GTA: Global Tracklet Association for Multi-Object Tracking in Sports

Jiacheng Sun

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Science

University of Washington
2024

Committee:
Jenq-Neng Hwang
James A. Ritcey

Program Authorized to Offer Degree:
Department of Electrical and Computer Engineering

University of Washington

Abstract

GTA: Global Tracklet Association for Multi-Object Tracking in Sports

Jiacheng Sun

Chair of the Supervisory Committee:

Jenq-Neng Hwang

Department of Electrical and Computer Engineering

Multi-object tracking in sports scenarios has become one of the focal points in computer vision, experiencing significant advancements through the integration of deep learning techniques. Despite these breakthroughs, challenges remain, such as accurately re-identifying players upon re-entry into the scene and minimizing ID switches.

In this paper, we propose an appearance-based global tracklet association algorithm designed to enhance tracking performance by splitting tracklets containing multiple identities and connecting tracklets seemingly from the same identity. This method can serve as a plug-and-play refinement tool for any multi-object tracker to further boost their performance.

The proposed method achieved a new state-of-the-art performance on the SportsMOT dataset with a HOTA score of 81.04%. Similarly, on the SoccerNet dataset, our method enhanced multiple trackers' performance, consistently increasing the HOTA score from 79.41% to 83.11%. These significant and consistent improvements across different trackers and datasets underscore our proposed method's potential impact on the application of sports player tracking.

We open-source our project codebase at https://github.com/sjc042/gta-link.git.

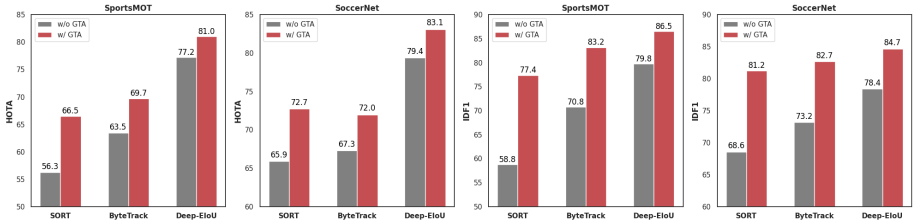

Fig. 1: Our proposed Global Tracklet Association (GTA) method significantly boosts the HOTA and IDF1 score of existing trackers, such as SORT, ByteTrack, and Deep-EIoU, on sports tracking datasets, including SportsMOT and SoccerNet.

1 Introduction

In recent years, advancements in computer vision and deep learning have revolutionized sports analytics, offering unprecedented insights into player performance and strategy. For example, sports video understanding [6, 10], sports field registration [13, 20], and 2D/3D human pose estimation for sports [21]. Among these innovations, sports player tracking systems have emerged as a cornerstone, providing coaches and analysts with valuable data on player movements, positioning, and interactions during game play [12]. These systems have become integral to modern sports, enabling data-driven decision-making and performance optimization across various disciplines.

However, despite significant progress in multi-object tracking technologies, challenges persist in accurately tracking players, including the irregular movements and similar appearances of sports players, and the lack of re-identification algorithm in handling re-entry situation after leaving camera field of view after certain amount of time.

Current state-of-the-art on-line tracking algorithms often fail in long-term object re-identification with the aforementioned challenges. While these trackers perform well in controlled environments or short-term scenarios, they struggle to maintain consistent player identities throughout entire matches or when players re-enter the field after substantial absences. This limitation significantly impacts the accuracy and reliability of player performance analysis, tactical evaluations, and automated game statistics.

To address these persistent issues, our paper proposes an effective plug-and-play post-processing algorithm named Global Tracklet Association (GTA), designed specifically for sports player tracking applications. GTA aims to refine the tracking results of on-line or off-line trackers by improving long-term re-identification capabilities and handling the unique challenges posed by sports environments. By leveraging global temporal information and advanced association techniques, GTA enhances the robustness and accuracy of player tracking, potentially bridging the gap between current tracking technologies and the demanding requirements of professional sports analytics.

2 Related Work

2.1 Sports Player Tracking

Fig. 2: Examples of **different players** on the same teams, highlighting the challenge of distinguishing between players with similar appearances in sports tracking.

With progress made in object detection and tracking, recent studies have focused on challenging multi-object tracking (MOT) scenarios like sports [7,8] and dancing [24]. Tracking sports players is more difficult than tracking pedestrians due to the complex nature of sports scenarios, including frequent occlusions, rapid direction changes, varying player densities, and similar appearance as in Figure 2, and re-entries to the camera view. Several works [4,18,19,29] have proposed methods to handle irregular object motion, improving tracking performance compared to traditional Kalman filter-based methods [3].

