MTMCT

Multi-Target Multi-Camera Tracking and Re-Identification from Detection to Tracking in Real-Time Scenarios

Luca Uckermann

Cologne University of Applied Sciences

2023-12-20

Table of Contents

Introduction

Background

Literature Review

Discussion

Conclusion

Introduction

Table of Contents

Introduction
- MTMCT
- Objectives

Background

Literature Review

Discussion

Conclusion

Introduction
MTMCT

- Object detection and tracking across multiple cameras

Introduction
MTMCT

- Object detection and tracking across multiple cameras
- Maintain global IDs of targets (re-ID)

Introduction
MTMCT

- Object detection and tracking across multiple cameras
- Maintain global IDs of targets (re-ID)
- Often people and vehicles (can be any moving object)

Introduction
MTMCT

- Object detection and tracking across multiple cameras
- Maintain global IDs of targets (re-ID)
- Often people and vehicles (can be any moving object)
- Important area in computer vision with many applications:
  - **Video Surveillance:** Monitor public places, e.g. airports, train stations, ...

Introduction
MTMCT

- Object detection and tracking across multiple cameras
- Maintain global IDs of targets (re-ID)
- Often people and vehicles (can be any moving object)
- Important area in computer vision with many applications:
  - **Video Surveillance:** Monitor public places, e.g. airports, train stations, ...
  - **Traffic Monitoring:** Monitor traffic, e.g. highways, intersections, parking lots, ...

Introduction
MTMCT

- Object detection and tracking across multiple cameras
- Maintain global IDs of targets (re-ID)
- Often people and vehicles (can be any moving object)
- Important area in computer vision with many applications:
  - **Video Surveillance:** Monitor public places, e.g. airports, train stations, ...
  - **Traffic Monitoring:** Monitor traffic, e.g. highways, intersections, parking lots, ...
  - **Sports Analysis:** Analyse sport events, e.g. football, basketball, ...

Introduction
MTMCT

- Object detection and tracking across multiple cameras
- Maintain global IDs of targets (re-ID)
- Often people and vehicles (can be any moving object)
- Important area in computer vision with many applications:
  - **Video Surveillance:** Monitor public places, e.g. airports, train stations, ...
  - **Traffic Monitoring:** Monitor traffic, e.g. highways, intersections, parking lots, ...
  - **Sports Analysis:** Analyse sport events, e.g. football, basketball, ...
  - **Crowd Management:** Analyse crowd behavior, e.g. demonstrations, concerts, ...
  - ...

Introduction

Objectives

- Literature review

Introduction

Objectives

- Literature review
- Focus on online and real-time tracking methods

Introduction

Objectives

- Literature review
- Focus on online and real-time tracking methods
- Explain background and fundamentals

Introduction

Objectives

- Literature review
- Focus on online and real-time tracking methods
- Explain background and fundamentals
- Discuss evolution (traditional $\rightarrow$ deep learning)

Introduction

Objectives

- Literature review
- Focus on online and real-time tracking methods
- Explain background and fundamentals
- Discuss evolution (traditional $\rightarrow$ deep learning)
- Identify:
  - Research gaps
  - Potential improvements
  - Future directions

Introduction

Objectives

- Literature review
- Focus on online and real-time tracking methods
- Explain background and fundamentals
- Discuss evolution (traditional $\rightarrow$ deep learning)
- Identify:
  - Research gaps
  - Potential improvements
  - Future directions
- Consider ethical and privacy aspects

Table of Contents

Introduction

Background
- Steps of MTMCT
- Intra- vs. Inter-Camera Tracking
- Tracking Process
- Fundamental Concepts
- Challenges in MTMCT
- Metrics and Evaluation

Literature Review

Discussion

Conclusion

Background

Steps of MTMCT

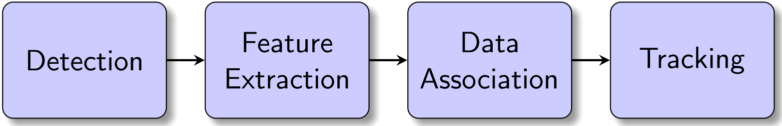

Figure: Steps of an MTMCT System

Background

Steps of MTMCT

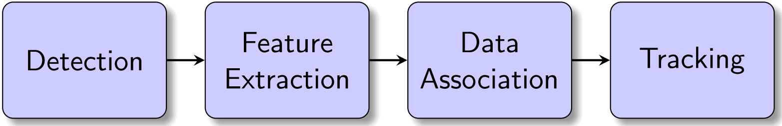

Figure: Steps of an MTMCT System

- **Detection:** Detect objects in each frame of each camera

Background

Steps of MTMCT

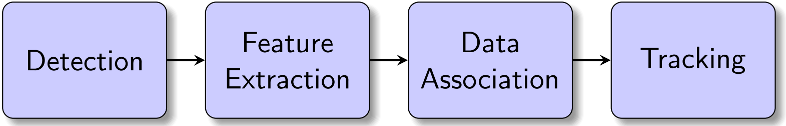

Figure: Steps of an MTMCT System

- **Detection:** Detect objects in each frame of each camera
- **Feature Extraction:** Extract features from each detection

