Iterative Scale-Up ExpansionIoU and Deep Features Association for Multi-Object Tracking in Sports

Hsiang-Wei Huang
hwhuang@uw.edu
Information Processing Lab, University of Washington
Seattle, Washington, USA

Cheng-Yen Yang
cycyang@uw.edu
Information Processing Lab, University of Washington
Seattle, Washington, USA

Jiacheng Sun
sjc042@uw.edu
Information Processing Lab, University of Washington
Seattle, Washington, USA

Jenq-Neng Hwang
hwang@uw.edu
Information Processing Lab, University of Washington
Seattle, Washington, USA

Chung-I Huang
1203033@narlabs.org.tw
National Center for High-Performance Computing
Hsinchu, Taiwan

ABSTRACT

Multi-object tracking algorithms have made significant advancements due to the recent developments in object detection. However, most existing methods primarily focus on tracking pedestrians or vehicles, which exhibit relatively simple and regular motion patterns. Consequently, there is a scarcity of algorithms that address the tracking of targets with irregular or non-linear motion, such as multi-athlete tracking. Furthermore, popular tracking algorithms often rely on the Kalman filter for object motion modeling, which fails to track objects when their motion contradicts the linear motion assumption of the Kalman filter. Due to this reason, we proposed a novel online and robust multi-object tracking approach, named Iterative Scale-Up ExpansionIoU and Deep Features for multi-object tracking. Unlike conventional methods, we abandon the use of the Kalman filter and propose utilizing the iterative scale-up expansion IoU. This approach achieves superior tracking performance without requiring additional training data or adopting a more robust detector, all while maintaining a lower computational cost compared to other appearance-based methods. Our proposed method demonstrates remarkable effectiveness in tracking irregular motion objects, achieving a score of 76.9% in HOTA. It outperforms all state-of-the-art tracking algorithms on the SportsMOT dataset, covering various kinds of sport scenarios.

CCS CONCEPTS

• **Computing methodologies → Tracking**.

KEYWORDS

deep learning, computer vision, multi-Object Tracking

MMAsia '23, Dec 06–08, 2023, Tainan, Taiwan

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

ACM Reference Format:
Hsiang-Wei Huang, Cheng-Yen Yang, Jiacheng Sun, Jenq-Neng Hwang, and Chung-I Huang. 2023. Iterative Scale-Up ExpansionIoU and Deep Features Association for Multi-Object Tracking in Sports. In *Proceedings of (MMAsia '23).* ACM, New York, NY, USA, 6 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Multi-Object Tracking is a fundamental computer vision task that aims to track multiple objects in a video and associate them with unique tracking IDs while localizing them in each frame. While most recent tracking algorithms focus on pedestrians or vehicles [5, 8, 9, 16], there is limited research on tracking objects with irregular motion, such as athletes. Given the growing demand for sports data analysis for applications like tactical analysis and athletes' movement statistics, this aspect requires significant attention.

To address this need, we propose a novel and robust online multi-object tracking algorithm specifically designed for objects with irregular and unpredictable motion. Our experimental results demonstrate that our algorithm effectively handles the irregular and unpredictable motion of athletes during the tracking process. It outperforms all tracking algorithms on the public benchmark [4] without introducing extra computational cost, while maintaining the algorithm online. Therefore, in this paper, we assert three main contributions:

- We present an association method, named ExpansionIoU, which has lower computational cost compared to the traditional Kalman filter-based tracking algorithms.
- We proposed a strong tracking algorithm leverage iterative scale-up expansionIoU and deep features association for robust multi-object tracking.
- Achieved the best performance among all the SOTA tracking algorithms with an HOTA of 76.9 on the SportsMOT [4] dataset, which consists of athlete tracking in three different sports including basketball, volleyball and football.

