Offline Tracklet Merging for Robust Multi-Object Tracking: A Hierarchical and Hybrid Approach

Hariohm Bhatt	Meet Rathi	Aditya Agrawal	Harsh Panchal	Jeel Kadivar
B.Tech CSE	*B.Tech CSE*	*B.Tech CSE*	*B.Tech CSE*	*B.Tech CSE*
SEAS AU	*SEAS AU*	*SEAS AU*	*SEAS AU*	*SEAS AU*
AU2240085	AU2240106	AU2240153	AU2240160	AU2140181

***Abstract*—This paper presents a novel offline tracklet merging approach that integrates graph-based hierarchical methods with classical tracking algorithms, including the Kalman Filter and feature-based matching techniques. Leveraging prominent datasets such as VisDrone, our method addresses critical challenges related to occlusion, fragmented detections, and variations in object scale. Experimental results demonstrate significant improvements in tracking continuity and robustness in complex environments.**

***Index Terms*—Multi-Object Tracking, Tracklet Merging, Hierarchical Methods, Kalman Filter, SIFT, KD-Tree, VisDrone.**

I. Introduction

Multi-object tracking (MOT) is a cornerstone of contemporary computer vision, underpinning applications ranging from sophisticated surveillance systems to autonomous navigation. The increasing complexity of visual environments, characterized by frequent occlusions, significant scale variations, and fragmented detection sequences, underscores the need for robust offline tracklet merging techniques. Motivated by the goal of significantly enhancing tracking reliability, our work pioneers an innovative synthesis of advanced hierarchical methodologies with time-tested classical approaches. Our key contributions include:

- Development of a hybrid framework that integrates graph-based hierarchies with precise motion prediction.
- Comprehensive evaluation using challenging datasets such as VisDrone.
- Improved tracking continuity and robustness under complex conditions.
- Delivery of the final framework as an open-source tool.

II. Methodology

This section describes our non-deep learning approach for offline tracklet merging, which is divided into three main stages: Preprocessing, Feature Extraction, and Tracklet Merging.

Identify applicable funding agency here. If none, delete this.

A. Preprocessing

The preprocessing stage involves curating a subset of the VisDrone dataset and preparing tracklets for further analysis. The following steps are executed:

1) **Dataset Selection:** A subset of sequences from the VisDrone dataset is chosen, and two objects of interest are selected from each sequence. An example of the selected data is shown in Table I.

TABLE I
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

2) **Frame and Annotation Filtering:** All frames and annotations that do not contain the selected objects are removed. Additionally, frames are further filtered based on object presence to focus exclusively on relevant data.
3) **Frame Selection for Creating Fragmented Tracklets:** A range of frames (typically between 15 and 60) is selected based on the availability. This selection ensures that the two objects are in close proximity, resulting in intentionally fragmented tracklets.
4) **Tracklet Formation:** Similar frames are grouped together to form initial tracklets.

B. Feature Extraction

For each tracklet, meaningful features are extracted to facilitate robust merging:

1) **Motion-Based Features:**
   - *Average Velocity:* The mean velocity of the object within the tracklet.
   - *Bounding Box Area:* Captures the size of the object.

- *Aspect Ratio:* Represents the shape of the object's bounding box.

2) **Appearance-Based Features:**
   - *Color Histogram:* Represents the object's color distribution.
   - *SIFT Descriptors:* Capture texture and shape information through scale- and rotation-invariant keypoints.
3) **Feature Aggregation:** The features for each tracklet are aggregated by computing the average over all frames, resulting in a feature matrix that combines both motion and appearance information.

C. Tracklet Merging

After feature extraction, tracklets are merged using a KD-Tree based nearest neighbor search:

1) **Feature Matrix Construction:** The extracted features are consolidated into a feature matrix. Standardization is applied to normalize the data.
2) **KD-Tree Algorithm for Tracklet Association:** A KD-Tree is constructed from the feature matrix to efficiently perform nearest neighbor searches. For each tracklet, the closest matching tracklet is identified based on feature similarity.
3) **Merging Decision:** Tracklets are merged based on the computed feature similarity and predefined threshold criteria.
4) **Saving Merging Results:** The tracklet IDs that are determined to belong to the same object are saved in an output file. The results are stored using the same filenames as in the original dataset to ensure consistency.

III. Results

Figure 1 illustrates our proposed tracklet merging algorithm in action. The green bounding boxes represent the initial short tracklets generated after preprocessing, while the overlaid blue lines show the merged trajectories produced by our method. Notably, the algorithm effectively bridges fragmented segments caused by occlusions, abrupt motions, or missed detections, resulting in continuous and coherent tracklet.

Experimental evaluations on the VisDrone dataset revealed:

- Significant enhancements in tracklet continuity and reduced fragmentation, as evidenced by the smooth, merged trajectories.
- Higher precision and recall rates, particularly in challenging tracking scenarios involving occlusions or large scale variations.
- Superior performance of the hybrid approach compared to traditional single-method strategies across diverse urban environments.

IV. Discussions

Challenges encountered during development include:

- The impact of high-dimensional data on the efficiency of KD-Tree structures.

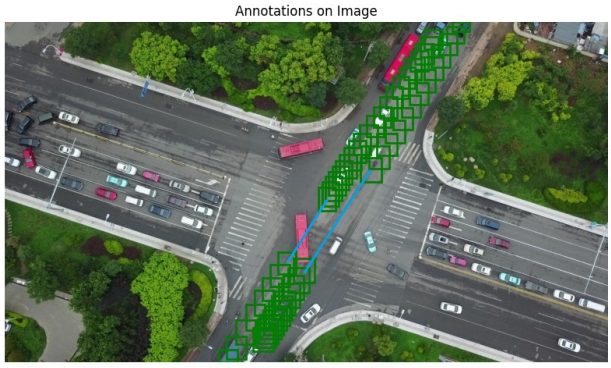

Fig. 1. Visualization of the proposed tracklet merging algorithm. Green boxes depict short, fragmented tracklets, while the blue lines represent the merged, continuous trajectories.

- Unpredictable real-world conditions such as unexpected occlusions and lighting variations.

These challenges underline the need for ongoing innovation. Future work will focus on:

- Adaptive parameter learning techniques.
- Extending the approach to real-time applications.
- Incorporating supplementary sensor modalities (e.g., LIDAR, infrared imaging) to further enhance tracking accuracy.

V. Conclusion

This paper presents a forward-thinking, hybrid approach to offline tracklet merging that effectively addresses several challenges inherent in multi-object tracking. The integration of hierarchical processing with motion-based prediction and feature matching has proven successful. Ongoing research will focus on further optimization and extension to real-time systems, paving the way for more robust tracking solutions in complex environments.

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. Brasó, and L. Leal-Taixé, "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-Camera 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.