Multi-Camera Multi-Object Tracking and Re-Identification

Research Project
Study program Computer Science & Engineering
Faculty of Information, Media and Electrical Engineering
Cologne University of Applied Sciences

presented by:	Luca Uckermann
matriculation number:	111 337 75
address:	Elisenstr. 29
	51149 Cologne
	luca_simon.uckermann@smail.th-koeln.de
submitted to:	Prof. Dr. Jan Salmen

Cologne, 2023-10-25

Declaration

I certify that I have written the submitted work independently. All passages taken verbatim or in spirit from the published or unpublished work of others, or from the author's own work, are marked as taken. All sources and tools used in the work are acknowledged. The work has not been submitted to any other examination authority with the same content or in substantial parts.

Place, Date

Signature

Abstract

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Contents

1 Introduction

1.1 Lorem ipsum

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

2 Structure

2.1 Citations

2.1.1 General

[1]: Current Trends in MCMOT. State of the Art. A lot of basic and advanced knowledge. Good for introduction. Analyzes 30 MCT algorithms.

[2]: General description of multi-camera tracking. State of the Art, Markov Process, graph partition theory, tracking by joint constraints.

[3]: Tracking people in multiple uncalibrated cameras. Discover spatial relationships between the camera FOVs. Tested on PETS 2001.

2.1.2 Begging

[4]: First approaches of tracking humans in multi camera network. Already done in 1999 with real-time tracking. Automatic camera switching. Bayesian classification schema.

[5]: Bayesian modality fusion to track multiple people in an indoor environment. Tries to fix already known occlusion problem.

2.1.3 Real-time

[6]: Toward Real-Time. Only multi-object tracking. Introduces JDE (Joint learning of detection and embedding). Very important paper (first real-time MOT system). Single-shot detector

[7]: Indoor scene, multiple top-view **fisheye** cameras. Possible to cover large space, less occlusion among objects. People detection and tracking. Calibrate cameras, real time (FPS of about 10) without GPU support.

[8]: Real-time distributed MCMOT system. City-scale scenario. Keeping communication and computing costs of each device low. Installs smart stations on the roadside and connects them to maintain communication. Decentralized Tracking. Kalman filter and hungarian algorithm. YoloX and DeepSORT.

[9]: FairMOT, one-shot tracker (anchor-free style). Tackles issue of object detection against re-ID. Re-ID often threated as secondary task. Reasons behind failure: anchors, feature sharing, feature dimension.

2.1.4 VOT

[10]: VOT21 Challenge Results. Considers single-camera, single-target, model-free tracking. VOT-RT2021 focuses on real-time RGB tracking. Requires predicting bounding boxes. Top two trackers: TrasT_M and STARK_RT.

[11]: VOT22 Challenge Results. Considers single.camera, single-target. VOT-RT2022 focuses on real-time RGB tracking, VOT-RTs by segmentation, VOT-RTb by bounding boxes. Goes beyond previous challenges (updating datasets). Real-time tracking at 20fps. Top trackers: MS_AOT and OSTrackSTB.

[12] VOTS23 Challenge Results. First year considering multiple-target tracking challenge. Explores short- and long-term at once. Only one challenge for all. Does not distinguish between these scenarios. Success is measured in IoU, tracking Quality mathbfQ, Accuracy, Robustness, NRE, DRE, ADQ. Dataset with challenging situations, wide range and diverse set of objects, object which are a part of other objects. Also longer videos. 77 trackers submitted, 47 valid. Most trackers applied uniform dynamic model, utilized transformers, general segmentation network SAM. Top tracker: DMAOT built upon VOT22 winner AOT. Best segmentation-based trackers outperformed all bound.box trackers.

2.1.5 Dynamic Cameras

[13]: Tracking multiple vehicles in the front view of an onboard monocular camera. Siamese network with a spatial pyramid pooling. Markov decision process. Effective for real-time long-term tracking. Hungarian algorithm, reinforcement learning.