There are two predominant types of errors in sports player tracking: the mix-up error (Figure 3) and the cut-off error (Figure 4). The first type, mix-up errors, occur due to irregular movements and occlusions during tracking, leading to a single tracklet mistakenly including multiple players (Figure 3). The second type, cut-off errors, arise because, unlike targets in traditional pedestrian tracking datasets [23], sports players often re-enter the camera's view after exiting. Assigning a new tracking ID to a previously tracked player results in a cut-off error, which fragments a single player's tracklet into multiple parts (Figure 4).

2.2 Person Re-Identification

When a player re-appears in the scene, a re-identification method is needed to assign the correct tracking ID. Although some trackers can re-identify players who re-enter shortly after exiting by lengthening the tracking buffer, cut-off errors from extended absences and re-entries are most prevalent in sports tracking. Person re-identification (ReID) is crucial in multi-object tracking for identifying individuals across video instances or different cameras.

Several methods have been proposed to address ReID challenges. OSNet [32] introduced a lightweight CNN-based method with state-of-the-art performance on various ReID datasets. Many tracking methods incorporate ReID models in the data association stage, such as DeepSORT [26] and BoTSORT [1]. In sports multi-player tracking, Deep-EIoU [19] has shown that using a ReID model can significantly improve tracking performance in sports player tracking scenarios.

2.3 Global Link Models

To address cut-off errors, where a tracker incorrectly assigns different tracking IDs to the same target, several previous works have proposed global linking models that utilize various types of information, such as appearance, spatial,

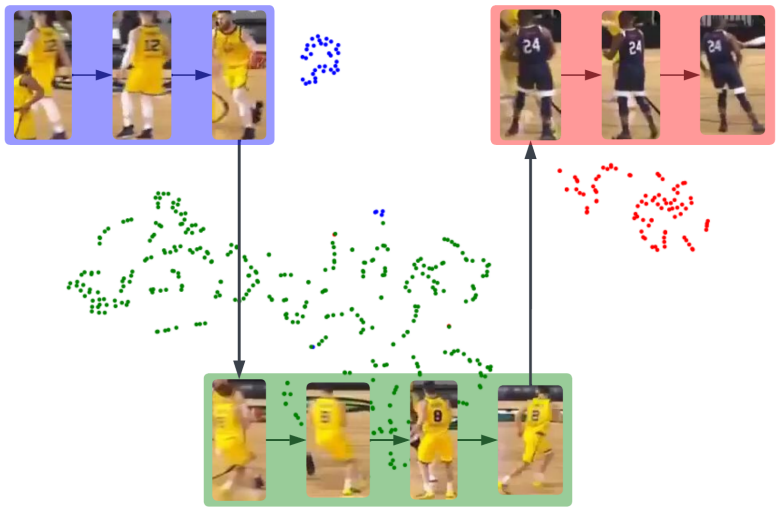

Fig. 3: An example of a mix-up error in a single tracklet. The tracklet output by the online tracking system contains **three** different identities, represented by purple, green, and red points. The figure illustrates the tracklet's features extracted by a ReID model and clustered using the DBSCAN clustering algorithm.

and temporal cues, to associate fragmented tracklets and reassign the correct tracking IDs after the online tracking process. These models either utilize motion or appearance features of tracklets for tracklet-level association. For example, Translink [30] incorporates a CNN and temporal attention network to extract and encode a tracklet's appearance features, treating the merging process of tracklet pairs as a binary classification task. AFLink [9] uses only spatial-temporal information. Some methods [5,17,28] utilize feature clustering methods to merge tracklets and boost the performance on multi-camera tracking scenarios. MambaTrack [15] proposed a motion model that serves as a motion predictor and extracts tracklet motion features for further global tracklet association.

Additionally, some methods exploit object moving direction [14,16] or metadata [27] as clues to conduct global tracklet association and enhance tracking performance. [25] proposed a universal tracklet booster based on CNN and temporal attention to address both mix-up and cut-off errors.

In this work, we propose a novel plug-and-play box-grained global tracklet association model, including a tracklet splitter and a connector. The proposed method is specifically designed to conduct player re-identification and boost the tracking performance in various sports scenarios.