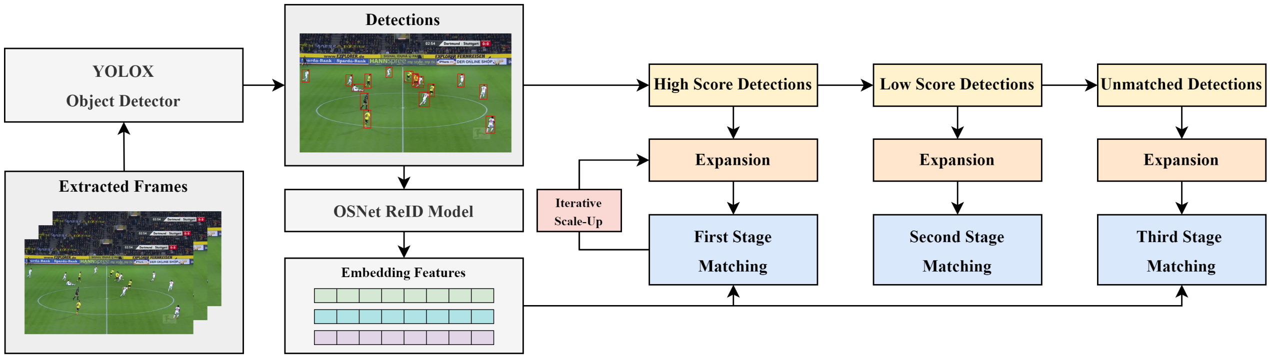

Figure 1: The proposed tracking pipeline.

2 RELATED WORK

2.1 Multi Object Tracking using Kalman Filter

Most of the existing tracking algorithms [2, 3, 9–12, 23, 24, 28, 30] incorporate Kalman filter [13] as a method for object motion modeling. Kalman filter can formulate object's motion as state and can be used to predict object's next frame location according to the object's motion from the previous frames. Kalman filter has shown effectiveness in multi-object tracking across several public benchmarks including [5, 16, 19]. However, due to the Kalman filter's linear motion assumption, Kalman filter might failed to track an object when dealing with nonlinear motion, due to this reason, OCSORT [3] proposed several methods including observation-centric re-update to fix the Kalman filter's parameters during the tracking process and prevent error accumulations when an object is not tracked. The performance has shown effectiveness for tracking objects with irregular motion in several public dataset [4, 19].

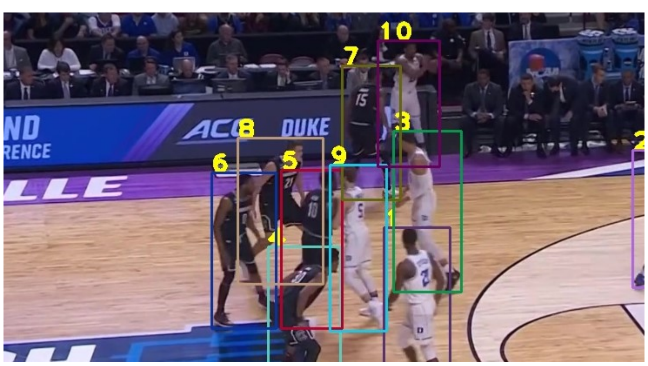

Figure 2: An example of the occlusion problem encountered during multi-athlete tracking. Occlusion can significantly hinder detection and tracking performance, and the occlusion issue in athlete tracking is particularly severe when compared to pedestrian tracking due to the high intensity of sports characteristics.

2.2 Location-based Multi-Object Tracking

Tracking can also be conducted based on the position information, given a high frame rate input video sequence, the object's position shifting between frames is going to be small due to the high frame rates, thus making the position information a reliable clue for association between frames. Several methods [12, 18] utilizes the bounding boxes' distance as the cost for bounding box association, while another recent method named Cascade BIoU Tracker [26] utilize BufferIoU, which compares object's buffered bounding boxes' IoU between adjacent frames for object tracking with irregular motion. By adding buffers toward the bounding boxes, Cascade BIoU Tracker expand the matching space of detections and tracks and shows decent performance on several public MOT benchmarks [5, 19].

2.3 Appearance-based Multi-Object Tracking

With the recent development and improvement of object ReID model [32] and training tricks [14], many tracking algorithms incorporate ReID into the association process. Some method uses joint detection and embedding approach [22, 30] to produce detection and object embedding at the same time to achieve realtime tracking. While the other methods [1, 23] applied another stand alone ReID model to extract detection's embedding features for association. The appearance-based tracking method improves the tracking robustness with an extra appearance clue, while sometimes the appearance can be unreliable due to several reasons including occlusions, appearance variation caused by object's rotation or the lighting condition.

2.4 Sports Player Tracking

Numerous studies have been conducted with the aim of monitoring team players in team sports during games. This monitoring serves not only to automate the recording of game statistics but also enables sports analysts to obtain comprehensive information from a video scene understanding perspective. The majority of studies utilize a tracking-by-detection method and integrate a re-identification network to generate an embedding feature for each player.