Background

Steps of MTMCT

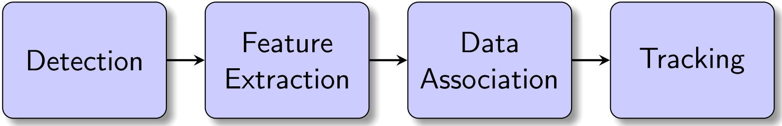

Figure: Steps of an MTMCT System

- **Detection:** Detect objects in each frame of each camera
- **Feature Extraction:** Extract features from each detection
- **Data Association:** Associate detections with existing trajectories (hierarchical)
  - **Intra-Camera:** Within each camera
  - **Inter-Camera:** Across cameras

Background

Steps of MTMCT

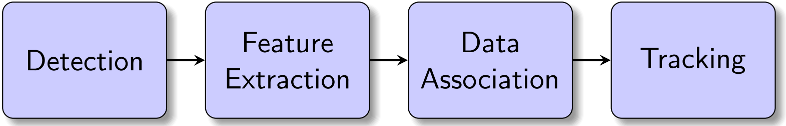

Figure: Steps of an MTMCT System

- **Detection:** Detect objects in each frame of each camera
- **Feature Extraction:** Extract features from each detection
- **Data Association:** Associate detections with existing trajectories (hierarchical)
  - **Intra-Camera:** Within each camera
  - **Inter-Camera:** Across cameras
- **Tracking:** Maintaining trajectories over time (create, update, delete)

Background

Intra- vs. Inter-Camera Tracking

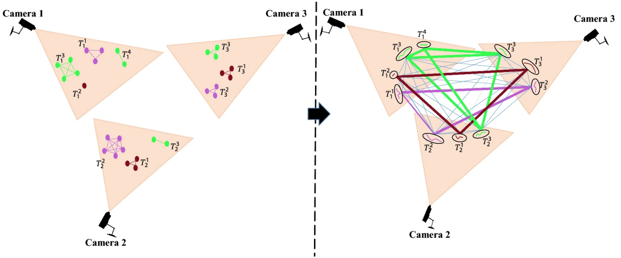

Figure: Intra- (left) and Inter-Camera (right) Tracking [1, Fig. 1]

Background

Tracking Process

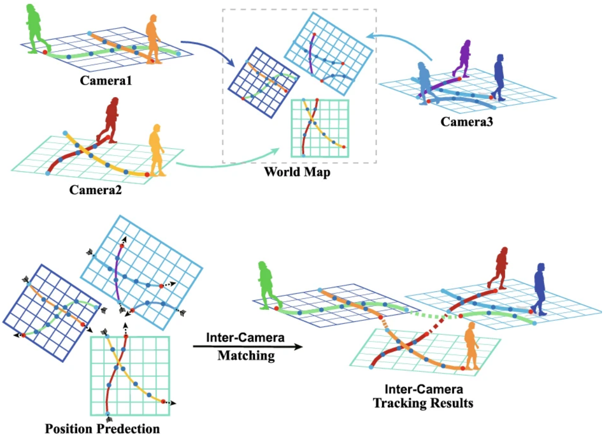

Figure: Tracking Process [2, Fig. 1]

Background

Fundamental Concepts

- Single-Target Single-Camera Tracking (STSCT)
  - Simplest form of tracking
  - Track a single target in FOV of a single camera
  - **Goal:** Maintain ID and trajectory of target

Background

Fundamental Concepts

- Single-Target Single-Camera Tracking (STSCT)
  - Simplest form of tracking
  - Track a single target in FOV of a single camera
  - **Goal:** Maintain ID and trajectory of target
- Multi-Target Single-Camera Tracking (MTSCT)
  - Builds on principles of STSCT
  - Adds complexity of multiple targets
  - **Goal:** Maintain IDs and trajectories of targets, avoid ID switches

Background

Challenges in MTMCT

- **Occlusions:** Targets can be occluded by other targets or objects

Background

Challenges in MTMCT

- **Occlusions:** Targets can be occluded by other targets or objects
- **Varying Lighting Conditions:** Lighting conditions can change over time and across cameras

Background

Challenges in MTMCT

- **Occlusions:** Targets can be occluded by other targets or objects
- **Varying Lighting Conditions:** Lighting conditions can change over time and across cameras
- **Camera Specifications:** Cameras can have different specifications (e.g. resolution, FPS, FOV, angle, ...)

