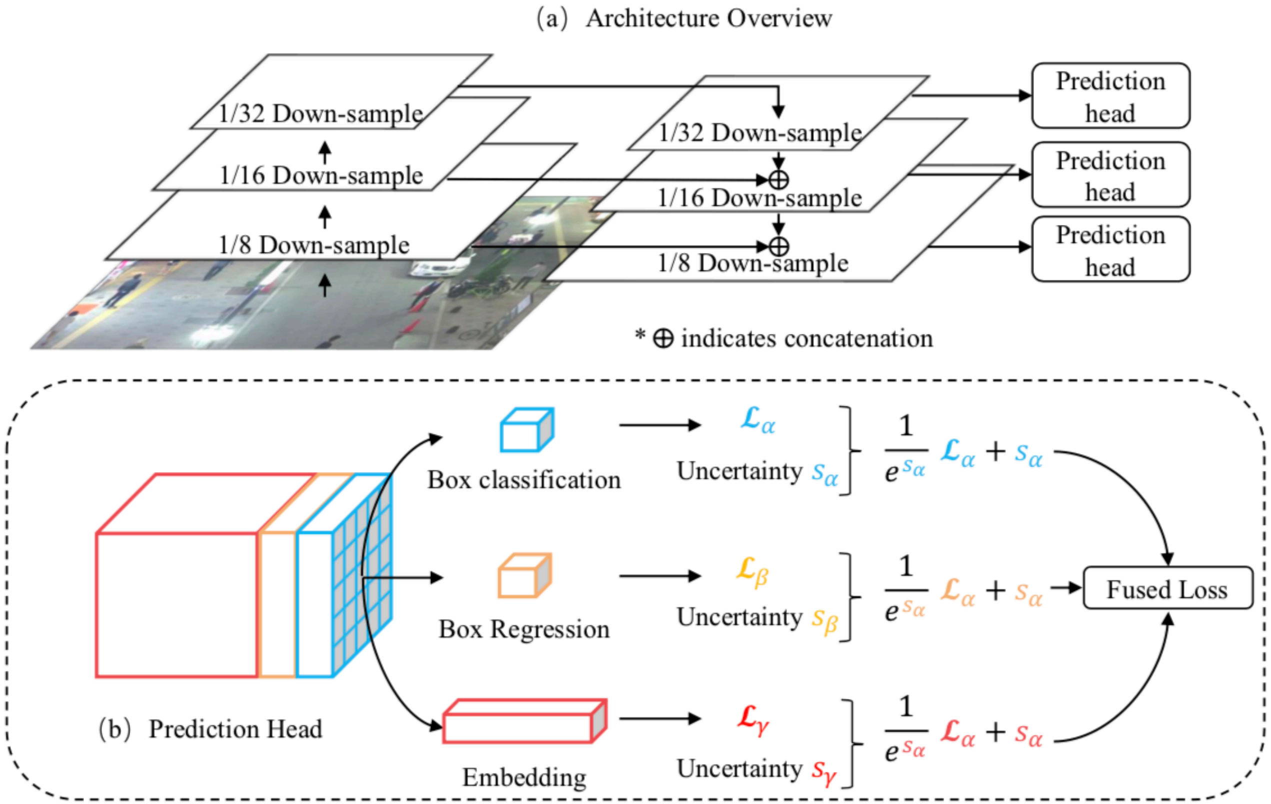
(1) Joint detection and embedding (JDE)

(2) Single camera based tracking

(3) Multi-camera based tracking

3 Method

3.1 Joint Detection and Embedding



$$L_{total} = \sum_{i}^{N} \sum_{j=\alpha,\beta,\gamma} \frac{1}{2}\left(\frac{1}{e^{s_j^i}} L_j^i + s_j^i\right)$$

Zhongdao Wang, Liang Zheng, Yixuan Liu, and Shengjin Wang. Towards real-time multi-object tracking. *arXiv preprint arXiv:1909.12605*, 2(3):4, 2019.
Alex Kendall, Yarin Gal, and Roberto Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 7482–7491, 2018.