Vats et al. [21] combined team identification and player identification approaches to improve the tracking process in hockey.

Similarly, Yang et al. [27] and Maglo et al. [15] demonstrated that by localizing the field and players, the tracking results in football can be more accurate. Additionally, Sangüesa et al. [17] proposed linking the semantics of poses and actions to the embedding features to enhance tracking in basketball. While Huang et al. [12] introduces the central distance recovery and an appearance based post processing to conduct tracking on multiple sports including basketball, volleyball and football [4].

3 PROPOSED METHODS

Our proposed method follows the tracking by detection paradigm and conduct tracklets and detections association every frame. We first apply the object detector YOLOX on each input frame, and then we conduct association based on several clue including the similarity between extracted appearance features and the expansionIoU between the tracklets and detections. After the association cost is obtained, Hungarian algorithm is conducted to get the association results between tracklets and detections.

3.1 Appearance-based Association

The appearance similarity is a strong clue for object association between frames, the similarity can be calculated by the cosine similarity between the appearance features, it can also be used to filtered out some impossible association. The cost for appearance association $Cost_A$ can be directly obtained from the cosine similarity with the following formula:

$$Cost_A = 1 - \text{Cosine Similarity} = 1 - \frac{a \cdot b}{\|a\|\|b\|} \tag{1}$$

Here, a and b are the tracklet's appearance feature and the detection's appearance feature. A higher cosine similarity denotes a higher similarity in appearance, while a lower cosine similarity means the tracklet's appearance and the detection's appearance are different.

3.2 Association with ExpansionIoU

Some previous works [25, 26] already show expanding the bounding box can benefit the association process of objects with irregular motion. Here, we proposed ExpansionIoU, a robust association clue for tracklets and detections matching under nonlinear motion. Different from the previous works [26], we found out expanding the bounding box even more during association can leads to a significant better performance in athlete tracking, and using two hyperparameters for expansion controlling can make the tracking process even robust. In this work, the expansion is controlled by two hyperparameters: E and λ. Where E is the expansion scale, and λ controls the relative contribution of the original width in the expansion calculation. The expand bounding box's new height $h_{expansion}$ and width $w_{expansion}$ can be derived from the following formulas:

$$E = \frac{h_{expansion} - \lambda h}{2h} = \frac{w_{expansion} - \lambda w}{2w} \tag{2}$$

where h,w denotes the original bounding box's height and width, respectively. Note that the operation of expanding bounding box does not change several important objects' original embedding features like the bounding box center, aspect ratio, or the appearance features. By expanding the search space, we can associate those tracklets and detections without no IoU between adjacent frame, which is considered a common situation when the target's movement is fast, especially in sport games.

3.3 Confidence Score Aware Cascade Matching

We perform matching with two stages, following ByteTrack [28], we gave the high confidence score detections with more priority during the matching process. The high score detections usually implies less occlusion, hence a higher chance to preserve more reliable appearance features. Due to this reason, the first stage of matching is based on the association cost of both appearance and ExpansionIoU, denoted as C_{stage1}. The first stage of matching is build upon several rounds of iterative associations with a gradually scale-up buffer scale, illustrated in 3.4. In the second round of matching with low score detections, only ExpansionIoU is used, the cost is denoted as C_{stage2}.

In our first matching stage, we follow [1] and abandoned the IoU-ReID weighted cost method used in several previous works [23, 29], where the cost is a weighted sum of the appearance cost C_A and IoU cost C_{IoU}:

$$C = \lambda C_A + (1 - \lambda) C_{IoU} \tag{3}$$

We first filtered out some impossible association by setting a cost threshold for both appearance and ExpansionIoU (EIoU). The appearance cost C_A will be set to 1 if either the cost is bigger than this threshold. Finally, the first stage association cost C_{stage1} will be set as the minimum of half of the appearance cost C_A and EIoU cost C_{EIoU}. With τ_A and τ_{EIoU} denotes the threshold for the cost filter, we can write the appearance cost C_A as:

$$C_A = \begin{cases} 1, & \text{if } C_A > \tau_A \text{ or } C_{EIoU} > \tau_{EIoU} \\ 0.5\ C_A, & \text{otherwise} \end{cases} \tag{4}$$

The final cost in the first stage of matching C_{stage1} will be the minimum between appearance cost C_A and EIoU cost C_{EIoU}.

$$C_{stage1} = \min(C_A, C_{EIoU}) \tag{5}$$

While the association cost in the second matching stage C_{stage2} will be equal to the EIoU cost C_{EIoU}.

