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, 2023-10-29

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 All passages taken verbatim or in spirit from the published or unpublished work of others, or from the author's own work, are marked as taken.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**
1.1 Definition of MTMCT 1
1.2 Importance of MTMCT 1
1.3 Objective of Review 1

2 Background **2**
2.1 Challenges and Issues 2
2.1.1 Occlusion 2
2.1.2 Varying Lighting Conditions 2
2.1.3 Camera Calibration 2
2.1.4 Camera Perspective 2
2.2 Metrics and Evaluation 2
2.2.1 MOTA 2
2.2.2 MOTP 2
2.2.3 IDF1 2
2.2.4 MT 2
2.3 Components of an MTMCT System 2
2.3.1 Detection 2
2.3.2 Feature Extraction 2
2.3.3 Data Association 2
2.3.4 Tracking 2

3 Literature Review **3**
3.1 Methods 3
3.1.1 Tracking-by-Detection 3
3.1.2 Single-Object Tracking 3
3.1.3 Multi-Object Tracking 3
3.1.4 Single-Camera Tracking 3
3.1.5 Multi-Camera Tracking 3
3.1.6 Re-Identification 3
3.2 Trends 3
3.2.1 Deep Learning 3
3.2.2 Graph Neural Networks 3
3.2.3 Edge Computing 3
3.3 Strengths and Weaknesses 3

4 Online and Real-Time Tracking 4
4.1 Significance and Benefits . . . 4
4.2 Algorithms and Technologies . . . 4
4.3 Challenges and Limitations . . . 4

5 Other Methods 5

6 Discussion 6
6.1 Comparison of Methods . . . 6
6.2 Gaps and Limitations . . . 6
6.3 Future Research . . . 6
6.4 Ethical and Privacy Concerns . . . 6

7 Conclusion 7
7.1 Summary . . . 7
7.2 Future Directions . . . 7

8 Structure 8
8.1 Citations . . . 8
8.1.1 General . . . 8
8.1.2 Beginning . . . 8
8.1.3 Real-time . . . 8
8.1.4 VOT . . . 9
8.1.5 Dynamic Cameras . . . 9
8.1.6 Person Tracking . . . 10
8.1.7 Vehicle Tracking (AI City) . . . 10
8.1.8 Re-ID, Data Association and Tracklet Matching . . . 11
8.1.9 Datasets . . . 11
8.1.10 Misc . . . 12
8.2 Approaches . . . 12
8.3 Die Beschics . . . 13
8.4 Composition . . . 14
8.5 Research . . . 14
8.6 Mentioned Papers . . . 14
8.6.1 Arising Questions . . . 15

List of Figures 17

List of Tables 18

Bibliography 19

1 Introduction

1.1 Definition of MTMCT

Define MTMCT

1.2 Importance of MTMCT

Importance in video surveillance, sports analysis, traffic monitoring, and other applications

Need for online and real-time tracking

1.3 Objective of Review

Provide a comprehensive overview of the current state-of-the-art in MTMCT with a focus on online and real-time tracking methods

2 Background

2.1 Challenges and Issues

2.1.1 Occlusion

2.1.2 Varying Lighting Conditions

2.1.3 Camera Calibration

2.1.4 Camera Perspective

2.2 Metrics and Evaluation

2.2.1 MOTA

2.2.2 MOTP

2.2.3 IDF1

2.2.4 MT

2.3 Components of an MTMCT System

2.3.1 Detection

2.3.2 Feature Extraction

2.3.3 Data Association

2.3.4 Tracking

3 Literature Review

Provide an overview of the main methods used in MTMCT

3.1 Methods

3.1.1 Tracking-by-Detection

3.1.2 Single-Object Tracking

3.1.3 Multi-Object Tracking

3.1.4 Single-Camera Tracking

3.1.5 Multi-Camera Tracking

3.1.6 Re-Identification

3.2 Trends

Discuss the trends and advancements in MTMCT

3.2.1 Deep Learning

3.2.2 Graph Neural Networks

3.2.3 Edge Computing

3.3 Strengths and Weaknesses

4 Online and Real-Time Tracking

4.1 Significance and Benefits

Explain the significance of online and real-time tracking in MTMCT and its benefits over offline tracking methods.

