CSE623 - Group XYZ

Offline Track association problem

Hariohm Bhatt
AU2240085
B.Tech CSE

Meet Rathi
AU2240106
B.Tech CSE

Aditya Agrawal
AU2240153
B.Tech CSE

Harsh Panchal
AU2240160
B.Tech CSE

Jeel Kadivar
AU2140181
B.Tech CSE

Problem Statement

In multi-object tracking, due to occlusion, missed detections and abrupt motion changes in varied environments, results in the fragmented tracklets. This offline tracklet association problem focuses on the merger of these fragmented tracklets to form them into continuous trajectories

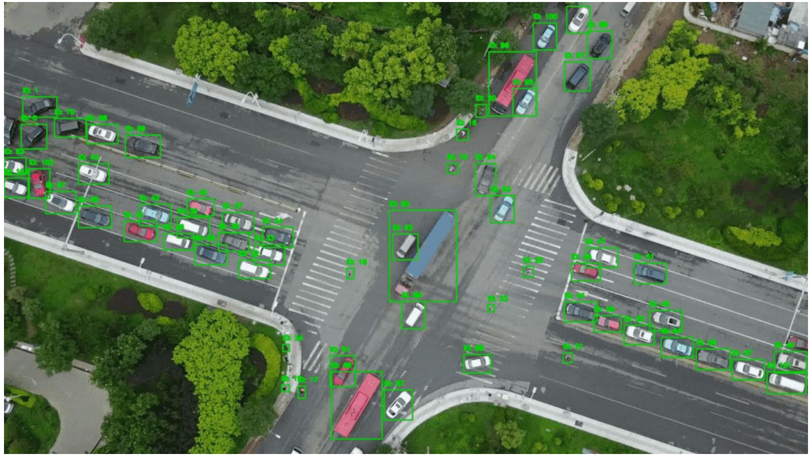
Key challenges:

- Identity Switches
- Object occlusion
- Feature consistency (like handling Abrupt motion changes)
- Missed detections

VisDrone Dataset

This dataset is captured in 14 cities of China, under urban and suburban settings under various conditions

- Video Clips: 116 video clips with 179,264 frames
- Static images: 10209 static images
- Annotations: overall 2.5 million bounding boxes are annotated



Dataset Features

- Scale variations
- Occlusion
- Motion Blur
- Viewpoint changes

Ahmedabad University

VisDrone Dataset

SAMPLE DATA FROM THE VISDRONE DATASET AFTER SELECTION.

FrameId	ObjectId	x	y	width	height	score	class	visibility
1	9	916	446	101	150	1	4	0
2	9	915	447	101	151	1	4	0
3	9	915	449	101	151	1	4	0
4	9	914	451	102	152	1	4	0
5	9	914	452	102	153	1	4	0
6	9	913	454	103	153	1	4	0
7	9	913	456	103	154	1	4	0
8	9	913	457	103	155	1	4	0
9	9	912	459	104	155	1	4	0
10	9	912	461	104	156	1	4	0

Annotations:

- **FrameId:** Unique identifier for the frames, it tracks the order of the frames
- **Bounding Box:** consists of **x, y, width, height**, which are the coordinates of the object
- **Object Id:** Unique identifier for object across the frames
- **Object Category:** class label for the object (in our case “cars”)
- **Score:** It is the confidence score of detection
- **Visibility:** indicating visibility level, which helps to detect occlusion
- **Ignored:** It is a flag to identify if the object is to be ignored

VisDrone Dataset

Used PCA:

- **Purpose:** Reduce high-dimensional feature set from annotations & images
- **Before PCA:** Included features like bounding box stats, motion, color histograms

Why PCA?

- Remove redundant/correlated features
- Improve DBSCAN & KD-Tree clustering
- Reduce computational load

Steps:

- Standardized the features (mean = 0, variance = 1)
- Applied PCA and retained principal components that explain ~95% variance
- Transformed the original feature matrix into the reduced PCA space

Result:

PCA helped in capturing the most important variance in fewer dimensions, making clustering more effective and efficient.

Methodology

Preprocessing:

- Select relevant sequences from the VisDrone dataset
- Remove unrelated frames and annotations
- Apply a temporal window (15–60 frames)
- Simulate occlusions by randomly dropping frames
- Split tracklets at large detection gaps
- Group continuous segments into tracklets
- Preserve mapping to original object IDs

Feature Extraction:

Annotation-Based Features

- **Geometry:** Mean/std of center, size, aspect ratio, area
- **Motion:** Velocity, acceleration, gaps, duration
- **Quality:** Detection score stats, class mode, visibility

Image-Based Features

- Color Histogram (HSV, reduced resolution)
- Optical Flow (mean/std of magnitude & direction between frames)

Methodology

Tracklet Merging:

Feature Matrix Construction:- Consolidate features (e.g., position, velocity, color, optical flow). Handle missing values & normalize features.

Similarity Graph Creation:-Compute pairwise Euclidean distances. Connect tracklets with distance below threshold.

DBSCAN Clustering:- Apply DBSCAN on normalized features. Groups similar tracklets, filters out noise.

Merging & Evaluation:- Merge tracklets using global & local cues.

Evaluate using Precision, Recall, F1-score.

Save Output:- Store merged results with tracklet-to-object ID mapping.

Methodology

Loss Function

Goal: Optimize tracklet merging by learning a **discriminative feature space**.

How it works:

- Operates on **pairs of tracklet embeddings**.
- **Minimizes** distance between **positive pairs** (same object).
- **Maximizes** distance (by a margin **m**) between **negative pairs** (different objects).

Outcome: Encourages clustering of similar tracklets and separation of dissimilar ones, boosting merge reliability.