Sport Type	# of tracks	# of frames	Track Len	Density
Basketball	10	845.4	767.9	9.1
Football	22	673.9	422.1	12.8
Volleyball	12	360.4	335.9	11.2

Table 1: Summary of the SportsMOT dataset split by the type of sport. The number of tracks, number of frames, track length, and track density are average numbers across all videos of the sport.

Method	Training Setup	HOTA↑	IDF1↑	AssA↑	MOTA↑	DetA↑	LocA↑	IDs↓	Frag↓
FairMOT [29]	Train	49.3	53.5	34.7	86.4	70.2	83.9	9928	21673
QDTrack [6]	Train	60.4	62.3	47.2	90.1	77.5	88.0	6377	11850
CenterTrack [33]	Train	62.7	60.0	48.0	90.8	82.1	90.8	10481	5750
TransTrack [20]	Train	68.9	71.5	57.5	92.6	82.7	91.0	4992	9994
BoT-SORT [1]	Train	68.7	70.0	55.9	94.5	84.4	90.5	6729	5349
ByteTrack [28]	Train	62.8	69.8	51.2	94.1	77.1	85.6	3267	4499
OC-SORT [3]	Train	71.9	72.2	59.8	94.5	86.4	92.4	3093	3474
ByteTrack [28]	Train+Val	64.1	71.4	52.3	95.9	78.5	85.7	3089	4216
OC-SORT [3]	Train+Val	73.7	74.0	61.5	**96.5**	**88.5**	**92.7**	2728	**3144**
MixSort-Byte [4]	Train+Val	65.7	74.1	54.8	96.2	78.8	85.7	**2472**	4009
MixSort-OC [4]	Train+Val	74.1	74.4	62.0	**96.5**	**88.5**	92.7	2781	3199
Deep-EIoU (Ours)	Train	74.1	75.0	63.1	95.1	87.2	92.5	3066	3471
Deep-EIoU (Ours)	Train+Val	**76.9**	**79.2**	**67.2**	96.2	88.1	92.4	3338	**3144**

Table 2: The performance comparison on the SportsMOT test set. Our algorithm outperforms all the popular and SOTA tracking algorithms in several major evaluation metrics. Part of the evaluation results are taken from the number reported in SportsMOT dataset paper [4].

3.4 Iterative Scale-Up ExpansionIoU

As illustrated by the previous work using expansion bounding box for association [26], the amount of the bounding box expansion is a crucial and sensitive hyperparameter in the tracking process and the performance of the tracker can be largely affected by the chosen of the hyperparameter. In the real world scenario, several factors might limit us from tuning the expansion scale and improve the tracking performance, including 1) the online and real time tracking requirements. One common requirement for athlete tracking system is the system needs to operate in an online and real time matter, tuning the expansion scale with experiments and twig the performance is not possible in such cases. 2) No access to the testing data. For real world scenario, the testing data's groundtruth is often not available, which makes finding the perfect expansion scale for association impossible. Due to the above reasons, we proposed a novel iterative scale-up ExpansionIoU association stage for robust tracking. Instead of doing hyperparameter tuning for the best expansion scale, we choose to iteratively conduct association based on a gradually bigger ExpansionIoU during the tracking process. By using this approach, we can first perform association to those trajectory and detection pairs with higher ExpansionIoU, and gradually search for those pairs with lower overlapping area, which enhances the robustness of our association process.

4 EXPERIMENTS AND RESULTS

4.1 Dataset

To show the effectiveness of our algorithm on athlete tracking, we evaluate our tracking algorithm on the new large scale multi-sports player tracking dataset SportsMOT [4]. SportsMOT consists of 240 video sequences with over 150K frames and over 1.6M bounding boxes collected from 3 different sports including basketball, football and volleyball. Different from the MOT dataset [5, 16], SportsMOT possess more difficulties including: 1) targets' fast and irregular motion, 2) larger camera movement, and 3) similar appearance between players in the same team.

