Offline track association problem

Kaushi Gohil, Devanshi Rathod, Richa Saraiya, Devasya Rajguru, Jay Golakiya
AU2444022, AU2444023, AU2444002, AU2444012, AU2444021

***Abstract*—Multi-object tracking (MOT) plays a crucial role in video analysis, particularly in applications such as surveillance and autonomous systems. However, consistent tracking of objects across video frames remains a challenging problem, exacerbated by issues like object occlusion, abrupt motion changes, missed detections, and false positives. These challenges lead to fragmented trajectories known as tracklets. This project focuses on developing an offline tracklet merging algorithm to address these issues by post-processing fragmented tracklets and creating continuous, accurate object trajectories.**

***Index Terms*—Multi-object tracking (MOT), tracklet merging, trajectory reconstruction, offline tracking, video analysis.**

I. Introduction

Trajectory reconstruction, object tracking, and multi-hypothesis tracking are critical challenges in the fields of computer vision, robotics, and autonomous systems. Accurately predicting and analyzing movement patterns plays a vital role in applications such as surveillance, autonomous navigation, and aerial tracking. This collection delves into various advanced methodologies designed to improve the precision and efficiency of these processes, including adaptive smoothing splines for trajectory estimation, graph-based multi-target tracking, and innovative approaches for drone-based monitoring.

The content explores a range of cutting-edge techniques, such as velocity-adaptive smoothing splines, Gaussian mixture models for tracking, and deep learning frameworks for multi-object tracking. These approaches address common challenges like noisy observations, occlusions, irregular sampling, and complex motion patterns, ultimately enhancing the reliability of trajectory estimation. By refining data processing methods and optimizing predictive algorithms, these advancements contribute to more robust and accurate tracking systems.

Beyond theoretical improvements, these methodologies have significant real-world implications. Their integration into security systems, autonomous vehicle navigation, and drone-based surveillance highlights their growing impact on AI-driven technologies. This compilation serves as a valuable resource for researchers and industry professionals aiming to advance tracking and reconstruction techniques in complex, dynamic environments.

II. Dataset Discussion

The VisDrone dataset consists of two folders: sequence (image frames) and annotations (ground truth labels). The sequence folder contains image frames used for object detection and tracking. Annotations include frame ID, object ID, width, height, x, y, center x, center y, and occlusion. OpenCV (cv2) is used to detect objects, draw bounding boxes, and plot trajectory lines.

To simulate occlusion, trajectories are randomly segmented into 3 to 5 pieces. Post-occlusion, objects are renamed (e.g., vehicle 0A to vehicle 0B) to indicate continuity.

Ground Truth Label File – Stores all object annotations.

Missing Trajectory CSV File – Assigns NaN values for occluded frames.

III. Methodology

1. MHT-method

The proposed approach utilizes Multi-Hypothesis Tracking (MHT) Tracklet Association to refine trajectory estimation from fragmented tracklets. Tracklets are first extracted from a CSV file, where each object's trajectory is represented as a sequence of (frame ID, center X, center Y).

The MHT association pipeline involves multiple steps. First, the association cost between tracklets is computed based on spatial distance and temporal gap, using an exponential likelihood model. Tracklets with an association cost below a predefined threshold are then connected in a hypothesis graph. To determine the best associations, the Hungarian algorithm is applied, ensuring optimal tracklet assignment. Once assigned, tracklets are merged to form continuous trajectories, reducing fragmentation and improving tracking consistency. Finally, the refined trajectories are visualized to validate the accuracy of the associations.

This methodology enhances object tracking by addressing challenges related to fragmentation and occlusion, leading to more accurate trajectory reconstruction.

2. Spatio-Temporal method

In this work, we propose a spatio-temporal merging approach to reconstruct complete object trajectories from fragmented tracklets, particularly in challenging video sequences with frequent occlusions or missed detections. The core idea is to identify and merge tracklet segments that likely belong to the same object based on both spatial closeness and temporal continuity. Each individual tracklet is treated as a sequence of detections, where each detection includes frame number, bounding box coordinates, center position (computed if missing), object ID, and segment label.

