Jump-Aware: Player Position Rectification and Identification in Dynamic Sports Using Jump Event Spotting

Yin May Oo[1,2] Ankhzaya Jamsrandorj[2] Vanyi Chao[1,2] Hoang Quoc Nguyen[1,2]
Yewon Hwang[2] Kyung-Ryoul Mun[2] Jinwook Kim[2]
[1]AI-Robotics, KIST School, University of Science and Technology, Seoul 02792, Republic of Korea
[2]Korea Institute of Science and Technology, Seoul 02792, Republic of Korea
{yinmay, ankhzaya, vanyi}@kist.re.kr, nqhoang@ust.ac.kr
{ywhwang, krmoon02}@kist.re.kr, jinwook.kim21@gmail.com

Abstract

Accurate player positioning and identification in broadcast sports videos are essential for sports analytics. However, jump-intensive sports such as basketball and volleyball pose significant challenges due to positional distortions caused by airborne motion and occlusions. To address these issues, we introduce Jump-Aware Position Rectification (JPR), a framework that integrates Jump Event Spotting (JES) and jersey-based player identification to improve spatial consistency and identity tracking. Our method first detects and validates jump events, then rectifies player positions in a top-view pitch coordinate system, reducing motion artifacts caused by temporary elevation changes in 2D image space. Additionally, jersey-based identification enhances identity tracking by leveraging jersey numbers, even under occlusions. To support our research, we present SportsJumpMotion, a dataset featuring frame-accurate jump annotations and jersey-based player identities for basketball and volleyball. Experimental results demonstrate that our JES model achieves a mean Average Precision (mAP) of 95.33, outperforming baseline models in jump event spotting. Furthermore, fine-tuning on sport-specific datasets significantly improves jersey-based identification, addressing variations in jersey visibility and motion patterns across sports. Our dataset and framework provide a comprehensive benchmark for advancing player tracking in dynamic sports scenarios. Our SportsJumpMotion dataset is publicly available at https://github.com/yinmayoo185/SportsJumpMotion.

1. Introduction

Advancements in computer vision algorithms have significantly transformed sports analytics, enabling automated player tracking, tactical evaluation, and performance assessment. Vision-driven analytics play a critical role in the formulation of game strategies, coaching decisions, and evaluation of athletes, offering data-driven insights to enhance competitive advantage. Among these tasks, accurate player positioning is particularly crucial, as it directly influences team formation, tactical strategies, and real-time game analysis.

Unlike multi-object tracking (MOT) systems used in pedestrian and vehicle scenarios, sports tracking presents unique challenges due to the dynamic nature of player movements and frequent interactions. Players experience severe occlusions, visual similarities between teammates, and rapid high-speed actions, such as basketball crossovers or volleyball spikes. One of the fundamental issues in player tracking is identity switch, where frequent occlusions, uniform similarities, and motion blur disrupt the continuity of identity assignment.

Another key limitation of existing tracking methods is their inability to handle airborne movements, which are common in jump-intensive sports. When players jump, their 2D projected positions shift significantly, causing spatial distortions in broadcast videos. While homography based sports field registration [7, 18, 21, 22] effectively maps player positions from image space to real-world pitch coordinates, it assumes a planar surface and fails to account for vertical displacements. As a result, current methods misrepresent player positions during jumps, reducing the reliability of trajectory and movement analysis.

Player identification is also essential for maintaining consistent identity labels throughout the game. However, in team sports, visually similar uniforms pose challenges for long-term re-identification (ReID) methods, especially under occlusions or motion blur. In contrast, jersey numbers provide a stable and robust identifier that remains constant throughout the game. Since jersey numbers are prominently displayed and visible across frames, recognizing them can significantly enhance player identification, even in crowded or occluded scenarios.

To address these challenges, we propose Jump-Aware Position Rectification (JPR), a framework that integrates Jump Event Spotting (JES) and jersey-based player identification for player tracking. Our approach explicitly detects jumping and landing frames, then corrects spatial inconsistencies using a top-view pitch coordinate system, effectively mitigating motion artifacts caused by airborne movements. Additionally, we leverage a Vision-Language Model (VLM) to assign unique identities to players, significantly improving robustness against occlusions and visually ambiguous conditions. Our key contributions can be summarized as follows:

1. Jump-Aware Rectification: We introduce a rectification mechanism that corrects spatial distortions caused by airborne movements, substantially improving the reliability of player localization.
2. Integrated Tracking Framework: We present a unified framework that combines player detection, tracking, sports field registration, jump event spotting, jump-aware rectification, and jersey-based identification, enabling robust and scalable player tracking for sports analytics.
3. SportsJumpMotion Dataset: We release SportsJumpMotion, the first dataset with frame-accurate jump annotations and jersey-based player identities for basketball and volleyball.