4.2 Detector

We choose YOLOX [7] as our object detector to achieve real time and high accuracy in detection performance. Several existing popular tracking algorithm [1, 3, 26, 28] also incorporate YOLOX as their detector, this also leads to a more fair comparison between these algorithms with ours. For the pretrained weight, we use the COCO pretrained YOLOX-X model provided by the official GitHub repositories of YOLOX. We trained the model with Sportsmot training and validation set for 80 epochs, following the YOLOX-X default training process of ByteTrack's [28] official GitHub repositories.

4.3 ReID Model

For player re-identification, we use the omni-scale feature learning proposed in OSNet [32]. The unified aggregation gate fuse the features from different scales and enhance the ability in human Re-ID. We use the pre-trained model from the Market-1501 dataset [31] and further trained the model on the SportsMOT dataset, the ReID training data is constructed based on the original SportsMOT dataset where we cropped out each player according to its ground-truth annotation of the bounding boxes. The sampled dataset include 31,279 training images, 133 query images and 1,025 gallery images.

4.4 Tracking Setting

The confidence score threshold for high and low score detection is 0.6, while the confidence threshold for low score detection filtering is 0.1. The cost filter threshold τ_A and τ_{BIoU} is set to 0.25 and 0.5, respectively. We also remove the constraint of aspect ratio in the detection bounding box, since sports like soccer and basketball might have the condition when a player is lying on the ground, which is different from the MOT datasets where most of the pedestrians are standing and walking. The initial value of expansion scale E in the first round matching is set to 0.7 and with a step of 0.1 in each iteration, the expansion scale is 0.8 in the second round, λ is set to 2 and the tracklet age is 60 frames.

Figure 3: The tracking demo images from three different sports in the SportsMOT dataset. With the iterative scale-up ExpansionIoU and deep features association, our algorithm can achieve robust multi-athlete tracking under severe occlusion condition in multiple diverse sport scenarios including basketball, football and volleyball.

4.5 Performance

To evaluate the performance of our method, we compared our tracking algorithm with several existing methods and the SOTA tracking algorithms on the SportsMOT testing set. All the testing result is evaluated on the SportsMOT evaluation server. The performance of our proposed method achieved 76.9 in HOTA, 79.2 in IDF1, 67.2 in AssA, which outperforms all the other existing tracking algorithms, showing the effectiveness of our algorithm in multi-athlete tracking.

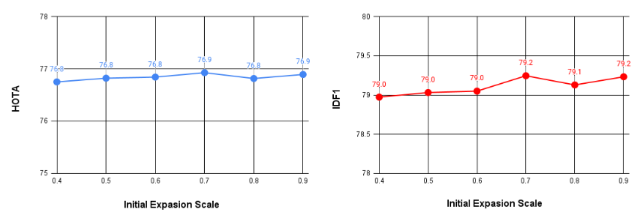

Figure 4: The performance comparison of different initial expansion scale on the SportsMOT test set. Different selection of initial expansion scale can always achieve SOTA performance in HOTA and IDF1, which demonstrate the robustness of our method.

5 ABLATION STUDIES

5.1 Robustness to initial expansion scale

To prove the effectiveness and robustness of our approach, we conduct experiment based on different initial expansion scale in the iterative scale-up process. We change the initial expansion scale from 0.4 to 0.9. The experiment results show that we can still achieve SOTA performance with different initial expansion scale because the iterative scale-up process can enhance the robustness and does not require any extra parameter tuning on the testing set. This proves our method's effectiveness in the real-world scenario, when groundtruth is often not available and the tracking parameter can not be tuned.

5.2 Combining Kalman Filter and EIoU

To test the effect ExpansionIoU on Kalman filter-based tracker, we also implemented several versions of our method by directly incorporating Kalman filter and ExpansionIoU. In our implementation, the Kalman filter is incorporated following the previous Kalman filter-based tracking methods [1, 28, 29]. The Kalman filter's prediction or the detection's bounding box will be expand for the association through ExpansionIoU, all the other parameters are the same.

Tracker	EIoU	HOTA
ByteTrack		62.8
ByteTrack	✓	67.5
BoTSORT		68.7
BoTSORT	✓	71.3

Table 3: We evaluated two popular Kalman filter based tracking algorithms including ByteTrack [28] and BoTSORT [1]. Experiment results show that Kalman filter based tracker can also benefited from incorporating ExpansionIoU during the tracking process.