To perform merging, the last detection of one tracklet is compared with the first detection of another. A distance metric is computed based on the Euclidean distance between their centers and the frame gap between them. If both the spatial distance is within a predefined threshold (max distance) and the time gap is within a tolerable range (max time gap), the tracklets are considered candidates for merging. This ensures that only geometrically and temporally consistent tracklets are merged, preventing incorrect associations between objects.

The merging process is greedy and non-overlapping. It iteratively processes each tracklet, appending other compatible tracklets and marking them as visited to avoid redundant merges. After merging, all reconstructed trajectories are sorted by frame number to maintain consistency over time. The final output is a dictionary of merged tracklets, each representing a continuous trajectory spanning across frames with occlusions or gaps effectively bridged.

To support qualitative analysis, a visualization module overlays these merged trajectories onto the original video frames. Unique colors are assigned to each trajectory for easy visual distinction. Bounding boxes and object labels are also drawn to validate the effectiveness of the merging process. This combination of spatio-temporal logic and visualization helps improve both the accuracy and interpretability of multi-object tracking in complex UAV video sequences.

IV. Evaluation Metrics

1. MHT Method

The graph visualizes the merged tracklet trajectories of multiple objects as observed in a tracking scenario using the Multiple Hypothesis Tracking (MHT) method. Each colored path represents the trajectory of a uniquely identified object over time in 2D space, where the X and Y axes denote spatial coordinates. The smooth and continuous lines reflect the system's ability to maintain consistent object identities across frames, while occasional abrupt connections or isolated dots may indicate identity switches or detection gaps.

In terms of evaluation metrics for MHT, this trajectory plot helps qualitatively assess tracking performance such as track purity, track continuity, and association accuracy. A dense, uninterrupted track implies high tracking accuracy, whereas broken or intersecting tracks may suggest challenges in maintaining hypothesis consistency or resolving occlusions. This visualization supports a clearer understanding of MHT's strength in maintaining multiple object tracks simultaneously even in complex scenes.

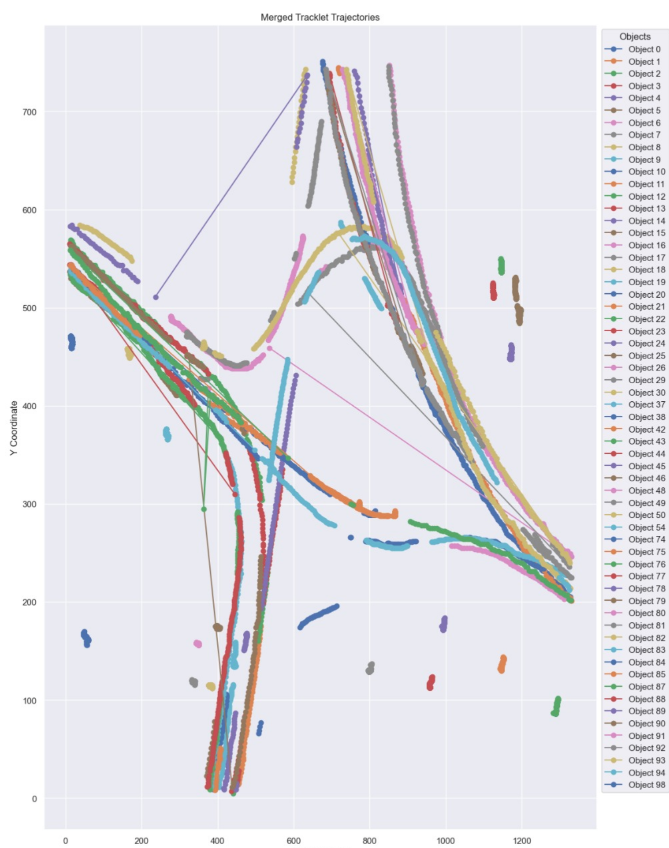

Fig. 1: Merged trajectories of tracked objects in 2D. Each object is represented with a unique color. Some trajectories appear smooth, while others have gaps or abrupt changes, indicating occlusions or tracking interruptions. Thin lines represent tracklet merging, ensuring continuity in fragmented paths. This visualization is useful for object tracking analysis, such as monitoring vehicles or pedestrian movement.