2. Related Works

This section briefly reviews the related work for the modules integrated into our framework to estimate player positions and identities.

2.1. Player Detection and Tracking

Multi-object tracking (MOT) in sports analytics follows the tracking-by-detection paradigm, where object detectors such as Faster R-CNN [23] and YOLO-X [10] detect players in each frame. DETR [4] replaced traditional region proposal-based methods with an end-to-end transformer architecture, showed promising results.

For tracking, motion-based approaches typically incorporate the Kalman filter [19] to predict player positions and use the Hungarian algorithm for association. SORT [2] employs IoU-based matching with Kalman filtering but struggles in occlusion-heavy scenarios. ByteTrack [31] refines associations by prioritizing high-confidence detections before handling low-confidence ones, improving robustness. OC-SORT [3] enhances motion modeling by mitigating non-linear motion errors, and DeepSORT [27] integrates appearance-based ReID features for improved identity tracking. Recently, Deep-EIoU [16] introduced an iterative expansion IoU-based association, reducing computational costs while outperforming Kalman filter-based methods. Huang *et al*. [15] combined OC-SORT with appearance-based refinements for enhanced tracking across sports, demonstrating state-of-the-art performance on SportsMOT [9] across different sports scenarios.

2.2. Player Identification

Consistent player identification is crucial for tracking individuals across frames and viewpoints. Recent appearance-based ReID [32, 33] models struggle with occlusions and visually similar uniforms. Jersey number recognition provides a more structured approach, directly linking players to their identities. Early methods, such as Chan *et al*. [5], used CNN+LSTM networks to predict jersey numbers but faced challenges when digits were obscured. Balaji *et al*. [1] introduced a Keyframe Identification (KfID) module to select frames where jersey numbers were most visible, improving recognition in low-resolution broadcasts. Transformer-based methods have further advanced player identification. Vats *et al*. [26] introduced a weakly supervised transformer-based framework for ice hockey, handling partial occlusions. Koshkina *et al*. [20] adapted scene text recognition models to extract jersey numbers without extensive manual annotation, refining predictions across tracklets. Habel *et al*. [11] explored CLIP-based embeddings for jersey number retrieval, highlighting the potential of vision-language models in sports analytics. Despite these advancements, publicly available player identification datasets remain scarce, with most relying on tracklets and partial annotations. This underscores the importance of integrating appearance-based ReID with jersey number recognition to ensure identity consistency in sports scenarios with heavy occlusions.

2.3. Event Spotting

Event spotting is crucial for understanding sports videos, as it localizes specific moments using a single timestamp. In jump-intensive sports, even small temporal errors (4-5 frames) can lead to incorrect motion interpretations. E2E-Spot [12] introduced an end-to-end learning framework that jointly trains feature extraction and event spotting heads, improving temporal precision. T-Deed [28] further advanced event spotting by enhancing token discriminability and multi-scale temporal reasoning.

2.4. Player Position Estimation

Public datasets for player position estimation remain scarce. SoccerNet-GSR [24] proposed Game State Reconstruction (GSR), integrating athlete localization and identification on a pitch minimap from soccer videos. Theiner *et al*. [25] developed a modular pipeline for 2D soccer player localization without focusing on identity tracking. Chen *et al*. [6] proposed a physics-based method for jump pattern recognition, estimating jump height and location but requiring prior knowledge of player heights and 3D pitch transformation,