3 Methods

Drawing inspiration from global link models presented in recent works [18] [29], we propose the Global Tracklet Association (GTA) method, a novel plug-and-

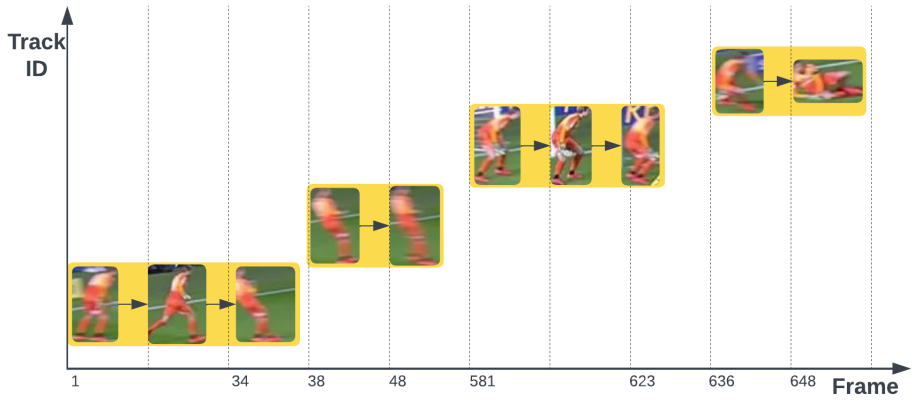

Fig. 4: An example of a cut-off error, where a player's tracklet is fragmented into four separate segments due to the player exiting and re-entering the camera view multiple times throughout the video sequence.

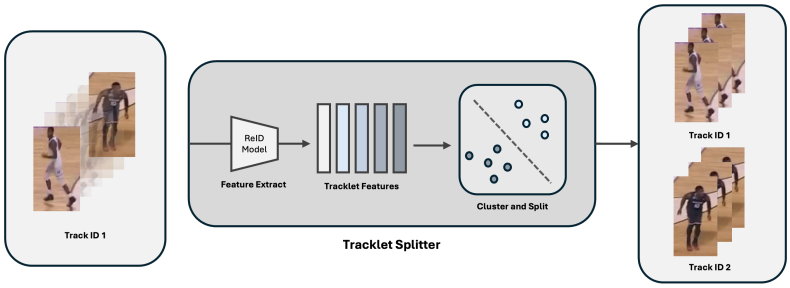

Fig. 5: Illustration of tracklet splitter.

play approach designed to address both mix-up and cut-off errors in multi-object tracking for sports scenarios. Our method consists of two key modules: a ***Tracklet Splitter*** and a ***Tracklet Connector***, which leverage deep feature representations and spatial constraints to enhance tracking accuracy and robustness.

Our proposed post-processing method follows a two-stage process to enhance tracking accuracy. Prior to post-processing, box-grained embedding features from online tracking results are generated by a CNN-based ReID model [32] for each tracklet. In the first stage of our tracklet association model, these tracklets are processed through a ***Tracklet Splitter*** to address mix-up errors, ensuring that instances of different identities are correctly separated, as illustrated in Figure 5. In the second stage, the split tracklets belonging to the same identities are further merged by the proposed ***Tracklet Connector*** to correct cut-off errors, as depicted in Figure 6.

3.1 Tracklet Splitter

The proposed tracklet splitter addresses mix-up errors within a single tracklet, $T = \{t_0, \ldots, t_n\}$, by splitting the tracklet into multiple fragments, t_i, ensuring that each fragment contains only bounding boxes with similar appearance features, measured by close cosine distances in the feature embedding space. We employ DBSCAN clustering [11] to split the tracklet into multiple clusters (tracklet fragments) based on their box-grained appearance embedding feature, which are generated by an OSNet ReID model [32]. The pipeline for our tracklet splitter is illustrated in Figure 5.

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a powerful clustering algorithm that operates by grouping points that are closely packed together while marking points that lie alone in low-density regions as outliers. It is important to note that, unlike traditional DBSCAN implementations, our adapted version assigns outliers to the nearest clusters at the end of the clustering process. This modification is based on the assumption that each bounding box contains a valid detection instance, ensuring that no potentially valuable data points are discarded. This approach allows for a more comprehensive analysis of the tracklet data, preserving information that might otherwise be lost. When applying DBSCAN to the process of tracklet splitting, we incorporate three crucial hyperparameters:

Minimum Samples (s): The Minimum Samples parameter, denoted as s, specifies the minimum number of points required to establish a densed region cluster. This parameter serves two critical purposes in the clustering process. First, it ensures that clusters are only formed when there is a sufficient concentration of points, effectively preventing the creation of clusters with too few instances. Second, it aids in noise reduction by initially labeling points that do not meet this threshold as noise or outliers, especially in the case of sport tracking.