[14]: Single-Stage Global Association Approach. Dynamic MCMOT (moving cameras in vehicle). Solves fragment-tracking issues. Not relevant for static MCMOT.

2.1.6 Person Tracking

[15]: Non-overlapping cameras. Pedestrian Tracking. Fix ID-switching issues with long-term feature extraction. OC-SORT + feature extraction.

[16]: Soccer Players. Raw detection heat maps. Google Research Football Environment. Multi camera, multi targets. Cameras have fixed positions. Do not use bounding boxes, instead raw input with heat maps. Graph Neural Network. No visual cues, such as jersey numbers. Player movement trajectories and interaction between neighborhood players.

[17]: Optical-based Pose Association (OPA). Online data association algorithm. Solve the occlusion problem. Take also human pose (see [18]) and optical flow into account, not only visual and spatial information. OpenPose, Object Keypoint Similarity, PWC-Net, Kunh-Munkras algorithm.

2.1.7 Vehicle Tracking (AI City)

[19]: Multi-camera vehicle tracking. No real-time tracking. Improve single-camera tracklets. 4th place in 2022 AI City Challenge. Track refinement module. Yolov5 pre-trained on COCO. Using GAN to generate synthetic data. Background filtering. Hierarchical clustering, zones, two rounds of clustering (tracklets separately each possible transition between cameras, akk tracks fro adjacent cameras).

[20]: Inspired [19]. First place in 2021 AI City Challenge. Yolov5 pre-trained on COCO. Most important: Introduces two step clustering (inter-zone, inter-camera clustering).

[21]: Fourth place in 2021 AI City Challenge (Track 3). Occlusion-aware tracking system. Inspired by Stadler.

[22]: Second place in 2022 AI City Challenge (Track 1). No new innovations made on first glance.

[23]: First place in 2020 AI City Challenge (Track 3). Electricity. Efficient vehicle tracking system. Aggregation loss and fast multi-target cross-camera tracking strategy. Weighted inter-class non-maximum suppression.

[24]: Graph Auto-Encoder and Self-Supervised Camera Link Model. First implementation of GAE in MTMCT. Very interesting paper. Network topology is learned automatically.

2.1.8 Re-ID, Data Association and Tracklet Matching

[25]: Locality aware appearance metric (LAAM). Intra- and inter-camera metric for re-ID. Can be applied on top of globally learned re-ID features. Improves tracking accuracy.

[18]: State-aware Re-ID. Human pose information is adopted to infer the target state including occlusion status and orientation. State-of-the-art result on Duke-MTMCT.

[26]: Proposes Mutual Information Temporal Weight Aggregated Person Re-ID Model (MI-TWA). Person re-identification. New algorithm. Not so interesting.

[27]: Dynamic Graph Model with Link Prediction. Tackles problem of data association with a dynamic graph model. Better feature representations and able to recover from lost tracks during camera transitions. Works for person and vehicle tracking for

overlapping and non-overlapping cameras. First time link prediction and dynamic graph are used together for MCMOT. Attention models.

[28]: Metadata-Aided Re-ID. Uses metadata information (car type, brand and color) for re-ID. Traffic-aware single-camera tracking. trajectory-based camera link model. Not so interesting.

[29]: Tracklet-to-Target Assignment. Solves cross-camera tracklet matching problem by TRACTA. Proposes the Restricted Non-negative Matrix Factorization (RNMF) algorithm. Estimates the number of targets in the whole network. Important paper.

2.1.9 Misc

[30]: Tracking framework for multiple interacting targets both overlapping and non-overlapping cameras, raw target trajectory with group state. SVMS, homography-based voting schema, networkflow problem, K-shortest paths algorithm.

[31]: Non-overlapping multiple cameras tracking based on similarity function. Data association method. Similarity based on color appearance and camera topology. Use superpixels for extracting color features generated by Simple Linear Iterative Clustering K-means camera topology learning.

[32]: Mathematical multi-camera tracking approach. Pre-clustering obtained from 3D geometry projections.