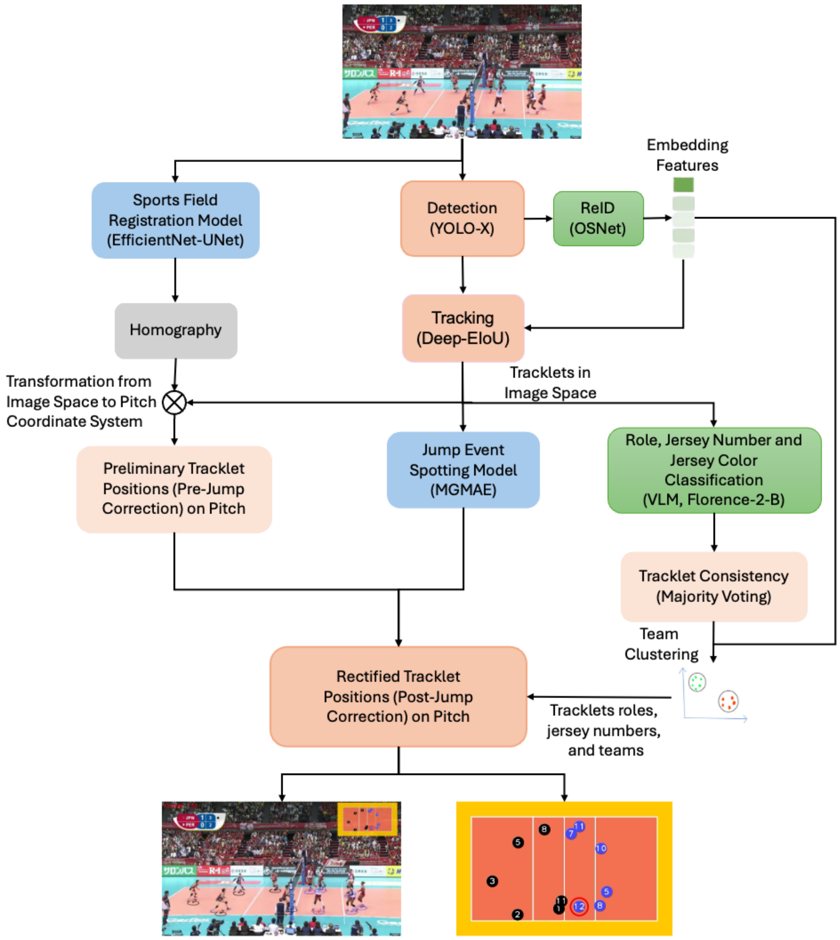


Figure 1. Overview of our proposed Jump-Aware Position Rectification (JPR) framework.

which is often unavailable in broadcast footage in volleyball. Hsu *et al*. [13] extended this by tracking jump trajectories over long video sequences, enabling motion pattern recognition in volleyball games.

3. Method

This section presents our Jump-Aware Position Rectification (JPR) framework, illustrated in Figure 1, which improves player tracking in jump-intensive sports by correcting positional distortions and identity inconsistencies. Broadcast sports videos introduce challenges such as camera motion, occlusions, and jersey similarities, making accurate tracking difficult. To address these, our framework integrates five key modules, each ensuring accurate spatial localization and identity tracking. **Player Detection and Tracking** extracts player positions in image space, while **Sports Field Registration** maps them to pitch coordinates. To rectify airborne distortions, **Jump Event Spotting (JES)** detects jumping and landing frames, guiding the **Jump-Aware Rectification** module to refine player trajectories. Finally, **Jersey-Based Identification and Team Clustering** ensures identity consistency across frames, mitigating occlusion and motion blur. Together, these modules enable robust tracking in dynamic sports environments.

3.1. Player Detection and Tracking

We adopt a tracking-by-detection paradigm to localize players across frames while ensuring consistent identity association. We employ YOLO-X [10] for efficient and robust multi-object detection, followed by Deep-EIoU [16], which enhances tracking robustness through iterative bounding-box expansion and deep feature matching, improving identity association under occlusions and rapid player movements. To further ensure identity consistency, OSNet [32] is used for appearance-based re-identification, generating discriminative embeddings that enhance short-term identity association, even under occlusions or jersey similarities. The resulting tracklets are passed as input for both jump event spotting module (Section 3.3) and jersey-based player identification and team clustering module (Section 3.5). The

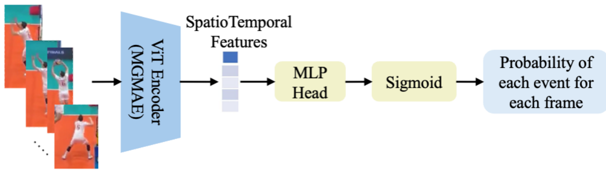


Figure 2. Architecture of our Jump Event Spotting (JES) model.

detected player positions are transformed into pitch coordinates using sports field registration (Section 3.2).