4.2 Algorithms and Technologies

Review the latest algorithms and technologies used for online and real-time tracking, such as Siamese networks, deep reinforcement learning, and edge computing.

4.3 Challenges and Limitations

Discuss the challenges specific to online and real-time tracking, such as handling large data streams and ensuring low latency.

Evaluate the performance and limitations of existing online and real-time tracking systems.

5 Other Methods

Briefly review other methods used in MTMCT, such as offline tracking, semi-supervised tracking, and unsupervised tracking. Compare and contrast these methods with online and real-time tracking. Highlight the scenarios where these methods may be more suitable or beneficial.

6 Discussion

6.1 Comparison of Methods

Compare and contrast the different methods reviewed in the previous chapters.

6.2 Gaps and Limitations

Identify the gaps and limitations in current research.

6.3 Future Research

Suggest areas for future research.

6.4 Ethical and Privacy Concerns

Discuss the ethical and privacy concerns related to MTMCT and the need for regulations and guidelines.

7 Conclusion

7.1 Summary

Summarize the main points made in your paper.

Highlight the importance of online and real-time tracking in MTMCT and its potential to revolutionize various applications.

7.2 Future Directions

Conclude by stating the future directions and challenges for research in this area.

8 Structure

8.1 Citations

8.1.1 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.

8.1.2 Beginning

[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.

8.1.3 Real-time

[6]: Faster R-CNN. Towards Real-Time Object Detection. Region Proposal Network (RPN). RPN is trained end-to-end. Attention mechanism. 5-17 fps on GPU. Two modules (first region proposal, second detector). Sharing convolutional features.

[7]: 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

[8]: 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.

[9]: 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.

[10]: 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.

[11]: Multiple non-overlapping cameras using fast-constrained dominant set clustering (FCDSC). Three-layer hierarchical approach. Orders of magnitudes faster than existing methods. Can be used in conjunction with re-id algorithms. Good graphics in paper.

8.1.4 VOT

[12]: 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.

[13]: 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.

[14] 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.

8.1.5 Dynamic Cameras

[15]: 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.

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

8.1.6 Person Tracking

[17]: Integrating social grouping behavior for tracking pedestrians. Online learned conditional random field (CRF). Non-overlapping cameras.

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

[19]: 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.

[20]: Optical-based Pose Association (OPA). Online data association algorithm. Solve the occlusion problem. Take also human pose (see [21]) and optical flow into account, not only visual and spatial information. OpenPose, Object Keypoint Similarity, PWC-Net, Kunh-Munkras algorithm.

8.1.7 Vehicle Tracking (AI City)

[22]: 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).

[23]: Inspired [22]. First place in 2021 AI City Challenge. Yolov5 pre-trained on COCO. Most important: Introduces two step clustering (inter-zone, inter-camera clustering).

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

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

[26]: 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.

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

8.1.8 Re-ID, Data Association and Tracklet Matching

[28]: Unsupervised cross-dataset transfer learning for person re-id. Unsupervised multi-task dictionary learning (UMDL) model. Uses latent attributes. Asymmetric multi-task learning approach.

[29]: First time use of hierarchical clustering for person re-id. No online method (needs neighboring frames).

[30]: Online-learning-based person re-id. Fully unsupervised learning method. Systematically builds camera link model. Two-way GMM fitting. Multi-kernel adaptive segmentation. Multi-shot framework.

[31]: Orientation-driven person re-id (ODPR). Leverages the orientation cuest and stable torso features to learn a discriminative representation. Also estimates camera topology.Entry/Exit zones are clustered with GMM.

[32]: 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.

[21]: 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.

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

[34]: 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.

[35]: 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.

[36]: 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.

8.1.9 Datasets

[37]: Largest annotated calibrated data set for MTMC (DukeMTMC).

[38]: 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.

8.1.10 Misc

[39]: 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.

[40]: 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.