[33]: Utilizes information regarding spatial and temporal consistency. Reconfigurable graph model. Two step approach: Associate all objects across cameras spatially then reconfig into a temporal graph model. Matching object across different views.

2.1.10 Datasets

[34]: Created MTMCT dataset in GTA V. No privacy issues. 6 cameras over 100 minutes per camera. Largest synthetic dataset for multi camera multi person tracking.

2.2 Approaches

Single vs Multi Camera Tracking

Static vs Dynamic MCMOT

Single-Stage vs Multi-Stage Tracking

Local and Global tracklets

Cross-camera tracklet matching problem

Graph Neural Networks, Self-Attention, Transformers

Hierarchical Clustering

Tracking by detections vs One-shot (Single-shot)

Challenges: Occlusion, perspective changes, changes in lighting, changes in appearances, unknown number of targets in the whole network, unknown number of cameras in which a certain target appears.

Common Pipeline:

- Detection
- Feature Extraction
- Single Camera Tracking
- Cross Camera Association
- Multi Camera Tracking

2.3 Die Beschics

Single Object Detection (SOD)

Multi Object Detection (MOD)

Object Re-Identification (ReID)

Single Camera Tracking (SCT)

Multi Camera Tracking (MCT)

Camera Link Model (CLM)

Trajectories and Tracklets

Fisheye vs Normal Cameras

Local neighborhood: Single-camera tracking: Consecutive frames. Multi-camera tracking: Neighboring cameras.

Different cameras have different technical characteristics.

Appearance features vs Motion features

2.4 Composition

- Introduction
- Motivation
- Technical Background
- Problem Statement
- State of the Art
- Approaches
- Challenges
- Papers
- Further Research
- Conclusion

2.5 Research

mentioned in [21]:

D. Stadler and J. Beyerer. Improving multiple pedestrian tracking by track management and occlusion handling. In IEEE Conf. Comput. Vis. Pattern Recog., 2021.

Online Tracking?

Hungarian algorithm?

Multi Object vs Multi Target (definitions)

List of Figures

List of Tables

Bibliography

[1] T. I. Amosa, P. Sebastian, L. I. Izhar, *et al.*, "Multi-camera multi-object tracking: A review of current trends and future advances," *Neurocomputing*, vol. 552, p. 126 558, 2023, ISSN: 0925-2312. DOI: `https://doi.org/10.1016/j.neucom.2023.126558`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S0925231223006811`.

[2] W. Tian, "Novel aggregated solutions for robust visual tracking in traffic scenarios," Ph.D. dissertation, Karlsruher Institut für Technologie (KIT), 2019, 146 pp., ISBN: 978-3-7315-0915-8. DOI: `10.5445/KSP/1000091919`.

[3] S. Khan, O. Javed, and M. Shah, "Tracking in uncalibrated cameras with overlapping field of view," in *2nd IEEE Workshop on Performance Evaluation of Tracking and Surveillance*, IEEE Computer Society Press Los Alamitos, vol. 5, 2001.

[4] Q. Cai and J. Aggarwal, "Tracking human motion in structured environments using a distributed-camera system," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 11, pp. 1241–1247, Nov. 1999, ISSN: 1939-3539. DOI: `10.1109/34.809119`.

[5] T.-H. Chang and S. Gong, "Tracking multiple people with a multi-camera system," in *Proceedings 2001 IEEE Workshop on Multi-Object Tracking*, Jul. 2001, pp. 19–26. DOI: `10.1109/MOT.2001.937977`.

[6] Z. Wang, L. Zheng, Y. Liu, Y. Li, and S. Wang, "Towards real-time multi-object tracking," in *Computer Vision – ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds., Cham: Springer International Publishing, 2020, pp. 107–122, ISBN: 978-3-030-58621-8.

[7] T. Wang, C.-H. Liao, L.-H. Hsieh, A. W. Tsui, and H.-C. Huang, "People detection and tracking using a fisheye camera network," in *2021 International Conference on Visual Communications and Image Processing (VCIP)*, Dec. 2021, pp. 1–5. DOI: `10.1109/VCIP53242.2021.9675451`.