3.2. Sports Field Registration

To transform player positions from image space to a top-view pitch representation, we integrate the sports field registration model from [22]. This module estimates a homography matrix H using a Residual Attention-based EfficientNet-UNet [22], which detects pre-defined pitch keypoints. We represent each player's location using the bottom-center point of their bounding box, a standard approximation that typically corresponds to the feet position. The estimated H transforms image-space coordinates $\tilde{p} \in \mathbb{R}^2$ into pitch coordinates $\hat{p} \in \mathbb{R}^2$:

$$\hat{p} = H^{-1}\tilde{p} \tag{1}$$

where $H \in \mathbb{R}^{3\times3}$ is the homography matrix.

These transformed positions are referred to as Preliminary Tracklet Positions (Pre-Jump Correction), providing an initial localization of players on the pitch. However, the assumption of a planar surface in homography does not account for airborne motion, leading to spatial displacement and positional distortions. To address this, jump event spotting module (Section 3.3) detects jumping and landing frames, enabling the jump-aware rectification process to refine player trajectories, ensuring reliable localization on the pitch.

3.3. Jump Event Spotting (JES)

To correct positional distortions from airborne motion, we introduce the **JES** model that automatically detects jumping and landing frames for jump-aware rectification. The model utilizes a Motion Guided Masking Autoencoder (MGMAE) encoder [14] to extract spatiotemporal representations from player tracklet sequences, capturing motion dynamics and structural cues related to jump transitions. The extracted features are processed through a MLP Head, followed by a Sigmoid activation, producing frame-wise probabilities for jumping and landing events:

$$\mathbf{P} = \sigma(\mathrm{MLP}(\mathrm{MGMAE}(\mathbf{F}; \theta^{MGMAE}); \theta^{MLP})) \tag{2}$$

where θ^{MGMAE} and θ^{MLP} are the trainable parameters of the MGMAE encoder and MLP head, respectively, $\mathbf{P} \in \mathbb{R}^{T\times2}$ represents event probabilities per frame, and $\mathbf{F} \in \mathbb{R}^{C\times T\times H\times W}$ denotes the input tracklet frames.

The jump-aware rectification process (Section 3.4) leverages detected jump events, which are identified by analyzing peaks in the JES probability outputs. This correction mitigates trajectory distortions caused by elevation changes. Figure 2 illustrates the JES model architecture.

3.4. Rectified Tracklet Positions (Post-Jump Correction)

Following JES, we refine player positions using physics-based validation to assess jump plausibility and improve trajectory consistency in the reconstructed pitch coordinates. After predicting the jumping f_J and landing f_L frame indices for each tracklet, we compute the jump duration t_J:

$$t_J = (f_L - f_J)\,\Delta t \tag{3}$$

where Δt is the per-frame time interval. Assuming a symmetric parabolic trajectory, where ascent and descent are of equal duration, we set the landing height to zero and calculate the theoretical maximum jump height h_J using:

$$h_J = \frac{1}{8} g t_J^2 \tag{4}$$

where $g = 9.81\,\mathrm{m/s^2}$ is the gravitational constant. By comparing h_J to pre-defined jump height thresholds $[h_{min}, h_{max}]$, we discard implausible or minor "hops" and retain only validated jumps. For instance, studies report that elite volleyball players may reach up to 120 cm [6, 13], while basketball jump heights often fall within 22–48 cm for female and 40–75 cm for male players [34]. Adjusting these thresholds accordingly allows the approach to accommodate different sports.

We refine airborne positions by interpolating between jumping and landing frames, smoothing discontinuities and preserving spatial coherence in pitch coordinates. Only physically plausible jumps are retained while reconstructing airborne trajectories. This rectification enhances 2D tracking accuracy and estimates jump height, a key metric for performance and tactical analysis in dynamic sports.