3 Method

3.2 Single Camera based Online Association

The matching cost between the *j*-th track and the *i*-th detection:

Appearance feature

Motion information

$$C = \lambda d_1(f_j, f_i^t) + (1 - \lambda) d_2(m_j, m_i^t)$$

Euclidean distance

Mahalanobis distance

Hungarian Algorithm

The update of the embedding of a tracklet at frame t:

$$f^t = \eta f^{t-1} + (1 - \eta)\tilde{f}$$

3 Method

3.3 Multi-Camera based Tracking

j-th tracklet in camera *b*

i-th tracklet in camera *a*

$$C = d(\frac{1}{T_1}\sum_{t=1}^{T_1} f_i^t, \frac{1}{T_2}\sum_{t=1}^{T_2} f_j^t)$$

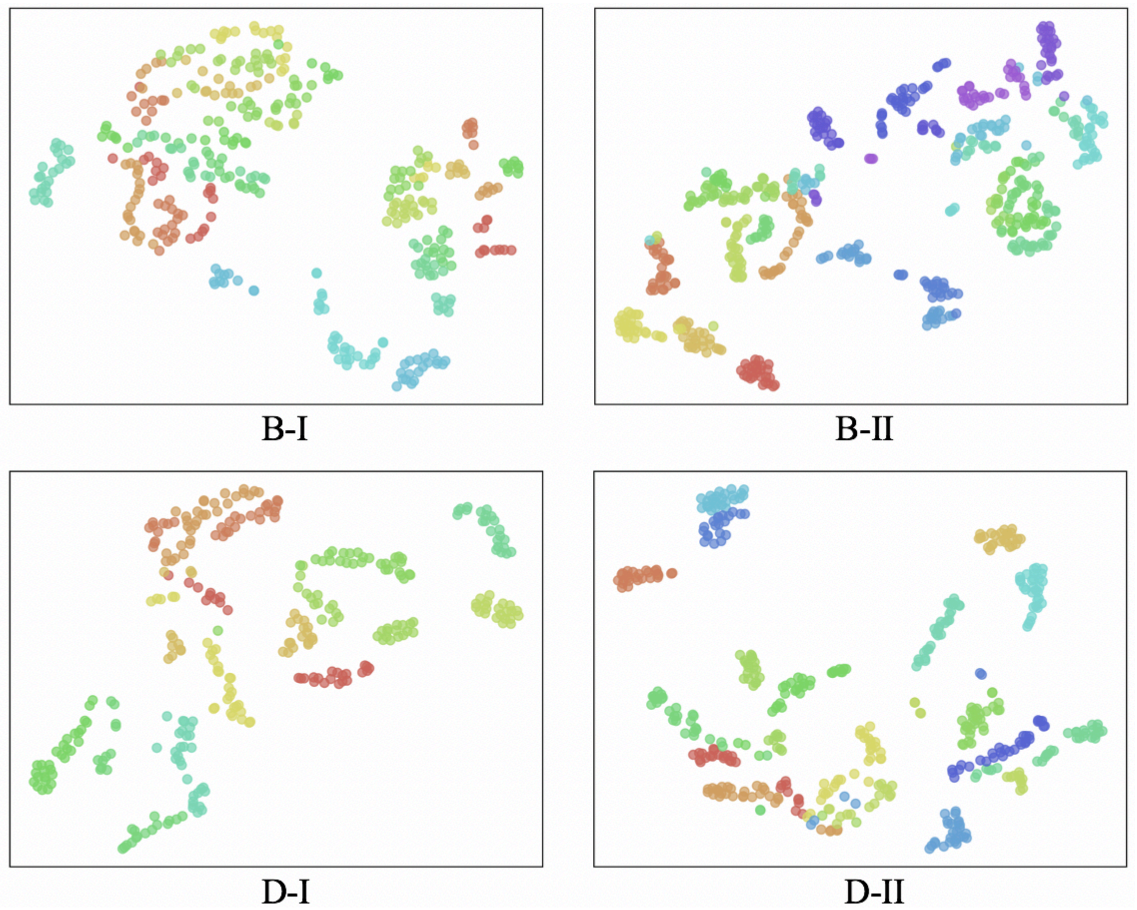
Euclidean distance

Jonker-Volgenant algorithm

Outline

- 1 Problem Statement
- 2 Method
- 3 Dataset
- **4 Experimental Results**
- 5 Conclusion

4 Experimental Results



Visualization of feature embedding for different identities using t-SNE.