Maximum Neighbor Distance (ϵ): The Maximum Neighbor Distance, represented by ϵ (epsilon), determines the radius within which points are considered neighbors. This parameter is fundamental in controlling the density requirement for cluster formation. In our implementation, we utilize cosine similarity as the distance metric between fragment features, offering several advantages in the context of tracklet splitting. It controls the density requirement for cluster formation, balancing the need to capture variations within a single identity while distinguishing between different identities.

Maximum Clusters (k): Unlike the original algorithm, we introduce a maximum clusters parameter, k, to limit the number of clusters and prevent excessive fragmentation of the tracklet. If the total number of final clusters exceeds k, clusters are progressively merged until only k clusters remain. This modification ensures that a given tracklet is not split into an excessive number of fragments, maintaining a balance between accuracy and fragmentation.

3.2 Tracklet Connector

Our tracklet connector is designed to merge fragmented tracklets from the same identity using a hierarchical clustering approach with a distance threshold α,

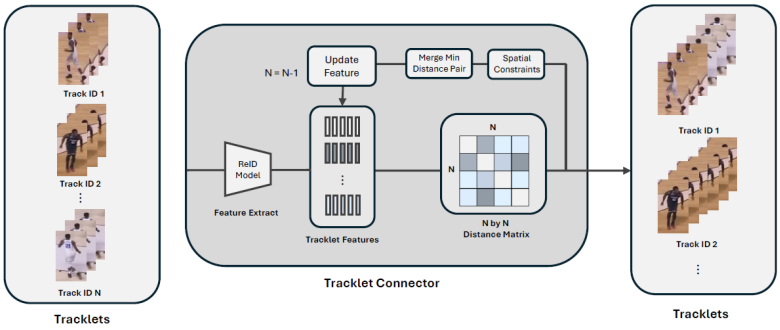

Fig. 6: Illustration of tracklet connector.

addressing the cut-off errors that occur when players exit and re-enter the field of view, or the ID switch during the tracking process. Our approach, as depicted in Figure 6, consists of tracklet clustering based on tracklets distance while applying temporal and spatial constraints to ensure accurate and reliable tracklet merging, and the three main components for the tracklet connector are listed below:

Constructing Tracklet Feature Distance Matrix. The first step of our tracklet connector is the initial construction of a symmetric cosine distance matrix that measures the similarity between all tracklet pairs within a video sequence. The distance matrix is constructed as follows:

$$D_{i,j} = \begin{cases} 1, \text{ if } i \neq j \ \ \& \ \ \Pi_i \cap \Pi_j \neq \emptyset \\ \frac{1}{N_i N_j} \sum_{i \in \Pi_i} \sum_{j \in \Pi_j} (1 - \frac{F_m^i \cdot F_n^j}{\|F_m^i\| \ \|F_n^j\|}), \text{ otherwise} \end{cases} \tag{1}$$

where $D_{i,j}$ denotes the tracklet distance between tracklet pair T_i and T_j; Π_i and Π_j represent temporal spans of tracklet T_i and T_j; F_m^i and F_n^j are tracklet T_i and T_j's embedding feature at frame m and n respectively; N_i and N_j are the length of the tracklet. This distance enables a nuanced representation of tracklet similarities based on the rich nature of deep features and temporal clues.

Enforcing Spatial Constraints. In a common sports game video where the camera position remains fixed, player movement is constrained by the field boundaries, and players do not exit and re-enter from opposite sides of the field. To reflect this, our method enforces spatial constraints for merging tracklets using:

$$\theta_{\text{hor}} = \beta \Delta_{\text{max,hor}}, \tag{2}$$

$$\theta_{\text{ver}} = \beta \Delta_{\text{max,ver}}, \tag{3}$$

where $\Delta_{\text{max,hor}}$ and $\Delta_{\text{max,ver}}$ denote the maximum horizontal and vertical distances from all bounding boxes in the current video. The spatial factor, $\beta \in (0, 1]$, sets thresholds of θ_{hor} and θ_{ver}, limiting the association distances threshold between temporally adjacent tracklet's exit and entry points using bounding box

center. Then, $D_{i,j}$ will be updated.

$$D_{i,j} = 1, \text{ if } \Delta_{i,j,hor} > \theta_{hor} \text{ or } \Delta_{i,j,ver} > \theta_{ver}, \tag{4}$$

where $\Delta_{i,j,hor}$ and $\Delta_{i,j,ver}$ are the horizontal and vertical distance between the beginning and ending of tracklets i and j, respectively. This approach filters out unreasonable associations between tracklets, enhancing the accuracy of the tracklets merging process.