Background

Challenges in MTMCT

- **Occlusions:** Targets can be occluded by other targets or objects
- **Varying Lighting Conditions:** Lighting conditions can change over time and across cameras
- **Camera Specifications:** Cameras can have different specifications (e.g. resolution, FPS, FOV, angle, ...)
- **Uncertainties (unknown number of):**
  - Targets in entire camera
  - Targets in single camera
  - Cameras a tracked target appears

Background

Metrics and Evaluation

- **MOTP:** Multiple Object Tracking Precision (accuracy of object localization) [3]
- **MOTA:** Multiple Object Tracking Accuracy (three in one: misses, false positives, ID switches) [3]

Background

Metrics and Evaluation

- **MOTP:** Multiple Object Tracking Precision (accuracy of object localization) [3]
- **MOTA:** Multiple Object Tracking Accuracy (three in one: misses, false positives, ID switches) [3]
- **IDF1:** ID F1 Score (harmonic mean of precision and recall) [4]

Background

Metrics and Evaluation

- **MOTP:** Multiple Object Tracking Precision
  (accuracy of object localization) [3]
- **MOTA:** Multiple Object Tracking Accuracy
  (three in one: misses, false positives, ID switches) [3]
- **IDF1:** ID F1 Score (harmonic mean of precision and recall) [4]
- **MT:** Mostly Tracked ($\geq$ 80% correctly tracked) [5]
- **ML:** Mostly Lost ($\leq$ 20% correctly tracked) [5]

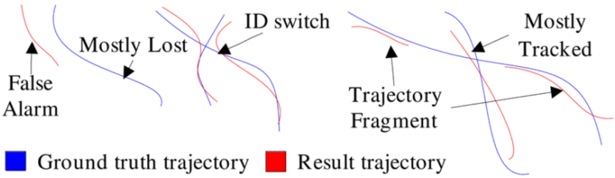

Figure: MT and ML [source image: 5, Fig. 5]

Literature Review

Table of Contents

Introduction

Background

Literature Review
- Milestones
- Tracking Paradigms
- Graph-Based
- Edge-Computing
- Online and Real-Time
- State-of-the-Art
- Honorable Mentions

Discussion

Conclusion

Literature Review

Milestones

- Detection

Literature Review

Milestones

- Detection
  - (Faster) R-CNN [6]–[8]
  - YOLO [9]
  - SSD [10]
  - ...

Literature Review

Milestones

- Detection
- Feature Extraction

Literature Review

Milestones

- Detection
- Feature Extraction
  - Scale-Invariant Feature Transform (SIFT) [11]
  - Histogram of Oriented Gradients (HOG) [12]
  - CNNs [13]
  - ...

Literature Review

Milestones

- Detection
- Feature Extraction
- Data Association

Literature Review
Milestones

- Detection
- Feature Extraction
- Data Association
  - Hungarian Algorithm [14]
  - Joint Probabilistic Data Association Filters (JPDAF) [15]
  - Probabilistic Occupancy Map (POM) [16]
  - RNNs [17]
  - ...

Literature Review

Milestones

- Detection
- Feature Extraction
- Data Association
- Tracking

Literature Review

Milestones

- Detection
- Feature Extraction
- Data Association
- Tracking
  - Kalman Filter [18]
  - Multiple Hypothesis Tracking (MHT) [19]
  - ...

Literature Review

Milestones

- Detection
- Feature Extraction
- Data Association
- Tracking
- Datasets and Challenges

Literature Review

Milestones

- Detection
- Feature Extraction
- Data Association
- Tracking
- Datasets and Challenges
  - see Table 1

Table: Overview of Datasets

Dataset	Environment	Num. of Scenarios	Num. of Cameras (Overlap)	FPS	IDs	Year	Class
PETS [20]	Outdoor	3	8 (✓)	25	—	2009	Person
MARS [21]	Mixed	Multiple	6 (✓)	—	1261	2016	Person
MOT16 [22]	Outdoor	14	1	25-30	—	2016	Person, Vehicle
DukeMTMC [4]	Outdoor	1	8 (✓)	60	2834	2016	Person
MOT17 [22]	Outdoor	14	1	25-30	—	2018	Person
Wildtrack [23]	Outdoor	Multiple	7 (✓)	2	313	2018	Person
MSMT17 [24]	Mixed	12	15 (✓)	15	4101	2018	Person
CityFlowV1 [25]	Outdoor	5	40 (✓)	10	666	2019	Vehicle
MOT20 [26]	Outdoor	8	1	25	—	2020	Person, Vehicle
CityFlowV2 [25]	Outdoor	6	46 (✓)	10	880	2021	Vehicle
MMPTRACK [27]	Indoor	5	23 (✓)	15	—	2023	Person
MEVID [28]	Mixed	17	33 (✓)	—	158	2023	Person