3.5. Jersey-Based Player Identification and Team Clustering

Florence-2-B [29], a lightweight Vision-Language Model (VLM), is used to predict structured text for jersey-based player identification. In the SoccerNet Fine-tuned dataset, the model classifies each individual's role (Player, Goalkeeper, Referee), jersey number, and jersey color directly from bounding-box images. In contrast, the SportsJump-Motion dataset contains only player tracklets and does not include role annotations, so role classification is not

Table 1. Statistics of the SportsJumpMotion dataset.

Split	Basketball Clips	Volleyball Clips	Num of Tracklets	Jumping Events	Landing Events
Train	26	99	1448	980	979
Valid	14	40	610	439	439
Test	12	44	648	463	458

performed for this dataset. To maintain identity consistency, we apply a majority-voting mechanism across frames within each tracklet, ensuring that occasional misclassifications do not disrupt long-term identity assignment. For team clustering, players are grouped based on jersey color and visual embeddings extracted from OSNet when jersey color alone is insufficient, while pitch position is used in volleyball to distinguish libero players.

4. Experiments

4.1. Datasets

We evaluate our framework on two datasets: SportsJumpMotion, introduced in this work, and SoccerNet Fine-tuned, which is extended from an existing benchmark.

4.1.1. SportsJumpMotion Dataset

SportsJumpMotion dataset consists of player tracklets with tracklet-level annotations. We collected basketball and volleyball broadcast video clips from SportsMOT [9], providing both raw frame images and bounding-box tracklet images (Figure 3(a)) for jump event spotting and jersey number classification. Each tracklet is manually annotated with frame-accurate jumping and landing events, along with jersey numbers, as shown in Figure 3(c). The dataset is divided into train, validation, and test sets for model evaluation. Table 1 presents the dataset statistics, and Table 2 describes the event types.

4.1.2. SoccerNet Fine-tuned Dataset

Since SoccerNet Jersey Number Recognition dataset [8] provides only jersey number annotations, we manually annotate additional attributes, including the role (Player, Goalkeeper, Referee, Other, Ball) and dominant jersey color for jersey-based player identification as illustrated in Figure 4.

Table 2. Event descriptions in the SportsJumpMotion dataset.

Event	Description
Jumping	When the tracklet starts jumping (both feet leave the ground).
Landing	When one of the tracklet's feet is back on the ground.

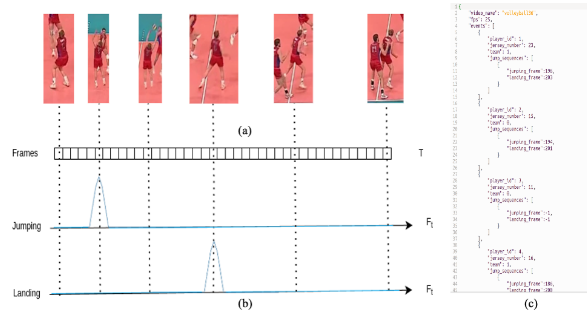


Figure 3. (a) Sample frames from the ground truth annotation of a video clip (b) Target construction for jump event spotting for each tracklet (c) Corresponding ground truth annotation for the example video clip.

For evaluation, we assess jersey number classification and role classification, while jersey color is used solely for team clustering.

4.2. Target Construction for Jump Event Spotting (JES)

The annotated events in the SportsJumpMotion dataset are micro-events, meaning they occur within a single-frame duration, making precise temporal localization crucial. However, distinguishing jumping and landing events is challenging due to the high similarity in pixel content between adjacent frames. To address this, we employ a Gaussian-based smoothing strategy. Specifically, for each event frame, we generate a probability distribution, centered on the annotated frame, forming two Gaussian curves (one for jumping and one for landing), as shown in Figure 3(b). This approach helps reduce misclassification risks and allows our model to learn more robust temporal patterns.

4.3. Implementation Details

We conducted the experiments on a single NVIDIA 4090 GPU with 24G memory size and implemented our approach in PyTorch. The implementation details for each module are explained as follows:

4.3.1. Player Detection and Tracking

For player detection, we adopt the YOLO-X model trained on SportsMOT from [9]. For tracking, we follow the configuration settings of the Deep-EIoU, which focuses on MOT for sports scenarios and integrates with the embedding features from the OSNet model that is trained on SportsMOT [16].

4.3.2. Sports Field Registration

For Residual Attention-based EfficientNet-UNet, we use weights trained on the Volleyball dataset [17] and the Basketball dataset [30] for sport-specific field registration.