Hierarchical Clustering. After the distance matrix is obtained using equation 1, we further conduct hierarchical clustering following [14] to merge fragment tracklets. We continuously merge the tracklets until no tracklet pair's distance is larger than the merging threshold α.

4 Experiments

4.1 Datasets

We evaluate our method on two large-scale sports player tracking datasets: SportsMOT [8] and SoccerNet [7]. These datasets are representative of athlete tracking in team sports scenarios, presenting unique challenges such as players with similar appearances, frequent re-entries into the camera's field of view, and abrupt changes in motion.

SportsMOT is a multi-object tracking dataset that contains over 240 video sequences spanning three team sports: basketball, football, and volleyball. Each sport presents its unique challenges, such as the fast-paced, close-quarters action of basketball, the wide-field dynamics of football, and the rapid vertical movements in volleyball. The dataset provides a robust foundation for developing and testing tracking algorithms in complex, real-world sports environments and has been widely used in benchmarking MOT for sport tracking.

SoccerNet focuses exclusively on videos captured from soccer matches, providing a collection of over 100 high-quality video clips extracted from professional games. For our experiments, we utilize the test set from the SoccerNet tracking dataset published in 2023.

4.2 Implementation details

Detector. For the SportsMOT test set, we use YOLOX as the detection model following [19], and for the SoccerNet 2023 test set, we directly apply oracle detection following others' implementations [19] for fair comparison.

Tracker. In our work, we test our tracklet refinement method on three trackers: SORT [3], ByteTrack [31], and Deep-EIoU [19]. SORT and ByteTrack are both implemented to track with spatial and motion cues. Deep-EIoU incorporates both spatial and appearance information to achieve state-of-the-art performance with a HOTA score of 77.2% on the SportsMOT test set and 85.4% on the SoccerNet test set published in 2022.

Table 1: Tracking performance on SportsMOT before and after applying our Global Tracklet Association (GTA) method.

Method	HOTA↑	AssA↑	IDF1↑	DetA↑	MOTA↑	IDs↓
SORT [26]	56.28	42.67	58.83	74.30	85.11	5180
SORT + GTA	**66.52** (+10.24)	**59.59** (+16.92)	**77.37** (+18.54)	74.29	**85.27**	**3547** (-1633)
ByteTrack [31]	63.46	51.81	70.76	77.81	94.91	3147
ByteTrack + GTA	**69.74** (+6.28)	**62.61** (+10.80)	**83.16** (+12.40)	77.72	**95.01**	**2107** (−1040)
Deep-EIoU [19]	77.21	67.63	79.81	88.22	96.30	2909
Deep-EIoU + GTA	**81.04** (+3.83)	**74.51** (+6.88)	**86.51** (+6.70)	88.21	**96.32**	**2737** (−172)

ReID Model. For our experiments, we use the OSNet [32] model trained on SportsMOT dataset. OSNet is chosen for its capability to capture discriminative features suitable for re-identification of athletes with similar appearances, which is critical for our tracklet refinement process.
Hyperparameters. We set minimum cluster samples s to 5, maximum neighbor distance threshold ϵ to 0.6, maximum clusters k to 3, merging threshold α to 0.4, and β to 1 for SportsMOT dataset and 0.7 for SoccerNet dataset, respectively.

4.3 Performance

Evaluation Metrics. We utilize commonly used tracking metrics, including HOTA [22] for its comprehensive evaluation of both detection and association accuracy (DetA and AssA); and CLEAR metrics [2], where IDF1 and MOTA serve as the standard benchmark for tracking performance across various scenarios, and IDs to verify the effectiveness of our method in reducing and associating the correct identities.
Performance on SportsMOT. In Table 1, our proposed method demonstrates significant performance improvements over existing trackers across various tracking metrics like HOTA, AssA, IDF1, IDs, and Frag. For the SportsMOT dataset, GTA achieved the highest HOTA improvement of 10.24% for SORT and 3.83% for Deep-EIoU, reaching a state-of-the-art HOTA of 81.04%. Demonstrating the GTA method is applicable for diverse kinds of sports tracking.
Performance on SoccerNet. In Table 2, GTA improved HOTA by 6.84% for SORT and 3.7% for Deep-EIoU. These results demonstrate the effectiveness of our tracklet refinement method in enhancing tracker performance on challenging sports player tracking datasets. Figure 8 illustrates the qualitative results, showing GTA effectively connecting tracklet fragments from online tracking.