Literature Review

Milestones - Datasets and Challenges

Table: Overview of Datasets

Dataset	Environment	Num. of Scenarios	Num. of Cameras (Overlap)	FPS	IDs	Year	Class
PETS [20]	Outdoor	3	8 (✓)	25	—	2009	Person
MARS [21]	Mixed	Multiple	6 (✓)	—	1261	2016	Person
MOT16 [22]	Outdoor	14	1	25-30	—	2016	Person, Vehicle
DukeMTMC [4]	Outdoor	1	8 (✓)	60	2834	2016	Person
MOT17 [22]	Outdoor	14	1	25-30	—	2018	Person
Wildtrack [23]	Outdoor	Multiple	7 (✓)	2	313	2018	Person
MSMT17 [24]	Mixed	12	15 (✓)	15	4101	2018	Person
CityFlowV1 [25]	Outdoor	5	40 (✓)	10	666	2019	Vehicle
MOT20 [26]	Outdoor	8	1	25	—	2020	Person, Vehicle
CityFlowV2 [25]	Outdoor	6	46 (✓)	10	880	2021	Vehicle
MMPTRACK [27]	Indoor	5	23 (✓)	15	—	2023	Person
MEVID [28]	Mixed	17	33 (✓)	—	158	2023	Person

Challenges:

- MOT [22], [26]
- AICity [29]
- VOT(S) [30], [31]
- ...

Literature Review

Milestones - Datasets and Challenges

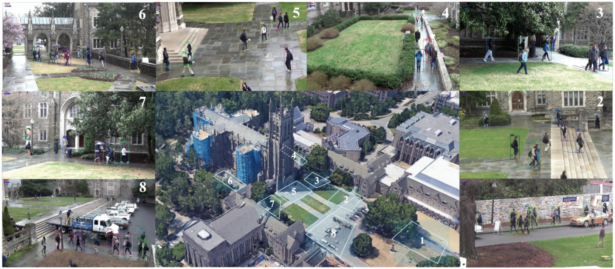

Figure: DukeMTMC [2, Fig. 2]

- **Tracking-by-Detection**

Literature Review

Tracking Paradigms

- **Tracking-by-Detection**
- Tracking-by-Regression

Literature Review

Tracking Paradigms

- **Tracking-by-Detection**
- Tracking-by-Regression
- Tracking-by-Segmentation

Literature Review

Tracking Paradigms

- **Tracking-by-Detection**
- Tracking-by-Regression
- Tracking-by-Segmentation
- Tracking-by-Attention

Literature Review

Tracking Paradigms

- **Tracking-by-Detection**
- Tracking-by-Regression
- Tracking-by-Segmentation
- Tracking-by-Attention
- **Single-Shot Approaches**

Literature Review

Tracking Paradigms

- **Tracking-by-Detection**
  - Sort [32]
  - DeepSORT [33]
- Tracking-by-Regression
- Tracking-by-Segmentation
- Tracking-by-Attention
- **Single-Shot Approaches**

Literature Review

Tracking Paradigms

- **Tracking-by-Detection**
  - Sort [32]
  - DeepSORT [33]
- Tracking-by-Regression
- Tracking-by-Segmentation
- Tracking-by-Attention
- **Single-Shot Approaches**
  - Tracktor [34]
  - Single-Shot Multi-Object Tracking (SMOT) [35]
  - Joint Detection and Embedding (JDE) [36]
  - FairMOT [37]

Literature Review

Tracking Paradigms

- **Tracking-by-Detection**
  - Sort [32]
  - DeepSORT [33]
- Tracking-by-Regression
- Tracking-by-Segmentation
- Tracking-by-Attention
- **Single-Shot Approaches**
  - Tracktor [34]
  - Single-Shot Multi-Object Tracking (SMOT) [35]
  - Joint Detection and Embedding (JDE) [36]
  - FairMOT [37]
- **Note:** Only refers to detection and intra-camera tracking (inter-camera tracking requires additional step)

Literature Review

Graph-Based

- Data association problem as graph

Literature Review

Graph-Based

- Data association problem as graph
- Nodes represent detections

Literature Review
Graph-Based

- Data association problem as graph
- Nodes represent detections
- Edges represent association costs

Literature Review

Graph-Based

- Data association problem as graph
- Nodes represent detections
- Edges represent association costs
- Recently use of Graph Neural Networks (GNNs) [38]

Literature Review

Edge-Computing

- **Advantages:**
  - Process data near source
  - Reduce latency and bandwidth
  - Improve security and privacy (data not stored)

Literature Review

Edge-Computing

- **Advantages:**
  - Process data near source
  - Reduce latency and bandwidth
  - Improve security and privacy (data not stored)
- **Drawback:** Limited resources