[41]: Multiple hypothesis tracking (MHT) for multi-camera tracking. Track hypothesis trees. Disjoint views. Status: tracking, searching, end-of-track. Real-time online method (15 fps). Also uses pose of person.

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

[43]: 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.

[44]: Equalized Global Graph Model-Based Approach. Improved similarity metric for single- and multiple-camera tracking. SCT and ICT in one step.

[45]: Joint person re-id and camera network topology inference. First framework which jointly solves both problems. Minimal prior knowledge about environment. Multi-shot method implemented as random-forest.

[46]: Joint learning of feature, affinity and multi-dimensional assignment (FAMNet). Online MOT. One deep-network for all three tasks. End-to-end learning.

8.2 Approaches

Single vs Multi Camera Tracking

Static vs Dynamic MCMOT

Single-Stage vs Multi-Stage Tracking

Intra camera vs Inter camera tracking

Local and Global tracklets

Cross-camera tracklet matching problem

Graph Neural Networks, Self-Attention, Transformers

Hierarchical Clustering

Gaussian Mixture Models (GMM)

Tracking by detections (Multi-shot) 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

8.3 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

Online vs Offline Tracking (Online: real-time and frame-by-frame, Offline: post-processing)

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

Different cameras have different technical characteristics.

Appearance features vs Motion features

Datasets:

- DukeMTMC
- MOTChallenge

- AI City Challenge
- PETS
- CityFlow

8.4 Composition

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

8.5 Research

8.6 Mentioned Papers

mentioned in [24]:

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

mentioned in [41]:

Ristani, E., Tomasi, C.: Tracking multiple people online and in real time. Proc. Asian Conf. Computer Vision, Singapore, 2014, pp. 444-459

Wei, S.-E., Ramakrishna, V., Kanade, T., et al.: Convolutional pose machines. Proc. IEEE Conf. Computer Vision and Pattern Recognition, Las Vegas, USA, 2016, pp. 4724-4732

mentioned in [29]:

Kuhn, H. W. 2010. The hungarian method for the as- signment problem. In 50 Years of Integer Programming.

Zhang, X.; Luo, H.; Fan, X.; Xiang, W.; Sun, Y.; Xiao, Q.; Jiang, W.; Zhang, C.; and Sun, J. 2017. Aligne-dreid: Surpassing human-level performance in person re-identification. arXiv preprint arXiv:1711.08184.

Zhong, Z.; Zheng, L.; Cao, D.; and Li, S. 2017. Re- ranking person re-identification with k-reciprocal encod- ing. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 3652-3661.

mentioned in [44]:

S. Yu, Y. Yang, and A. Hauptmann, "Harry Potters Marauders Map: Localizing and tracking multiple persons-of-interest by nonnegative dis- cretization," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2013, pp. 3714-3720.

mentioned in [17]:

X. Chen, K. Huang, and T. Tan, "Object tracking across non-overlapping views by learning inter-camera transfer models," Pattern Recognit., vol. 47, no. 3, pp. 1126-1137, 2014.

E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," IEEE Comput. Graph. Appl., vol. 21, no. 5, pp. 34-41, Sep./Oct. 2001.

M. Moussaiid, N. Perozo, S. Garnier, D. Helbing, and G. Theraulaz, The walking behaviour of pedestrian social groups and its impact on crowd dynamics

W. Ge, R. T. Collins, and R. B. Ruback, "Vision-based analysis of small groups in pedestrian crowds," IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 5, pp. 1003-1016, May 2012.

D. Helbing and P. Molnar, Social force model for pedestrian dynamics, Phys. Rev. E, vol. 51, pp. 4282-4286, May 1995.

8.6.1 Arising Questions

Online Tracking?

Hungarian algorithm?

Multi Object vs Multi Target (definitions)

Attention mechanisms

Detection Frameworks:

- YOLO
- Faster R-CNN

- R-CNN

Tracking Frameworks:

- OpenCV
- DeepSORT
- SORT
- MOTSA

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] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, ISSN: 1939-3539. DOI: 10.1109/TPAMI.2016.2577031.