4.4 Ablation Study

To evaluate the effectiveness of the splitter and connector modules of our proposed method, we conduct ablation studies on the performance gain of SORT [3], ByteTrack [31], and DeepEIoU [19] after applying each module. The ablation study summarized in Table 3 highlights the effectiveness of different modules of the Global Tracklet Association (GTA) method on the performance of the three

Table 2: Tracking performance on SoccerNet before and after applying our Global Tracklet Association (GTA) method.

Method	HOTA↑	AssA↑	IDF1↑	DetA↑	MOTA↑	IDs↓
SORT [26]	65.89	57.15	68.56	76.11	82.59	3281
SORT + GTA	**72.73** (+6.84)	**69.62** (+12.47)	**81.24** (+12.68)	76.04	**82.93**	**1374** (-1907)
ByteTrack [31]	67.30	60.38	73.22	75.14	84.66	4558
ByteTrack + GTA	**71.97** (+4.67)	**69.03** (+8.65)	**82.67** (+9.45)	75.10	**84.91**	**3149** (-1409)
Deep-EIoU [19]	79.41	71.55	78.40	88.14	87.92	2803
Deep-EIoU + GTA	**83.11** (+3.70)	**78.38** (+6.83)	**84.66** (+6.26)	88.13	**88.03**	**2188** (-615)

trackers across the SportsMOT and SoccerNet datasets. The proposed Connector alone results in notable improvements for SORT in HOTA, IDF1, and AssA scores, with an increase of 9.15% in HOTA on SportsMOT and 5.85% on SoccerNet. When the Splitter and Connector are both applied for SORT, we obtained HOTA improvements of 10.24% on SportsMOT and 6.84% on SoccerNet, along with substantial increases in IDF1 and AssA, demonstrating the effectiveness of both modules.

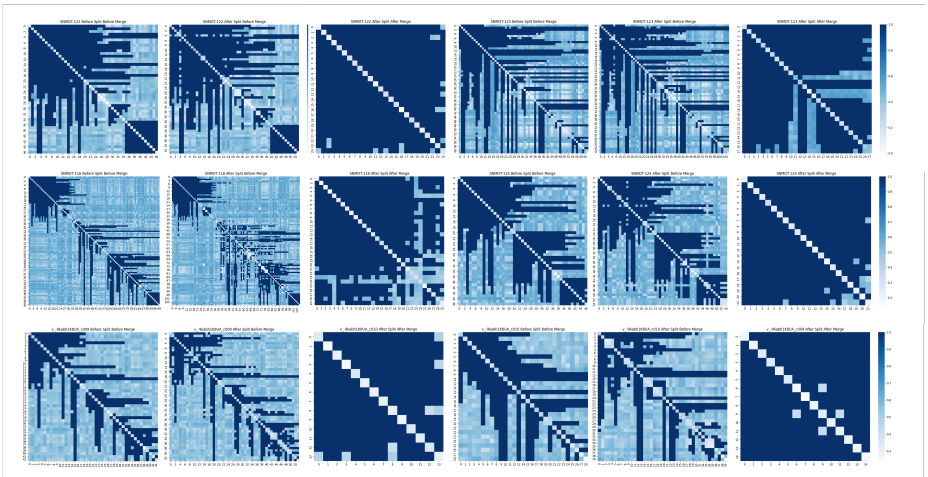

Fig. 7: The cosine distance matrix of the embeddings of three stages: (1) Before Split, (2) After Split, and (3) After Connect. Both the x-axis and y-axis represent the IDs, while the darker color represents the farther distance.

5 Conclusion

In this paper, we proposed the Global Tracklet Association (GTA), a novel tracklet refinement method to enhance the performance of existing trackers in challenging sports player tracking scenarios. Our approach effectively addresses common issues such as mix-up errors and cut-off errors by leveraging a combination

Table 3: Effectiveness of ***Splitter*** and ***Connector*** on different MOT algorithms and datasets.

