BRNO UNIVERSITY OF TECHNOLOGY FACULTY OF INFORMATION TECHNOLOGY

Excel@FIT 2023
http://excel.fit.vutbr.cz

Multi-Target Multi-Camera tracking

Dominik Kroupa*

Abstract
This paper focuses on Multi-Target Multi-Camera tracking (MTMC) problem, specifically on MTMC vehicle and pedestrian tracking challenges issued by *AI City Challenge*, where the aim is to track multiple objects across multiple cameras. A framework consisting of three stages to deal with the MTMC pedestrian tracking problem is proposed. The stages are single-camera tracking, tracklet refinement along with tracklet completion and inter-camera association (ICA). Best result on Test set B reached 0.2533 IDF1 score, resulting in 21st place in the challenge, making it a baseline solution for MTMC tracking without the usage of deep features in online single-camera tracker.

*xkroup12@stud.fit.vutbr.cz, *Faculty of Information Technology, Brno University of Technology*

1. Introduction

Multi-Target Multi-Camera tracking (MTMC) is a challenging task in computer vision, where the aim is to track multiple objects across multiple cameras. This helps in obtaining more information about tracked scenes than with the usage of single camera tracking. The information can be further used for crowd or traffic analysis.

This paper focuses on MTMC tracking challenges issued by *AI City Challenge* [1]. Previous years of *AI City Challenge* the task was aimed at traffic analysis, but this year the task is aimed at people movement monitoring. Compared to vehicle tracking, pedestrian movement is more chaotic, therefore some movement rules cannot be applied, such as clustering by turn direction.

Existing solutions [2, 3, 4] follow basic MTMC tracking pipeline consisting of object detection followed by feature extraction. These informations are used in single-camera tracking. Single-camera tracklets are then post-processed to reduce tracklet fragmentation and identity switches. Used object trackers employ appearance features. All these works use traffic rules in tracklet clustering to reduce search space both in single-camera (SC) and multi-camera (MC) tracking. After clustering, similarity matrix is computed for tracklet pairs that could be merged together (SC) or, in case of MC tracking, possibly belong to the same object. Best solution [2] achieved IDF1 score of 0.8095.

This paper introduces created vehicle detection dataset and proposes a framework consisting of three stages to deal with the MTMC pedestrian tracking problem. Firstly, single-camera tracklets are generated along with extracted appearance features. Secondly, refinement and completion for extracted tracklets is performed by using time constraint conditions, appearance features, and information about potential identity switches when tracklets move close to each other. Finally, Inter-Camera Association (ICA) is performed by using appearance features.

Best result on Test set B reached 0.2533 IDF1 score, resulting in 21st place in the challenge, making it a baseline solution for MTMC tracking without the usage of deep features in online single-camera tracker.

2. Created vehicle detection dataset

Inspired by [5], the provided training and validation sets from MTMC vehicle tracking dataset are used for improving detection accuracy. First, all camera videos are processed by background extraction method *labgen-of* [6]. Then, video frames from provided dataset are manually extracted based on new information, for example 100 follow-up frames containing only non-moving vehicles are skipped. Best *YOLOv5* model *yolov5x6* is used to detect vehicles in each extracted frame. In spite of model accuracy, the predicted bounding boxes (*BBoxes*) do not perfectly fit every object, so they were manually corrected. After correction of *BBoxes*, annotated vehicles were cropped and placed onto corresponding extracted backgrounds.

The created dataset contains 8106 manually annotated objects in total of 3418 images. Due to imbalance between classes (6470 cars, 1594 trucks and only 39 buses) and due to challenge task not requiring vehicle type information, all objects were merged into one *car* class. Some examples from created dataset can be seen on Figure 2.

3. Scene examples from provided pedestrian dataset

The *AI City Challenge* MTMC tracking dataset consists mainly of synthetic data, generated using the *NVIDIA Omniverse Platform*, and a small portion of real data, totaling 1491 minutes of Full-HD videos at 30 FPS from a total of 130 cameras. The videos are divided into 22 subsets, 10 for training, 5 for validation, and 7 for testing. The subsets in the dataset are captured from many different scenarios, such as from a store or warehouse. Some scene examples from provided dataset can be seen on Figure 3.

4. Created pedestrian re-identification dataset

To be compliant with the *AI City Challenge* rules that forbid the usage of external datasets (except for MS-COCO [7] and ImageNet [8]), provided ground-truth informations in training and validation sets were used to extract pedestrians from videos into images for person re-identification dataset creation. Extracted pedestrian images were manually processed and non-representative pictures were deleted.