[7] 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.

[8] 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.

[9] 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`.

[10] 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`.

[11] Y. T. Tesfaye, E. Zemene, A. Prati, M. Pelillo, and M. Shah, "Multi-target tracking in multiple non-overlapping cameras using Fast-Constrained dominant sets," *International Journal of Computer Vision*, vol. 127, no. 9, pp. 1303–1320, Sep. 2019.

[12] 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`.

[13] 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.

[14] 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.

[15] 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`.

[16] 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]`.

[17] X. Chen and B. Bhanu, "Integrating social grouping for multitarget tracking across cameras in a crf model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 11, pp. 2382–2394, Nov. 2017, ISSN: 1558-2205. DOI: `10.1109/TCSVT.2016.2565978`.

[18] 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`.

[19] 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`.

[20] 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`.

[21] 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`.

[22] 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`.

[23] 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`.

[24] 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`.

[25] 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`.

[26] 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`.

[27] 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`.

[28] P. Peng, T. Xiang, Y. Wang, *et al.*, "Unsupervised cross-dataset transfer learning for person re-identification," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 1306–1315. DOI: `10.1109/CVPR.2016.146`.

[29] Z. Zhang, J. Wu, X. Zhang, and C. Zhang, "Multi-target, multi-camera tracking by hierarchical clustering: Recent progress on dukemtmc project," 2017. arXiv: `1712.09531 [cs.CV]`.

[30] Y.-G. Lee, Z. Tang, and J.-N. Hwang, "Online-learning-based human tracking across non-overlapping cameras," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 10, pp. 2870–2883, Oct. 2018, ISSN: 1558-2205. DOI: `10.1109/TCSVT.2017.2707399`.

[31] N. Jiang, S. Bai, Y. Xu, C. Xing, Z. Zhou, and W. Wu, "Online inter-camera trajectory association exploiting person re-identification and camera topology," in *Proceedings of the 26th ACM International Conference on Multimedia*, ser. MM '18, Seoul, Republic of Korea: Association for Computing Machinery, 2018, pp. 1457–1465, ISBN: 9781450356657. DOI: `10.1145/3240508.3240663`. [Online]. Available: `https://doi.org/10.1145/3240508.3240663`.

[32] 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]`.

[33] 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`.

[34] 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`.

[35] 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`.

[36] 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`.

[37] E. Ristani, F. Solera, R. S. Zou, R. Cucchiara, and C. Tomasi, "Performance measures and a data set for multi-target, multi-camera tracking," in *European conference on computer vision*, Springer, 2016, pp. 17–35. arXiv: `1609.01775 [cs.CV]`.

[38] 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`.

[39] 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.

[40] 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.

[41] K. Yoon, Y.-m. Song, and M. Jeon, "Multiple hypothesis tracking algorithm for multi-target multi-camera tracking with disjoint views," *IET Image Processing*, vol. 12, no. 7, pp. 1175–1184, 2018. DOI: https://doi.org/10.1049/iet-ipr.2017.1244. eprint: https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/iet-ipr.2017.1244. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-ipr.2017.1244.

[42] 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.

[43] 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].

[44] W. Chen, L. Cao, X. Chen, and K. Huang, "An equalized global graph model-based approach for multicamera object tracking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 11, pp. 2367–2381, Nov. 2017, ISSN: 1558-2205. DOI: 10.1109/TCSVT.2016.2589619.

[45] Y.-J. Cho, S.-A. Kim, J.-H. Park, K. Lee, and K.-J. Yoon, "Joint person re-identification and camera network topology inference in multiple cameras," *Computer Vision and Image Understanding*, vol. 180, pp. 34–46, 2019, ISSN: 1077-3142. DOI: https://doi.org/10.1016/j.cviu.2019.01.003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1077314219300037.

[46] P. Chu and H. Ling, "Famnet: Joint learning of feature, affinity and multi-dimensional assignment for online multiple object tracking," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2019, pp. 6171–6180. DOI: 10.1109/ICCV.2019.00627.