Dataset	Method	*Connector*	*Splitter*	HOTA↑	AssA↑	IDF1↑
SportsMOT	SORT [3]			56.28	42.67	58.83
		✓		65.43 (+9.15)	57.77 (+15.10)	76.13 (+17.30)
		✓	✓	66.52 (+10.24)	59.59 (+16.92)	77.37 (+18.54)
	ByteTrack [31]			63.46	51.81	70.76
		✓		69.51 (+6.05)	62.21 (+10.40)	82.88 (+12.12)
		✓	✓	69.74 (+6.28)	62.61 (+10.80)	83.16 (+12.40)
	DeepEIoU [19]			77.21	67.63	79.81
		✓		80.48 (+3.27)	73.50 (+5.87)	85.76 (+5.95)
		✓	✓	81.04 (+3.83)	74.51 (+6.88)	86.51 (+6.70)
SoccerNet	SORT			65.89	57.15	68.56
		✓		71.74 (+5.85)	67.74 (+10.59)	79.73 (+11.17)
		✓	✓	72.73 (+6.84)	69.62 (+12.47)	81.24 (+12.68)
	ByteTrack			67.30	60.38	73.22
		✓		71.05 (+3.75)	67.30 (+6.92)	78.22 (+4.61)
		✓	✓	71.97 (+4.67)	69.03 (+8.65)	82.67 (+9.45)
	DeepEIoU			79.41	71.55	78.40
		✓		82.01 (+2.60)	76.32 (+4.77)	83.13 (+4.73)
		✓	✓	83.11 (+3.70)	78.38 (+6.83)	84.66 (+6.26)

of ReID model and unsupervised clustering techniques for tracklet splitting and merging. The integration of our GTA method with trackers like SORT, ByteTrack, and Deep-EIoU has demonstrated significant improvements in various tracking performance metrics, particularly in metrics that related to associations like HOTA, AssA, and IDF1, while also reducing the number of ID switches (IDs) and tracklet fragments (Frag). Our proposed module achieves state-of-the-art performance on SportsMOT and SoccerNet datasets. Future work will focus on exploring additional enhancements and adaptations of the GTA method for other multi-object tracking scenarios beyond sports.

References

1. Aharon, N., Orfaig, R., Bobrovsky, B.Z.: Bot-sort: Robust associations multi-pedestrian tracking. arXiv preprint arXiv:2206.14651 (2022)
2. Bernardin, K., Stiefelhagen, R.: Evaluating multiple object tracking performance: The clear mot metrics. J. Image Video Process. **2008** (2008)
3. Bewley, A., Ge, Z., Ott, L., Ramos, F., Upcroft, B.: Simple online and realtime tracking. In: 2016 IEEE International Conference on Image Processing (ICIP). pp. 3464–3468 (2016). `https://doi.org/10.1109/ICIP.2016.7533003`
4. Cao, J., Pang, J., Weng, X., Khirodkar, R., Kitani, K.: Observation-centric sort: Rethinking sort for robust multi-object tracking. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9686–9696 (2023)
5. Cherdchusakulchai, R., Phimsiri, S., Trairattanapa, V., Tungjitnob, S., Kudisthalert, W., Kiawjak, P., Thamwiwatthana, E., Borisuitsawat, P., Tosawadi, T., Choppradi, P., et al.: Online multi-camera people tracking with spatial-temporal mechanism and anchor-feature hierarchical clustering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7198–7207 (2024)

Fig. 8: Tracking visualization of athletes before and after applying Global Tracklet Association (GTA). In rows one and three, dashed lines indicate association errors, showing inconsistent athlete IDs across frames. In contrast, solid lines represent correct associations with consistent IDs after applying GTA (rows two and four). The comparison highlights how the algorithm improves ID continuity across frames in both basketball (first two rows) and soccer (second two rows) sequences.

6. Cioppa, A., Deliège, A., Giancola, S., Ghanem, B., Van Droogenbroeck, M., Gade, R., Moeslund, T.B.: A context-aware loss function for action spotting in soccer videos. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2020)
7. Cioppa, A., Giancola, S., Deliege, A., Kang, L., Zhou, X., Cheng, Z., Ghanem, B., Van Droogenbroeck, M.: Soccernet-tracking: Multiple object tracking dataset and benchmark in soccer videos. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3491–3502 (2022)
8. Cui, Y., Zeng, C., Zhao, X., Yang, Y., Wu, G., Wang, L.: Sportsmot: A large multi-object tracking dataset in multiple sports scenes. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9921–9931 (2023)
9. Du, Y., Zhao, Z., Song, Y., Zhao, Y., Su, F., Gong, T., Meng, H.: Strongsort: Make deepsort great again. IEEE Transactions on Multimedia (2023)
10. Duan, H., Zhao, Y., Chen, K., Lin, D., Dai, B.: Revisiting skeleton-based action recognition. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (Jun 2022). https://doi.org/10.1109/cvpr52688.