Literature Review

Edge-Computing

- **Advantages:**
  - Process data near source
  - Reduce latency and bandwidth
  - Improve security and privacy (data not stored)
- **Drawback:** Limited resources
- **Examples:**
  - Multi-Camera TrackingChain (MCTChain) [39]
  - Multi-Camera Tracking using Edge-Computing and Low-Power Communication [40]

Literature Review

Online and Real-Time

- "People Detection and Tracking Using a Fisheye Camera Network" [41]
  - Simulate checkout-free store
  - Fisheye cameras
  - Enter and exit store by scanning QR code
  - POM for data association
  - About 10 FPS without GPU

- “People Detection and Tracking Using a Fisheye Camera Network” [41]
  - Simulate checkout-free store
  - Fisheye cameras
  - Enter and exit store by scanning QR code
  - POM for data association
  - About 10 FPS without GPU
- Fast-Constrained Dominant Set Clustering (FCDSC) [1]
  - Graph-based approach
  - Consider only a sub-graph at each step
  - Solve intra- and inter-camera tracking simultaneously
  - About 18 FPS

Literature Review

State-of-the-Art

- Self-supervised Camera Link Model (SCLM) [42]
  - Graph Auto-Encoder (GAE) [43]
  - Zone generation algorithm
  - State-of-the-art on CityFlow 2019 and 2020

Literature Review

State-of-the-Art

- Self-supervised Camera Link Model (SCLM) [42]
  - Graph Auto-Encoder (GAE) [43]
  - Zone generation algorithm
  - State-of-the-art on CityFlow 2019 and 2020
- Lifted Multicut Meets Geometry Projections (LMGP) [44]
  - Use of POM
  - Bottom center of bounding box projection
  - State-of-the-art on Wildtrack

Literature Review

State-of-the-Art

- Self-supervised Camera Link Model (SCLM) [42]
  - Graph Auto-Encoder (GAE) [43]
  - Zone generation algorithm
  - State-of-the-art on CityFlow 2019 and 2020
- Lifted Multicut Meets Geometry Projections (LMGP) [44]
  - Use of POM
  - Bottom center of bounding box projection
  - State-of-the-art on Wildtrack
- EarlyBird [45]
  - Early fusion in bird's eye view
  - Encoder network
  - Projection onto ground plane
  - Second best on Wildtrack

Literature Review

State-of-the-Art

- Self-supervised Camera Link Model (SCLM) [42]
  - Graph Auto-Encoder (GAE) [43]
  - Zone generation algorithm
  - State-of-the-art on CityFlow 2019 and 2020
- Lifted Multicut Meets Geometry Projections (LMGP) [44]
  - Use of POM
  - Bottom center of bounding box projection
  - State-of-the-art on Wildtrack
- EarlyBird [45]
  - Early fusion in bird's eye view
  - Encoder network
  - Projection onto ground plane
  - Second best on Wildtrack
- AOT, DeAOT, DMAOT [46], [47]
  - Transformer architecture
  - Segmentation-based tracking
  - Winner of VOTS2023 challenge

Literature Review

Honorable Mentions

- Harry Potter's Marauder's Map [48]
  - Draws parallels to Marauder's Map from Harry Potter
  - Localizes and tracks people
  - Uses color information and face recognition
  - Real-world nursing home, 15 cameras

Literature Review

Honorable Mentions

- Harry Potter's Marauder's Map [48]
  - Draws parallels to Marauder's Map from Harry Potter
  - Localizes and tracks people
  - Uses color information and face recognition
  - Real-world nursing home, 15 cameras
- MTA Dataset [49]
  - MTMCT dataset
  - Virtual environment (GTA V)
  - Innovative approach
  - No privacy concerns

Introduction

Background

Literature Review

Discussion
- Summary
- Gaps and Limitations
- Future Research
- Ethical and Privacy Concerns

Conclusion

Discussion

Summary

- High complexity of MTMCT systems

Discussion

Summary

- High complexity of MTMCT systems
- Many challenges

Discussion

Summary

- High complexity of MTMCT systems
- Many challenges
- A lot of different approaches

Discussion

Gaps and Limitations

- Existing datasets and challenges focus primarily on intra-camera tracking

Discussion

Gaps and Limitations

- Existing datasets and challenges focus primarily on intra-camera tracking
- No challenge focuses solely on real-time inter-camera tracking

Discussion

Gaps and Limitations

- Existing datasets and challenges focus primarily on intra-camera tracking
- No challenge focuses solely on real-time inter-camera tracking
- Intra-camera tracking frameworks not optimized for inter-camera tracking

Discussion

Gaps and Limitations

- Existing datasets and challenges focus primarily on intra-camera tracking
- No challenge focuses solely on real-time inter-camera tracking
- Intra-camera tracking frameworks not optimized for inter-camera tracking
- Trade-off between accuracy and speed