[8] Y. Chen, L. Ma, S. Liu, M. Liu, C. Wu, and M. Li, "A real-time distributed multi-camera multi-object tracking system," in *2022 2nd International Conference on Electrical Engineering and Mechatronics Technology (ICEEMT)*, Jul. 2022, pp. 146–149. DOI: `10.1109/ICEEMT56362.2022.9862731`.

[9] Y. Zhang, C. Wang, X. Wang, W. Zeng, and W. Liu, "Fairmot: On the fairness of detection and re-identification in multiple object tracking," *International Journal of Computer Vision*, vol. 129, no. 11, pp. 3069–3087, Nov. 1, 2021, ISSN: 1573-1405. DOI: 10.1007/s11263-021-01513-4. [Online]. Available: https://doi.org/10.1007/s11263-021-01513-4.

[10] M. Kristan, J. Matas, A. Leonardis, *et al.*, "The ninth visual object tracking vot2021 challenge results," in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, Oct. 2021, pp. 2711–2738. DOI: 10.1109/ICCVW54120.2021.00305.

[11] M. Kristan, A. Leonardis, J. Matas, *et al.*, "The tenth visual object tracking vot2022 challenge results," in *Computer Vision – ECCV 2022 Workshops*, L. Karlinsky, T. Michaeli, and K. Nishino, Eds., Cham: Springer Nature Switzerland, 2023, pp. 431–460, ISBN: 978-3-031-25085-9.

[12] M. Kristan, J. Matas, M. Danelljan, *et al.*, "The first visual object tracking segmentation vots2023 challenge results," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, Oct. 2023, pp. 1796–1818.

[13] Y. Zou, W. Zhang, W. Weng, and Z. Meng, "Multi-vehicle tracking via real-time detection probes and a markov decision process policy," *Sensors*, vol. 19, no. 6, 2019, ISSN: 1424-8220. DOI: 10.3390/s19061309. [Online]. Available: https://www.mdpi.com/1424-8220/19/6/1309.

[14] P. Nguyen, K. G. Quach, C. N. Duong, S. L. Phung, N. Le, and K. Luu, "Multi-camera multi-object tracking on the move via single-stage global association approach," 2022. arXiv: 2211.09663 [cs.CV].

[15] D.-J. Huang, P.-Y. Chou, B.-Z. Xie, and C.-H. Lin, "Multi-target multi-camera pedestrian tracking system for non-overlapping cameras," in *2023 International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan)*, Jul. 2023, pp. 629–630. DOI: 10.1109/ICCE-Taiwan58799.2023.10227006.

[16] J. Komorowski and G. Kurzejamski, "Graph-based multi-camera soccer player tracker," in *2022 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2022, pp. 1–8. DOI: 10.1109/IJCNN55064.2022.9892562.

[17] S. You, H. Yao, and C. Xu, "Multi-target multi-camera tracking with optical-based pose association," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 8, pp. 3105–3117, Aug. 2021, ISSN: 1558-2205. DOI: 10.1109/TCSVT.2020.3036467.

[18] P. Li, J. Zhang, Z. Zhu, Y. Li, L. Jiang, and G. Huang, "State-aware re-identification feature for multi-target multi-camera tracking," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2019, pp. 1506–1516. DOI: 10.1109/CVPRW.2019.00192.

[19] A. Specker, L. Florin, M. Cormier, and J. 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)*, Jun. 2022, pp. 3198–3208. DOI: 10.1109/CVPRW56347.2022.00361.

[20] C. Liu, Y. Zhang, H. Luo, *et al.*, "City-scale multi-camera vehicle tracking guided by crossroad zones," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2021, pp. 4124–4132. DOI: 10.1109/CVPRW53098.2021.00466.