2022.00298, http://dx.doi.org/10.1109/CVPR52688.2022.00298
11. Ester, M., Kriegel, H.P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). pp. 226–231. AAAI Press (1996)
12. Gade, R., Moeslund, T.B.: Constrained multi-target tracking for team sports activities. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (June 2020)
13. Gutiérrez-Pérez, M., Agudo, A.: No bells just whistles: Sports field registration by leveraging geometric properties. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 3325–3334 (June 2024)
14. Hsu, H.M., Huang, T.W., Wang, G., Cai, J., Lei, Z., Hwang, J.N.: Multi-camera tracking of vehicles based on deep features re-id and trajectory-based camera link models. In: CVPR workshops. pp. 416–424 (2019)
15. Huang, H.W., Yang, C.Y., Chai, W., Jiang, Z., Hwang, J.N.: Exploring learning-based motion models in multi-object tracking. arXiv preprint arXiv:2403.10826 (2024)
16. Huang, H.W., Yang, C.Y., Hwang, J.N.: Multi-target multi-camera vehicle tracking using transformer-based camera link model and spatial-temporal information. arXiv preprint arXiv:2301.07805 (2023)
17. Huang, H.W., Yang, C.Y., Jiang, Z., Kim, P.K., Lee, K., Kim, K., Ramkumar, S., Mullapudi, C., Jang, I.S., Huang, C.I., Hwang, J.N.: Enhancing multi-camera people tracking with anchor-guided clustering and spatio-temporal consistency id re-assignment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 5239–5249 (June 2023)
18. Huang, H.W., Yang, C.Y., Ramkumar, S., Huang, C.I., Hwang, J.N., Kim, P.K., Lee, K., Kim, K.: Observation centric and central distance recovery for athlete tracking. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 454–460 (2023)
19. Huang, H.W., Yang, C.Y., Sun, J., Kim, P.K., Kim, K.J., Lee, K., Huang, C.I., Hwang, J.N.: Iterative scale-up expansioniou and deep features association for multi-object tracking in sports. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 163–172 (2024)
20. Jiang, W., Higuera, J.C.G., Angles, B., Sun, W., Javan, M., Yi, K.M.: Optimizing through learned errors for accurate sports field registration. In: 2020 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE (2020)
21. Jiang, Z., Ji, H., Menaker, S., Hwang, J.N.: Golfpose: Golf swing analyses with a monocular camera based human pose estimation. In: 2022 IEEE International Conference on Multimedia and Expo Workshops (ICMEW). pp. 1–6. IEEE (2022)
22. Luiten, J., Osep, A., Dendorfer, P., Torr, P., Geiger, A., Leal-Taixé, L., Leibe, B.: Hota: A higher order metric for evaluating multi-object tracking. International journal of computer vision **129**, 548–578 (2021)
23. Milan, A., Leal-Taixe, L., Reid, I., Roth, S., Schindler, K.: Mot16: A benchmark for multi-object tracking (2016), arXiv preprint arXiv:1603.00831
24. Sun, P., Cao, J., Jiang, Y., Yuan, Z., Bai, S., Kitani, K., Luo, P.: Dancetrack: Multi-object tracking in uniform appearance and diverse motion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20993–21002 (2022)

25. Wang, G., Wang, Y., Gu, R., Hu, W., Hwang, J.N.: Split and connect: A universal tracklet booster for multi-object tracking. IEEE Transactions on Multimedia **25**, 1256–1268 (2022)
26. Wojke, N., Bewley, A., Paulus, D.: Simple online and realtime tracking with a deep association metric (2017), in 2017 IEEE international conference on image processing(ICIP), pages 3645–3649. IEEE
27. Yang, C.Y., Huang, H.W., Jiang, Z., Kuo, H.C., Mei, J., Huang, C.I., Hwang, J.N.: Sea you later: Metadata-guided long-term re-identification for uav-based multi-object tracking. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) Workshops. pp. 805–812 (January 2024)
28. Yang, C.Y., Huang, H.W., Kim, P.K., Jiang, Z., Kim, K.J., Huang, C.I., Du, H., Hwang, J.N.: An online approach and evaluation method for tracking people across cameras in extremely long video sequence. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7037–7045 (2024)
29. Yang, F., Odashima, S., Masui, S., Jiang, S.: Hard to track objects with irregular motions and similar appearances? make it easier by buffering the matching space. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 4799–4808 (2023)
30. Zhang, Y., Wang, S., Fan, Y., Wang, G., Yan, C.: Translink: Transformer-based embedding for tracklets' global link. In: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1–5. IEEE (2023)
31. Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., Wang, X.: Bytetrack: Multi-object tracking by associating every detection box. In: European conference on computer vision. pp. 1–21. Springer (2022)
32. Zhou, K., Yang, Y., Cavallaro, A., Xiang, T.: Omni-scale feature learning for person re-identification (2019), proceedings of the IEEE/CVF International Conference on Computer Vision