6 CONCLUSIONS

In this paper, we proposed an iterative scale-up ExpansionIoU and deep features association method for multi-athlete tracking, which achieves 76.9 in HOTA, 79.2 in IDF1 and 96.2 in MOTA on the SportsMOT dataset, successfully tackle down the challenges of irregular movement during multi-athlete tracking. Our algorithm outperforms most of the existing tracking algorithms, showing the effectiveness and robustness of the proposed method.

REFERENCES

[1] Nir Aharon, Roy Orfaig, and Ben-Zion Bobrovsky. 2022. BoT-SORT: Robust associations multi-pedestrian tracking. *arXiv preprint arXiv:2206.14651* (2022).

[2] Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, and Ben Upcroft. 2016. Simple online and realtime tracking. In *2016 IEEE International Conference on Image Processing (ICIP)*. 3464–3468. https://doi.org/10.1109/ICIP.2016.7533003

[3] Jinkun Cao, Xinshuo Weng, Rawal Khirodkar, Jiangmiao Pang, and Kris Kitani. 2022. Observation-Centric SORT: Rethinking SORT for Robust Multi-Object Tracking. arXiv preprint arXiv:2203.14360.

[4] Yutao Cui, Chenkai Zeng, Xiaoyu Zhao, Yichun Yang, Gangshan Wu, and Limin Wang. 2023. SportsMOT: A Large Multi-Object Tracking Dataset in Multiple Sports Scenes. *arXiv preprint arXiv:2304.05170* (2023).

[5] Patrick Dendorfer, Hamid Rezatofighi, Anton Milan, Javen Shi, Daniel Cremers, Ian Reid, Stefan Roth, Konrad Schindler, and Laura Leal-Taixé. 2020. Mot20: A benchmark for multi object tracking in crowded scenes. *arXiv preprint arXiv:2003.09003* (2020).

[6] Tobias Fischer, Jiangmiao Pang, Thomas E Huang, Linlu Qiu, Haofeng Chen, Trevor Darrell, and Fisher Yu. 2022. QDTrack: Quasi-Dense Similarity Learning for Appearance-Only Multiple Object Tracking. *arXiv preprint arXiv:2210.06984* (2022).

[7] Zheng Ge, Songtao Liu, Feng Wang, Zeming Li, and Jian Sun. 2021. Yolox: Exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430.

[8] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. 2013. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research* 32, 11 (2013), 1231–1237.

[9] Hsiang-Wei Huang, Cheng-Yen Yang, and Jenq-Neng Hwang. 2023. Multi-target multi-camera vehicle tracking using transformer-based camera link model and spatial-temporal information. *arXiv preprint arXiv:2301.07805* (2023).

[10] Hsiang-Wei Huang, Cheng-Yen Yang, Jenq-Neng Hwang, Pyong-Kun Kim, Kwangju Kim, and Kyoungoh Lee. 2022. Observation Centric and Central Distance Recovery on Sports Player Tracking. *arXiv preprint arXiv:2209.13154* (2022).

[11] Hsiang-Wei Huang, Cheng-Yen Yang, Zhongyu Jiang, Pyong-Kun Kim, Kyoungoh Lee, Kwangju Kim, Samartha Ramkumar, Chaitanya Mullapudi, In-Su Jang, Chung-I Huang, et al. 2023. Enhancing Multi-Camera People Tracking with Anchor-Guided Clustering and Spatio-Temporal Consistency ID Re-Assignment. *arXiv preprint arXiv:2304.09471* (2023).

[12] Hsiang-Wei Huang, Cheng-Yen Yang, Samartha Ramkumar, Chung-I Huang, Jenq-Neng Hwang, Pyong-Kun Kim, Kyoungoh Lee, and Kwangju Kim. 2023. Observation Centric and Central Distance Recovery for Athlete Tracking. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 454–460.

[13] R. E. Kalman. 1960. A new approach to linear filtering and prediction problems. J. Fluids Eng., 82(1):35–45.

[14] Hao Luo, Youzhi Gu, Xingyu Liao, Shenqi Lai, and Wei Jiang. 2019. Bag of tricks and a strong baseline for deep person re-identification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. 0–0.

[15] Adrien Maglo, Astrid Orcesi, and Quoc-Cuong Pham. 2022. Efficient tracking of team sport players with few game-specific annotations. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (2022), 3460–3470.