Table 3. Comparison of Jump Event Spotting performance on the SportsJumpMotion dataset. Our model outperforms all baselines in mAP and class-wise AP at different temporal tolerances ($\delta = 1, 2, 3$).

Method	**Backbone**	**mAP**			**AP (Jumping)**			**AP (Landing)**		
		$\delta = 1$	$\delta = 2$	$\delta = 3$	$\delta = 1$	$\delta = 2$	$\delta = 3$	$\delta = 1$	$\delta = 2$	$\delta = 3$
E2E-Spot [12]	RegNet-Y	89.39	92.99	94.05	90.88	93.57	94.32	87.89	92.41	93.78
T-Deed [28]	RegNet-Y	89.36	92.90	94.48	91.10	94.17	95.03	87.61	91.63	93.93
Ours	ViT-B	**90.80**	**94.41**	**95.33**	**91.60**	**95.29**	**95.78**	**89.99**	**93.52**	**94.88**

4.3.3. Jump Event Spotting (JES)

The input is 16 frames of each tracklet and resized to 224x224 pixels, and the JES model is trained for 60 epochs using AdamW (learning rate $3e^{-4}$, batch size of 8). We approach jump event spotting as a regression problem using a sliding-window method with a frame stride of step S. In our experiments, we set S=5, which provides a balance between temporal resolution and computational efficiency. We use Mean Squared Error (MSE) loss. Data augmentations, such as random cropping, color jittering, horizontal flipping, and random rotation within $\pm 10°$, were applied consistently across frame sequences rather than individual frames, preserving temporal coherence. During inference, we identify predicted events by analyzing the probability distributions of output by the model. Peaks in these distributions are detected at a threshold of 0.6 to determine the frames containing jump events.

4.3.4. Jersey-Based Player Identification and Team Clustering

To adapt Florence-2-B for sports broadcast videos, we fine-tune the model separately on small training subsets of the SportsJumpMotion and SoccerNet Fine-tuned datasets. We apply identical training settings (5 epochs, learning rate $2e^{-4}$) to evaluate its generalization across different sports. During inference, each bounding-box image is processed with the prompt: "Classify Role, Jersey Number, and Jersey Color". If the jersey number is not visible, we assign the placeholder "100" to denote an unreadable label. Since Florence-2-B already demonstrates strong OCR capabilities, we apply minimal fine-tuning to adapt its output to the **"Role,Jersey Number,Jersey Color"** format, while preserving its pre-trained text recognition strengths.

4.4. Evaluation Metrics

4.4.1. Jump Event Spotting (JES)

The prediction accuracy is assessed using Average Precision (AP) with a tolerance of δ frames (AP @δ). AP is computed for each event class and mean Average Precision (mAP) is the mean across classes. We emphasize tight tolerances, ($\delta = 1, 2, 3$), to ensure a precise jump event spotting accuracy evaluation.

4.4.2. Jersey-Based Player Identification

For jersey-based player identification, we evaluate jersey number and role classification at the tracklet-level using classification accuracy and F1-score. A prediction is considered correct if the majority-voted jersey number for a tracklet matches the ground truth.

4.5. Results

We evaluate our framework's effectiveness through both quantitative and qualitative analysis, focusing on jump event spotting and jersey-based player identification.

4.5.1. Quantitative Evaluation

As shown in Table 3, our JES model achieves the highest mAP of 95.33 at $\delta = 3$, outperforming E2E-Spot and T-Deed. It also records the highest Jumping AP and Landing AP, demonstrating superior capture of jump-related motion. Our model outperforms T-Deed by 0.85% in mAP and 0.75% in Jumping AP at $\delta = 3$, highlighting its effectiveness in improving motion consistency. While T-Deed performs competitively, it exhibits lower Landing AP, sug-