4 Experimental Results

Table 3. Results of single camera tracking

Sequence	IDF1↑	IDP↑	IDR↑	MOTA↑
A-I	38.1%	57.8%	28.1%	15.2%
A-II	62.0%	58.6%	65.9%	74.6%
B-I	72.0%	76.8%	67.9%	76.1%
B-II	60.8%	65.4%	56.9%	76.8%
C-I	69.5%	87.7%	57.6%	61.6%
C-II	65.1%	70.9%	60.1%	61.8%
D-I	65.3%	72.6%	59.4%	57.6%
D-II	56.5%	64.4%	50.3%	49.5%
E-I	79.1%	94.8%	67.8%	64.7%
E-II	85.0%	80.6%	89.9%	68.2%
E-III	91.2%	98.4%	85.0%	83.6%
F-I	70.2%	77.8%	64.0%	48.4%
F-II	74.1%	77.9%	63.8%	48.1%
F-III	28.7%	32.2%	25.9%	29.2%
OVERALL	66.5%	73.1%	60.8%	62.3%

Table 5. Comparisons of single-camera tracking methods

Sequence	Deepsort		Tracktor	
	IDF1	MOTA	IDF1	MOTA
A-I	5.6%	-1.8%	15.9%	4.7%
A-II	25.8%	0.3%	35.6%	-17.1%
B-I	22.2%	23.3%	51.1%	53.8%
B-II	21.6%	15.8%	53.0%	51.7%
C-I	37.4%	19.2%	46.4%	39.4%
C-II	22.3%	23.4%	35.5%	33.2%
D-I	38.2%	23.5%	54.9%	36.5%
D-II	25.1%	10.0%	51.9%	26.0%
E-I	56.9%	56.1%	71.9%	58.0%
E-II	94.5%	89.2%	95.8%	91.9%
E-III	64.0%	75.1%	88.0%	78.6%
F-I	20.4%	11.3%	59.5%	33.9%
F-II	43.7%	25.8%	67.5%	38.0%
F-III	13.2%	18.4%	50.7%	29.7%
OVERALL	33.2%	26.3%	55.5%	41.3%

Deepsort: Nicolai Wojke et al. Simple online and realtime tracking with a deep association metric. In 2017 IEEE international conference on image processing (ICIP), pages 3645–3649. IEEE, 2017.
Tractor: Philipp Bergmann et al. Tracking without bells and whistles. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 941–951, 2019.

4 Experimental Results

Table 4. Results of multiple cameras tracking

Scene	IDF1↑	IDP↑	IDR↑
A	48.7%	51.6%	46.0%
B	59.6%	63.8%	55.9%
C	60.9%	67.2%	55.7%
D	56.5%	63.7%	50.8%
E	63.3%	69.8%	57.9%
F	48.8%	52.9%	43.2%
OVERALL	57.8%	63.6%	52.8%

Figure 6. The same pedestrian in different cameras being assigned to the same identity through multi-camera based tracking methodology. Two examples from Scene B (two cameras) and E (three cameras) are given.

4 Experimental Results

Outline

- 1 Problem Statement
- 2 Dataset
- 3 Method
- 4 Experimental Results
- **5 Conclusion**

5 Conclusion

- ✓ We propose a multi-target and multi-moving camera dataset, called “DHU-MTMMC”, which is collected for multiple object tracking across different moving cameras. It bridges the gap between the increasing need for correlating moving vehicles on the road and lacking of such a dataset in the community.
- ✓ We carry out a joint object detection and embedding extraction, and use the Hungarian algorithm for single camera based tracking. We explore to use the Jonker Volgenant algorithm for tracklets assignment across cameras. It is simple but effective for association.

THANK YOU