2. Spatio-Temporal method

To assess the effectiveness of our spatio-temporal tracklet merging approach, we utilize the motmetrics library for Multi-Object Tracking (MOT) evaluation. Our evaluation framework compares the merged tracklets against ground truth annotations, focusing on accuracy and consistency of object identities across frames. Both prediction and ground truth files are loaded from CSV format, with object positions represented using center coordinates (cx, cy) and associated with unique object IDs and frame numbers.

The evaluation process begins by parsing both ground truth and predicted trajectories into dictionaries keyed by object IDs. These are further converted into frame-wise mappings to allow comparison of predicted and actual object positions at each frame. We calculate the Euclidean distance between each pair of predicted and ground truth objects within the same frame to determine their association cost. These distances are fed into a MOTAccumulator object, which handles identity matching over time.

For every frame where both prediction and ground truth data are present, pairwise distances are computed and submitted to the accumulator. This allows us to evaluate how well the merged trajectories align with the ground truth in terms of object identity preservation and spatial consistency. Key MOT metrics such as MOTA (Multiple Object Tracking Accuracy), MOTP (Precision), ID switches, and false positives are computed using the motmetrics summary module, which offers a standardized view of the tracker's performance.

This approach ensures that the evaluation reflects not only the spatial accuracy of object localization but also the temporal continuity of identity tracking. The integration of the motmetrics toolkit provides a reliable and interpretable benchmark for analyzing improvements due to spatio-temporal merging, making it easier to identify where the tracker succeeds and where it might still struggle, such as during prolonged occlusions or fast motion.

V. Key Learnings

Robustness of MHT for Occlusion Handling: The MHT method provided strong performance in maintaining identity consistency during brief occlusions and interactions between targets. By maintaining multiple possible hypotheses and delaying hard decisions, MHT was effective in reducing ID switches and fragmentation, especially in crowded or noisy scenes.

1) Efficiency and Simplicity of Spatio-Temporal Merging:
The spatio-temporal method demonstrated a simpler yet efficient mechanism for associating tracklets based on their spatial proximity and temporal continuity. This approach was particularly useful in scenarios where long-term object motion could be inferred from stable motion patterns.
2) Trade-off Between Complexity and Real-time Performance:
While MHT offers higher accuracy in ambiguous cases, it is computationally more expensive due to hypothesis tracking. In contrast, the spatio-temporal method is lightweight and easier to implement, making it suitable for real-time UAV-based tracking systems where speed is critical.
3) Importance of Accurate Evaluation:
The use of motmetrics provided critical insights into how tracklet merging affected performance metrics such as MOTA, ID switches, and False Positives. Frame-wise association and proper evaluation helped in diagnosing strengths and weaknesses of each method effectively.

VI. Conclusion

Both MHT and Spatio-Temporal merging approaches contribute valuable strengths to multi-object tracking in UAV scenarios. MHT excels in maintaining identity under uncertainty but comes with increased computational cost. On the other hand, the spatio-temporal method offers a practical, efficient alternative that leverages motion consistency for merging short-term tracklets.

Integrating both methods in a hybrid pipeline or choosing based on application requirements can lead to more robust and scalable MOT systems. Future work can focus on combining appearance features or deep learning embeddings with spatio-temporal cues for even more accurate tracklet merging.

VII. References

1) Dehghan, A., Sukkarieh, S. A., Shah, M. (2015). GMMCP Tracker: Globally Optimal Generalized Maximum Multi Clique Problem for Multiple Object Tracking. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*
2) Patterson, D. J., Hauptmann, A. G. (2015). Multi-Hypothesis Tracking for Robust Multi-Person Tracking. In *IEEE International Conference on Computer Vision (ICCV).*

3) (2023). SMFMOT: A Novel Approach for UAV-Based Object Tracking. In *Drones, 7*(681).
4) (2022). ByteTrack: Multi-Object Tracking by Associating Every Detection Box. In *European Conference on Computer Vision (ECCV).*
5) (2016). Simple Online and Realtime Tracker (SORT). In *IEEE International Conference on Image Processing (ICIP).*