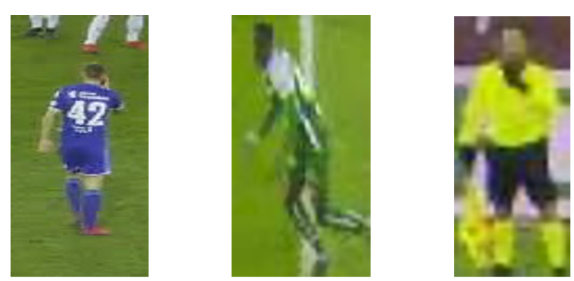


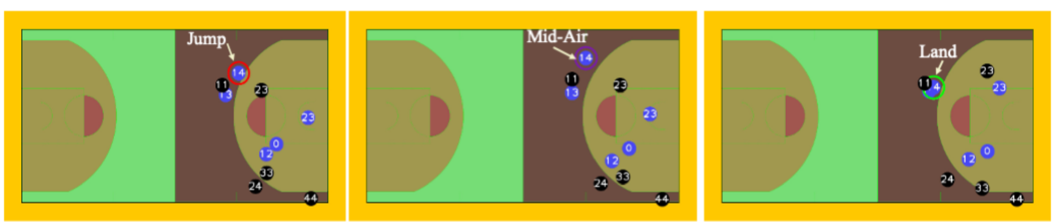
Figure 4. Examples images and ground truth annotations from the SoccerNet Fine-tuned Dataset used for fine-tuning Florence-2-B. The first image shows a player with a visible jersey number, the second depicts a player with an invisible number (placeholder: "100"), and the third represents a referee with role and jersey color annotations. These annotations support jersey-based identification.

(a) Visualization with rectified tracklet positions (post-jump correction)

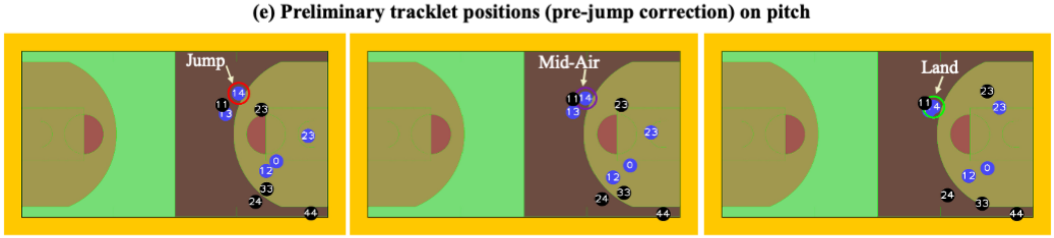
(b) Preliminary tracklet positions (pre-jump correction) on pitch

(c) Rectified tracklet positions (post-jump correction) on pitch

(d) Visualization with rectified tracklet positions (post-jump correction)



(e) Preliminary tracklet positions (pre-jump correction) on pitch



(f) Rectified tracklet positions (post-jump correction) on pitch

Figure 5. Qualitative results of our JPR framework. (a) and (d) show rectified tracklet positions overlaid on broadcast video frames, where red triangles represent reprojected rectified positions, aligning airborne players with their estimated true locations after jump correction. (b) and (e) display preliminary tracklet positions before jump correction on the pitch, with jersey numbers and team clustering applied for identity tracking while (c) and (f) visualize rectified tracklet positions after jump correction, with mid-air positions reconstructed. Red circles indicate jumping events, green circles mark landing events, and mid-air positions are captured in between. Jersey-based player identification ensures consistent identity tracking, while team clustering distinguishes opposing teams.

Table 4. Jersey-Based Player Identification Performance. Evaluation of Florence-2-B on SoccerNet Fine-tuned and SportsJumpMotion test sets, reporting jersey number classification accuracy and F1-score. Role classification is omitted for SportsJumpMotion as it includes only player tracklets.

Method	Dataset	Jersey Number Classification		Role Classification	
		Acc	F1	Acc	F1
Florence-2-B (Fine-tuned on SportsJumpMotion)	SoccerNet Fine-tuned	61.23	43.52	-	-
	SportsJumpMotion	76.38	54.99	-	-
Florence-2-B (Fine-tuned on SoccerNet)	SoccerNet Fine-tuned	84.69	56.52	98.93	95.77
	SportsJumpMotion	69.29	51.77	-	-

gesting limitations in detecting landing frames precisely. In contrast, our model improves Landing AP by 0.95%, further validating the robustness of our approach.