Loss Formula:

$$\mathcal{L} = \frac{1}{2N} \sum_{i,j=1}^{N} \left[y_{ij} \, \|f_i - f_j\|_2^2 + (1 - y_{ij}) \, \max\left(0, m - \|f_i - f_j\|_2\right)^2 \right]$$

Terms Explained:

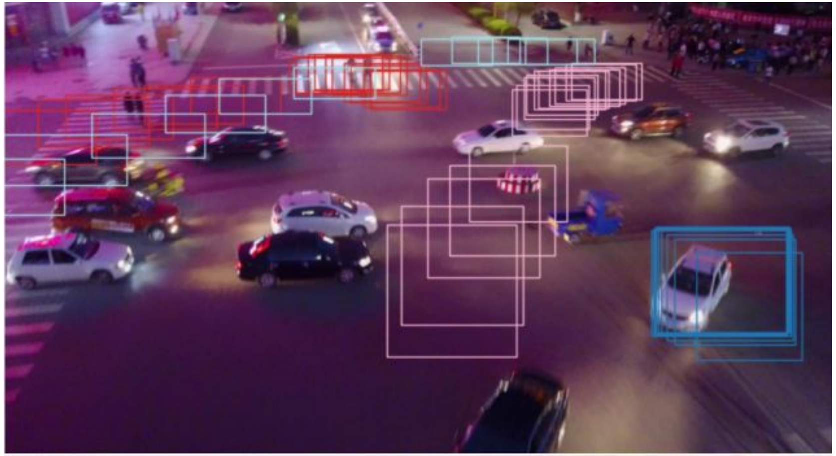
fi,fjf_i, f_jfi,fj: Feature vectors of tracklets i and j

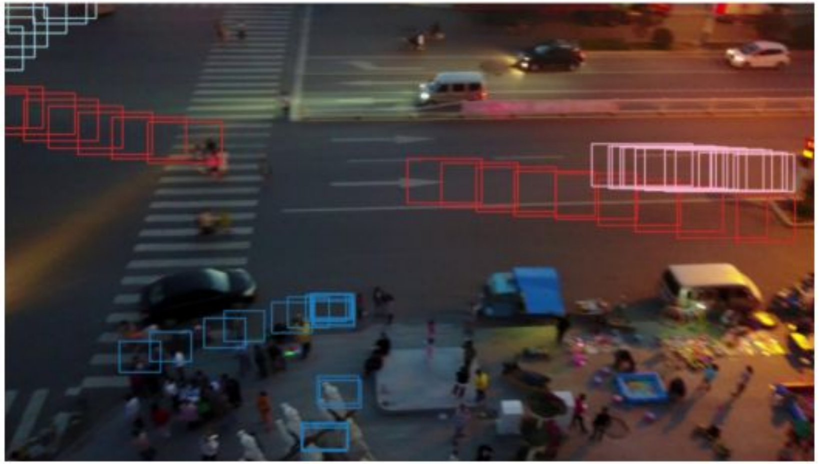
yijy_{ij}yij: 1 if same object, 0 otherwise

mmm: Margin for dissimilar pairs

NNN: Total number of tracklet pairs

Results





Ahmedabad University

Future work

1. Adaptive Parameter Learning: Use **reinforcement learning** or **Bayesian optimization** to dynamically tune thresholds and DBSCAN's ε for improved accuracy.

2. Real-Time Processing Enhancements: Leverage **GPU acceleration**, **approximate nearest neighbors**, and **distributed computing** for real-time performance.

3. Deep Feature Learning: Apply **Siamese** or **triplet networks** to learn more discriminative and robust feature embeddings.

4. Dimensionality Reduction: Use **PCA** or **autoencoders** to reduce feature dimensionality and boost KD-Tree efficiency.

5. Scalability & Robustness: Design scalable and **distributed algorithms** to handle larger datasets and complex environments.

6. Enhanced Evaluation Metrics: Beyond **Precision, Recall, F1**, also measure **MOTA** and **MOTP** for deeper tracking performance insight.

References

[1] P. Zhu, L. Wen, D. Du, X. Bian, H. Fan, Q. Hu, and H. Ling, "Detection and Tracking Meet Drones Challenge," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. X, no. Y, pp. 1–15, 2025.

[2] Q. Ren, J. He, Z. Liu, and M. Xu, "Traffic Flow Characteristics and Traffic Conflict Analysis in the Downstream Area of Expressway Toll Station Based on Vehicle Trajectory Data," Asian Transport Studies, vol. 10, 100138, 2024.

[3] O. Cetintas, G. Bras´o, and L. Leal-Taix´e, "Unifying Short and Long- Term Tracking with Graph Hierarchies," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. X, no. Y, pp. 1–15, 2025.

[4] X. Zhang, H. Yu, Y. Qin, X. Zhou, and S. Chan, "Video-Based Multi-Cacmera Vehicle Tracking via Appearance-Parsing Spatio-Temporal Trajectory Matching Network," IEEE Transactions on Circuits and Systems for Video Technology, vol. 34, no. 10, pp. 10077, Oct. 2024.

[5] D. B. Reid, "An Algorithm for Tracking Multiple Targets," IEEE Transactions on Automatic Control, vol. 24, no. 6, pp. 843–854, 1979.

[6] Y. Bar-Shalom and T. E. Fortmann, Tracking and Data Association, Academic Press, 1988.

[7] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple Online and Realtime Tracking," in Proc. IEEE International Conference on Image Processing (ICIP), 2016, pp. 3464–3468.

[8] J. L. Bentley, "Multidimensional Binary Search Trees Used for Associative Searching," Communications of the ACM, vol. 18, no. 9, pp. 509–517, 1975.

[9] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp.91–110, 2004.

[10] C. de Boor, A Practical Guide to Splines, Springer-Verlag, 1978.