The created dataset contains 15 808 pictures with a total of 104 identities and example of dataset can be seen on Figure 4.

5. Torchreid training results

Figures 5 and 6 visualize loss function and mAP during training for 60 epochs with *osnet_x1_0* [9] model using Torchreid [10] implementation. The final model mAP is 92.3 % with stable loss function during training.

6. YOLOv7 training results

YOLOv7 [11] detector was used to improve vehicle detection with created vehicle dataset, specifically model *yolov7*. Graphs visualising training and validation process for 100 epochs can be seen on Figure 7. Final achieved mAP is 0.744 (using main COCO [7] metric).

7. Proposed solution

This section focuses on modules in proposed solution for MTMC pedestrian tracking challenge, as shown in Figure 8.

YOLOv7 [11] and ByteTrack [12] were used for object detection and tracking respectively. In contrast to existing solutions, ByteTrack does not work with appearance features during training and only relies on detection results while using low confidence detections in association process as well, resulting in precise pedestrian tracking.

Although ByteTrack offers precise tracking with reduced computational time, it generates more fragmented tracklets when pedestrian is occluded for longer period of time. Thus, this solution relies more on post-processing, which handles identity switches and tracklet merging, using feature vectors extracted during single-camera tracking phase and spatio-temporal conditions. Both for tracklet merging and inter-camera-association (ICA), similarity matrix is constructed, containing mean cosine distances between tracklets.

8. Conclusions

Proposed solution provides a baseline solution for MTMC tracking with the usage of tracker without CNN module. Solutions to increase IDF1 metric, such as ground-plane projection, improving spatio-temporal constraints and finding best configuration between YOLOv7 model and its confidence threshold will be examined in future work.

Acknowledgements

I would like to thank my supervisor prof. Ing. Adam Herout Ph.D. for his help.

References

[1] M. Naphade, S. Wang, D. C. Anastasiu, Z. Tang, M. Chang, Y. Yao, L. Zheng, M. Shaiqur Rahman, A. Venkatachalapathy, A. Sharma, Q. Feng, V. Ablavsky, S. Sclaroff, P. Chakraborty, A. Li, S. Li, and R. Chellappa. The 6th ai city challenge. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 3346–3355. IEEE Computer Society, June 2022.

[2] Chong Liu, Yuqi Zhang, Hao Luo, Jiasheng Tang, Weihua Chen, Xianzhe Xu, Fan Wang, Hao Li, and Yi-Dong Shen. City-scale multi-camera vehicle tracking guided by crossroad zones. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops)*, pages 4124–4132, 2021.

[3] Andreas Specker, Lucas Florin, Mickael Cormier, and Jürgen Beyerer. Improving multi-target multi-camera tracking by track refinement and completion. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 3198–3208, 2022.

[4] Fei Li, Zhen Wang, Ding Nie, Shiyi Zhang, Xingqun Jiang, Xingxing Zhao, and Peng Hu. Multi-camera vehicle tracking system for ai city challenge 2022. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 3264–3272, 2022.

[5] Minghu Wu, Yeqiang Qian, Chunxiang Wang, and Ming Yang. A multi-camera vehicle tracking system based on city-scale vehicle re-id and spatial-temporal information. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 4072–4081, 2021.

[6] B. Laugraud and M. Van Droogenbroeck. Is a memoryless motion detection truly relevant for background generation with LaBGen? In *Advanced Concepts for Intelligent Vision Systems (ACIVS)*, Lecture Notes in Computer Science, Antwerp, Belgium, September 2017. Springer.

[7] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: common objects in context. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision – ECCV 2014*, pages 740–755, Cham, 2014. Springer International Publishing.

[8] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.

[9] Kaiyang Zhou, Yongxin Yang, Andrea Cavallaro, and Tao Xiang. Omni-scale feature learning for person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019.

[10] Kaiyang Zhou and Tao Xiang. Torchreid: A library for deep learning person re-identification in pytorch. *arXiv preprint arXiv:1910.10093*, 2019.

[11] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, 2022.

[12] Yifu Zhang, Peize Sun, Yi Jiang, Dongdong Yu, Fucheng Weng, Zehuan Yuan, Ping Luo, Wenyu Liu, and Xinggang Wang. Bytetrack: Multi-object tracking by associating every detection box, 2022.