[16] A. Milan, L. Leal-Taixe, I. Reid, S. Roth, and K. Schindler. 2016. Mot16: A benchmark for multi-object tracking. arXiv preprint arXiv:1603.00831.

[17] Adrià Arbués Sangüesa, Coloma Ballester, and Gloria Haro. 2019. Single-Camera Basketball Tracker through Pose and Semantic Feature Fusion. *CoRR* abs/1906.02042 (2019). arXiv:1906.02042 http://arxiv.org/abs/1906.02042

[18] Robert Stone et al. 2000. CenterTrack: An IP overlay network for tracking DoS floods.. In *USENIX Security Symposium*, Vol. 21. 114.

[19] Peize Sun, Jinkun Cao, Yi Jiang, Zehuan Yuan, Song Bai, Kris Kitani, and Ping Luo. 2022. Dancetrack: Multi-object tracking in uniform appearance and diverse motion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 20993–21002.

[20] Peize Sun, Jinkun Cao, Yi Jiang, Rufeng Zhang, Enze Xie, Zehuan Yuan, Changhu Wang, and Ping Luo. 2020. Transtrack: Multiple object tracking with transformer. *arXiv preprint arXiv:2012.15460* (2020).

[21] Kanav Vats, Pascale Walters, Mehrnaz Fani, David A Clausi, and John S. Zelek. 2021. Player Tracking and Identification in Ice Hockey. *ArXiv* abs/2110.03090 (2021).

[22] Zhongdao Wang, Liang Zheng, Yixuan Liu, Yali Li, and Shengjin Wang. 2020. Towards real-time multi-object tracking. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16*. Springer, 107–122.

[23] Nicolai Wojke, Alex Bewley, and Dietrich Paulus. 2017. Simple online and realtime tracking with a deep association metric. In 2017 IEEE international conference on image processing(ICIP), pages 3645–3649. IEEE.

[24] Cheng-Yen Yang, Alan Yu Shyang Tan, Melanie J. Underwood, Charlotte Bodie, Zhongyu Jiang, Steve George, Karl Warr, Jenq-Neng Hwang, and Emma Jones. 2023. Multi-Object Tracking by Iteratively Associating Detections with Uniform Appearance for Trawl-Based Fishing Bycatch Monitoring. arXiv:2304.04816 [cs.CV]

[25] Fan Yang, Shigeyuki Odashima, Shoichi Masui, and Shan Jiang. 2022. The Second-place Solution for ECCV 2022 Multiple People Tracking in Group Dance Challenge. *arXiv preprint arXiv:2211.13509* (2022).

[26] Fan Yang, Shigeyuki Odashima, Shoichi Masui, and Shan Jiang. 2023. 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*. 4799–4808.

[27] Yukun Yang, Ruiheng Zhang, Wanneng Wu, Yu Peng, and Min Xu. 2021. Multi-camera Sports Players 3D Localization with Identification Reasoning. *2020 25th International Conference on Pattern Recognition (ICPR)* (2021), 4497–4504.

[28] Yifu Zhang, Peize Sun, Yi Jiang, Dongdong Yu, Zehuan Yuan, Ping Luo, Wenyu Liu, and Xinggang Wang. 2021. Bytetrack: Multi-object tracking by associating every detection box. arXiv preprint arXiv:2110.06864.

[29] Yifu Zhang, Chunyu Wang, Xinggang Wang, Wenjun Zeng, , and Wenyu Liu. 2021. Fairmot: On the fairness of detection and re-identification in multiple object tracking. International Journal of Computer Vision, 129(11):3069–3087.

[30] Yifu Zhang, Chunyu Wang, Xinggang Wang, Wenjun Zeng, and Wenyu Liu. 2021. Fairmot: On the fairness of detection and re-identification in multiple object tracking. *International Journal of Computer Vision* 129 (2021), 3069–3087.

[31] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. 2015. Scalable Person Re-identification: A Benchmark. In *Computer Vision, IEEE International Conference on*.

[32] Kaiyang Zhou, Yongxin Yang, Andrea Cavallaro, and Tao Xiang. 2019. Omni-scale feature learning for person re-identification. Proceedings of the IEEE/CVF International Conference on Computer Vision.

[33] Xingyi Zhou, Vladlen Koltun, and Philipp Krähenbühl. 2020. Tracking Objects as Points. *ArXiv* abs/2004.01177 (2020).

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009