Discussion

Future Research

- Evaluation metric for MTMCT systems
  (currently based on single-camera tracking metrics)

Discussion

Future Research

- Evaluation metric for MTMCT systems
  (currently based on single-camera tracking metrics)
- Challenge for real-time inter-camera tracking

Discussion

Future Research

- Evaluation metric for MTMCT systems
  (currently based on single-camera tracking metrics)
- Challenge for real-time inter-camera tracking
- Investigate semi- and unsupervised learning approaches
  (reduce amount of labeled data)

Discussion

Future Research

- Evaluation metric for MTMCT systems
(currently based on single-camera tracking metrics)
- Challenge for real-time inter-camera tracking
- Investigate semi- and unsupervised learning approaches
(reduce amount of labeled data)
- Processing of heterogeneous data

Discussion

Ethical and Privacy Concerns

- Balance technological progress with ethical and privacy concerns

Discussion

Ethical and Privacy Concerns

- Balance technological progress with ethical and privacy concerns
- Synthetical data generation

Discussion
Ethical and Privacy Concerns

- Balance technological progress with ethical and privacy concerns
- Synthetical data generation
- YOLO developer stopped working due to ethical concerns [50]

Discussion

Ethical and Privacy Concerns

- Balance technological progress with ethical and privacy concerns
- Synthetical data generation

- YOLO developer stopped working due to ethical concerns [50]
- DukeMTMC has been withdrawn due to privacy concerns [51]

Conclusion

Table of Contents

Introduction

Background

Literature Review

Discussion

Conclusion

- General
- Next Steps

Conclusion
General

- Intra-camera tracking well researched

Conclusion
General

- Intra-camera tracking well researched
- Inter-camera tracking still open research area with many challenges, especially in real-time scenarios

Conclusion

General

- Intra-camera tracking well researched
- Inter-camera tracking still open research area with many challenges, especially in real-time scenarios
- Need for robust data association methods across cameras

Conclusion

General

- Intra-camera tracking well researched
- Inter-camera tracking still open research area with many challenges, especially in real-time scenarios
- Need for robust data association methods across cameras
- Edge computing could lead to more comprehensive and versatile systems

Conclusion

General

- Intra-camera tracking well researched
- Inter-camera tracking still open research area with many challenges, especially in real-time scenarios
- Need for robust data association methods across cameras
- Edge computing could lead to more comprehensive and versatile systems
- Ethical and privacy concerns need to be considered in the development of MTMCT systems

Conclusion

General

- Intra-camera tracking well researched
- Inter-camera tracking still open research area with many challenges, especially in real-time scenarios
- Need for robust data association methods across cameras
- Edge computing could lead to more comprehensive and versatile systems
- Ethical and privacy concerns need to be considered in the development of MTMCT systems
- MTMCT will, as it already does, play an important role in many applications

Conclusion

Next Steps

- Implement a simple MTMCT system

Conclusion

Next Steps

- Implement a simple MTMCT system
- Follow a multi-shot tracking approach

Conclusion

Next Steps

- Implement a simple MTMCT system
- Follow a multi-shot tracking approach
- Use latest version of our tracker for intra-camera tracking

Conclusion

Next Steps

- Implement a simple MTMCT system
- Follow a multi-shot tracking approach
- Use latest version of our tracker for intra-camera tracking
- Find out which methods suit best for our use cases

Conclusion

Next Steps

- Implement a simple MTMCT system
- Follow a multi-shot tracking approach
- Use latest version of our tracker for intra-camera tracking
- Find out which methods suit best for our use cases
- Perform inter-camera tracking

Conclusion

Next Steps

- Implement a simple MTMCT system
- Follow a multi-shot tracking approach
- Use latest version of our tracker for intra-camera tracking
- Find out which methods suit best for our use cases
- Perform inter-camera tracking
- Evaluate performance and optimize system

The End

Thank you for your attention!

Questions? Comments?

[1] 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, 2019. DOI: 10.1007/s11263-019-01180-6.

[2] C. Ma, F. Yang, Y. Li, H. Jia, X. Xie, and W. Gao, "Deep trajectory post-processing and position projection for single & multiple camera multiple object tracking,", vol. 129, 2021. DOI: 10.1007/s11263-021-01527-y.

[3] K. Bernardin and R. Stiefelhagen, "Evaluating multiple object tracking performance: The clear mot metrics,", vol. 2, 2008. DOI: 10.1155/2008/246309.

[4] 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*, 2016, pp. 17–35. DOI: 10.48550/arXiv.1609.01775.

[5] B. Wu and R. Nevatia, "Tracking of multiple, partially occluded humans based on static body part detection," in

IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, 2006, pp. 951–958. DOI: 10.1109/CVPR.2006.312.