[21] A. Specker, D. Stadler, L. Florin, and J. Beyerer, "An occlusion-aware multi-target multi-camera tracking system," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2021, pp. 4168–4177. DOI: 10.1109/CVPRW53098.2021.00471.

[22] F. Li, Z. Wang, D. Nie, *et al.*, "Multi-camera vehicle tracking system for ai city challenge 2022," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2022, pp. 3264–3272. DOI: 10.1109/CVPRW56347.2022.00369.

[23] Y. Qian, L. Yu, W. Liu, and A. G. Hauptmann, "Electricity: An efficient multi-camera vehicle tracking system for intelligent city," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2020, pp. 2511–2519. DOI: 10.1109/CVPRW50498.2020.00302.

[24] H.-M. Hsu, Y. Wang, J. Cai, and J.-N. Hwang, "Multi-target multi-camera tracking of vehicles by graph auto-encoder and self-supervised camera link model," in *2022 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*, Jan. 2022, pp. 489–499. DOI: 10.1109/WACVW54805.2022.00055.

[25] Y. Hou, L. Zheng, Z. Wang, and S. Wang, "Locality aware appearance metric for multi-target multi-camera tracking," 2019. arXiv: 1911.12037 [cs.CV].

[26] J. Li and Y. Piao, "Multi-target multi-camera tracking based on mutual information-temporal weight aggregation person re-identification," in *2022 IEEE 5th International Conference on Electronic Information and Communication Technology (ICEICT)*, Aug. 2022, pp. 149–151. DOI: 10.1109/ICEICT55736.2022.9908659.

[27] K. G. Quach, P. Nguyen, H. Le, *et al.*, "Dyglip: A dynamic graph model with link prediction for accurate multi-camera multiple object tracking," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2021, pp. 13 779–13 788. DOI: 10.1109/CVPR46437.2021.01357.

[28] H.-M. Hsu, J. Cai, Y. Wang, J.-N. Hwang, and K.-J. Kim, "Multi-target multi-camera tracking of vehicles using metadata-aided re-id and trajectory-based camera link model," *IEEE Transactions on Image Processing*, vol. 30, pp. 5198–5210, 2021, ISSN: 1941-0042. DOI: 10.1109/TIP.2021.3078124.

[29] Y. He, X. Wei, X. Hong, W. Shi, and Y. Gong, "Multi-target multi-camera tracking by tracklet-to-target assignment," *IEEE Transactions on Image Processing*, vol. 29, pp. 5191–5205, 2020, ISSN: 1941-0042. DOI: 10.1109/TIP.2020.2980070.

[30] S. Zhang, Y. Zhu, and A. Roy-Chowdhury, "Tracking multiple interacting targets in a camera network," *Computer Vision and Image Understanding*, vol. 134, pp. 64–73, 2015, Image Understanding for Real-world Distributed Video Networks, ISSN: 1077-3142. DOI: https://doi.org/10.1016/j.cviu.2015.01.002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1077314215000168.

[31] H. Choi and M. Jeon, "Data association for non-overlapping multi-camera multi-object tracking based on similarity function," in *2016 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)*, Oct. 2016, pp. 1–4. DOI: 10.1109/ICCE-Asia.2016.7804834.

[32] D. M. H. Nguyen, R. Henschel, B. Rosenhahn, D. Sonntag, and P. Swoboda, "Lmgp: Lifted multicut meets geometry projections for multi-camera multi-object tracking," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2022, pp. 8856–8865. DOI: 10.1109/CVPR52688.2022.00866.

[33] C.-C. Cheng, M.-X. Qiu, C.-K. Chiang, and S.-H. Lai, "Rest: A reconfigurable spatial-temporal graph model for multi-camera multi-object tracking," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 10 051–10 060. arXiv: 2308.13229 [cs.CV].

[34] P. Köhl, A. Specker, A. Schumann, and J. Beyerer, "The mta dataset for multi target multi camera pedestrian tracking by weighted distance aggregation," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2020, pp. 4489–4498. DOI: 10.1109/CVPRW50498.2020.00529.