[6] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587. DOI: 10.1109/CVPR.2014.81.

[7] R. Girshick, "Fast r-cnn," in *IEEE International Conference on Computer Vision*, 2015, pp. 1440–1448. DOI: 10.1109/ICCV.2015.169.

[8] 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, 2017. DOI: 10.1109/TPAMI.2016.2577031.

[9] J. Redmon, S. K. Divvala, R. B. Girshick, and A. Farhadi, *You only look once: Unified, real-time object detection*, 2015. DOI: 10.48550/arXiv.1506.02640.

[10] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. E. Reed, C. Fu, and A. C. Berg, "SSD: single shot multibox detector," in *Computer Vision – ECCV*, Springer International Publishing, 2016, pp. 21–37. DOI: 10.1007/978-3-319-46448-0_2.

[11] D. G. Lowe, "Distinctive image features from scale-invariant keypoints,", vol. 60, no. 2, pp. 91–110, 2004. DOI: 10.1023/B:VISI.0000029664.99615.94.

[12] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, 2005, pp. 886–893. DOI: 10.1109/CVPR.2005.177.

[13] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition,", vol. 86, no. 11, pp. 2278–2324, 1998. DOI: 10.1109/5.726791.

[14] H. W. Kuhn, "The hungarian method for the assignment problem,", vol. 2, no. 1–2, pp. 83–97, 1955. DOI: 10.1002/nav.3800020109.

[15] T. Fortmann, Y. Bar-Shalom, and M. Scheffe, "Sonar tracking of multiple targets using joint probabilistic data association,", vol. 8, no. 3, pp. 173–184, 1983. DOI: 10.1109/JOE.1983.1145560.

[16] F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua, "Multicamera people tracking with a probabilistic occupancy map," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 267–282, 2008. DOI: 10.1109/TPAMI.2007.1174.

[17] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors,", vol. 323, no. 6088, pp. 533–536, 1986. DOI: 10.1038/323533a0.

[18] R. E. Kalman, "A new approach to linear filtering and prediction problems,", vol. 82, no. 1, pp. 35–45, 1960. DOI: 10.1115/1.3662552.

[19] S. Blackman, "Multiple hypothesis tracking for multiple target tracking,", vol. 19, no. 1, pp. 5–18, 2004. DOI: 10.1109/MAES.2004.1263228.

[20] J. Ferryman and A. Shahrokni, "Pets2009: Dataset and challenge," in *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, 2009, pp. 1–6. DOI: 10.1109/PETS-WINTER.2009.5399556.

[21] L. Zheng, Z. Bie, Y. Sun, J. Wang, C. Su, S. Wang, and Q. Tian, "Mars: A video benchmark for large-scale person re-identification," in *Computer Vision – ECCV*, Springer International Publishing, 2016, pp. 868–884. DOI: 10.1007/978-3-319-46466-4_52.

[22] A. Milan, L. Leal-Taixé, I. Reid, S. Roth, and K. Schindler, *MOT16: A benchmark for multi-object tracking*, 2016. DOI: 10.48550/arXiv.1603.00831.

[23] T. Chavdarova, P. Baqué, S. Bouquet, A. Maksai, C. Jose, T. Bagautdinov, L. Lettry, P. Fua, L. Van Gool, and F. Fleuret, "Wildtrack: A multi-camera hd dataset for dense

unscripted pedestrian detection," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5030–5039. DOI: 10.1109/CVPR.2018.00528.

[24] L. Wei, S. Zhang, W. Gao, and Q. Tian, "Person transfer gan to bridge domain gap for person re-identification," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 79–88. DOI: 10.1109/CVPR.2018.00016.

[25] Z. Tang, M. Naphade, M.-Y. Liu, X. Yang, S. Birchfield, S. Wang, R. Kumar, D. Anastasiu, and J.-N. Hwang, "Cityflow: A city-scale benchmark for multi-target multi-camera vehicle tracking and re-identification," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8789–8798. DOI: 10.1109/CVPR.2019.00900.

[26] P. Dendorfer, H. Rezatofighi, A. Milan, J. Shi, D. Cremers, I. Reid, S. Roth, K. Schindler, and L. Leal-Taixé, *Mot20: A benchmark for multi object tracking in crowded scenes*, 2020. DOI: 10.48550/arXiv.2003.09003.

[27] X. Han, Q. You, C. Wang, Z. Zhang, P. Chu, H. Hu, J. Wang, and Z. Liu, "Mmptrack: Large-scale densely annotated multi-camera multiple people tracking benchmark," in *IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 4849–4858. DOI: 10.1109/WACV56688.2023.00484.

[28] D. Davila, D. Du, B. Lewis, C. Funk, J. Van Pelt, R. Collins, K. Corona, M. Brown, S. McCloskey, A. Hoogs, and B. Clipp, "Mevid: Multi-view extended videos with identities for video person re-identification," in *IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 1634–1643. DOI: 10.1109/WACV56688.2023.00168.

[29] M. Naphade, S. Wang, D. C. Anastasiu, Z. Tang, M.-C. Chang, Y. Yao, L. Zheng, M. S. Rahman, M. S. Arya, A. Sharma, Q. Feng, V. Ablavsky, S. Sclaroff, P. Chakraborty, S. Prajapati, A. Li, S. Li, K. Kunadharaju, S. Jiang, and R. Chellappa, "The 7th ai city challenge," in *The IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2023. DOI: 10.48550/arXiv.2304.07500.

[30] M. Kristan, A. Leonardis, J. Matas, *et al.*, "The tenth visual object tracking vot2022 challenge results," in *Computer Vision – ECCV Workshops*, Springer Nature Switzerland, 2022, pp. 431–460. DOI: 10.1007/978-3-031-25085-9_25.

[31] 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 Workshops*, 2023, pp. 1796–1818.

[32] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *IEEE International Conference on Image Processing*, 2016, pp. 3464–3468. DOI: 10.1109/ICIP.2016.7533003.

[33] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *IEEE International Conference on Image Processing*, 2017, pp. 3645–3649. DOI: 10.1109/ICIP.2017.8296962.

[34] P. Bergmann, T. Meinhardt, and L. Leal-Taixé, "Tracking without bells and whistles," in *IEEE/CVF International*

Conference on Computer Vision, 2019, pp. 941–951. DOI: 10.1109/ICCV.2019.00103.

[35] W. Li, Y. Xiong, S. Yang, S. Deng, and W. Xia, *SMOT: single-shot multi object tracking*, 2020. DOI: 10.48550/arXiv.2010.16031.

[36] Z. Wang, L. Zheng, Y. Liu, Y. Li, and S. Wang, "Towards real-time multi-object tracking," in *Computer Vision – ECCV*, Springer International Publishing, 2020, pp. 107–122. DOI: 10.1007/978-3-030-58621-8_7.

[37] 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, 2021. DOI: 10.1007/s11263-021-01513-4.

[38] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model,", vol. 20, no. 1, pp. 61–80, 2009. DOI: 10.1109/TNN.2008.2005605.

[39] S. Wang, H. Sheng, Y. Zhang, D. Yang, J. Shen, and R. Chen, "Blockchain-empowered distributed multi-camera multi-target tracking in edge computing,", pp. 1–10, 2023. DOI: 10.1109/TII.2023.3261890.

[40] M. Nikodem, M. Słabicki, T. Surmacz, P. Mrówka, and C. Dołęga, "Multi-camera vehicle tracking using edge computing and low-power communication,", vol. 20, no. 11, 2020. DOI: 10.3390/s20113334.

[41] 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 *International Conference on Visual Communications and Image Processing*, 2021, pp. 1–5. DOI: 10.1109/VCIP53242.2021.9675451.

[42] 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 *IEEE/CVF Winter Conference on Applications of Computer Vision Workshops*, 2022, pp. 489–499. DOI: 10.1109/WACVW54805.2022.00055.

[43] T. N. Kipf and M. Welling, *Variational graph auto-encoders*, 2016. DOI: 10.48550/arXiv.1611.07308.

[44] 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 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 8856–8865. DOI: 10.1109/CVPR52688.2022.00866.

[45] T. Teepe, P. Wolters, J. Gilg, F. Herzog, and G. Rigoll, *Earlybird: Early-fusion for multi-view tracking in the bird's eye view*, 2023. DOI: 10.48550/arXiv.2310.13350.

[46] Z. Yang, Y. Wei, and Y. Yang, *Associating objects with transformers for video object segmentation*, 2021. DOI: 10.48550/arXiv.2106.02638.

[47] Z. Yang and Y. Yang, *Decoupling features in hierarchical propagation for video object segmentation*, 2022. DOI: 10.48550/arXiv.2210.09782.

[48] S.-I. Yu, Y. Yang, and A. Hauptmann, "Harry potter's marauder's map: Localizing and tracking multiple persons-of-interest by nonnegative discretization," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3714–3720. DOI: 10.1109/CVPR.2013.476.

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

[50] Synced, "Yolo creator says he stopped cv research due to ethical concerns," (2020), [Online]. Available: https://medium.com/syncedreview/yolo-creator-says-he-stopped-cv-research-due-to-ethical-concerns-b55a291ebb29 (visited on 11/01/2023).

[51] J. Harvey Adam. LaPlace, "Exposing.ai dukemtmc," (2019), [Online]. Available: https://exposing.ai/duke_mtmc/ (visited on 11/01/2023).