For jersey-based player identification, we evaluate Florence-2-B on both SportsJumpMotion and SoccerNet Fine-tuned test sets, measuring accuracy and F1-score at the tracklet-level (Table 4). Fine-tuning on SportsJumpMotion achieves 76.38% accuracy (54.99 F1-score) on its own test set but struggles on SoccerNet Fine-tuned (61.23% accuracy, 43.52 F1-score). Conversely, fine-tuning on SoccerNet Fine-tuned results in 84.69% accuracy (56.52 F1-score) on its own test set but generalizes less effectively to SportsJumpMotion (69.29% accuracy, 51.77 F1-score). Role classification is omitted for SportsJumpMotion, which contains only player tracklets, but the model achieves 98.93% accuracy (95.77 F1-score) on SoccerNet Fine-tuned. These results highlight the challenges of cross-sport jersey number classification. The performance drop when transferring the model from SoccerNet Fine-tuned to SportsJumpMotion highlights the impact of dataset-specific variations in jersey visibility, motion blur, and occlusion frequency on jersey recognition accuracy. This underscores the need for sport-specific adaptation strategies, particularly when applying models trained on one domain to another.

4.5.2. Qualitative Evaluation

Direct evaluation of rectified player positions against ground truth pitch coordinates is not possible due to the lack of sensor-based ground truth positional data in broadcast videos. Instead, we assess the effectiveness of our rectification through qualitative visualization and spatial consistency improvements. Figure 5 presents qualitative results of our JPR framework applied to basketball and volleyball. Figure 5(a) and (d) show the broadcast video frames with rectified tracklet positions reprojected as red triangles, illustrating how jump-aware correction aligns airborne players with their actual locations. Meanwhile, (b) and (e) show preliminary tracklet positions before jump correction, while (c) and (f) visualize the rectified positions post-correction on a top-view pitch with jersey numbers. The red circles highlight jumping event frames, green circles denote landing events, and mid-air positions are captured in between. The rectification process produces smoother positional transitions, reducing distortions caused by airborne motion. Furthermore, jersey-based identification helps maintain identity consistency, ensuring that detected jersey numbers align with players across frames. The framework also facilitates team clustering, effectively distinguishing opposing teams using jersey colors. These visualizations demonstrate the robustness of our approach in addressing occlusions, motion ambiguities, and positional inconsistencies. In the supplementary material, we also demonstrate additional broadcast videos in volleyball and basketball.

5. Conclusion and Future Work

We introduced a jump-aware player positioning and identification framework that mitigates positional distortions caused by airborne motion in dynamic sports analytics. By integrating JES with jersey-based player identification, our framework ensures smooth spatial transitions and robust identity tracking, even under occlusions. This enables structured player tracking and effective team clustering in dynamic sports scenarios. To validate our approach, we introduced SportsJumpMotion, a dataset with frame-accurate jump annotations and jersey-based player identities for basketball and volleyball. Experimental results demonstrate that our method outperforms baselines in jump event spotting, while domain-specific fine-tuning enhances identity consistency across frames in jersey-based identification.

Despite these advancements, our framework assumes predictable airborne motion and relies on homography-based localization, which introduces perspective distortions under extreme camera angles. Future work could explore learning-based rectification models to improve adaptability across diverse broadcast settings and mitigate perspective variation errors. Additionally, more advanced domain adaptation strategies could further improve jersey-based identification across different sports. Evaluating the JPR framework's impact on tracking accuracy would also provide deeper insights into its effectiveness in sports analytics. By addressing these challenges, future work can enhance player localization and identity tracking, advancing automated sports analysis in real-world broadcast environments.

6. Acknowledgement

This work was supported by Athletes' training/matches data management and AI-based performance enhancement solution technology development Project (No.1375027432) and Korea Institute of Science and Technology (KIST) Institutional Program (No. 2E33841)

References

[1] Bavesh Balaji, Jerrin Bright, Harish Prakash, Yuhao Chen, David A Clausi, and John Zelek. Jersey number recognition using keyframe identification from low-resolution broadcast videos. In *Proceedings of the 6th International Workshop on Multimedia Content Analysis in Sports*, pages 123–130, 2023. 2

[2] Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, and Ben Upcroft. Simple online and realtime tracking. In *2016 IEEE international conference on image processing (ICIP)*, pages 3464–3468. IEEE, 2016. 2

[3] Jinkun Cao, Jiangmiao Pang, Xinshuo Weng, Rawal Khirodkar, and Kris Kitani. Observation-centric sort: Rethinking sort for robust multi-object tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9686–9696, 2023. 2

[4] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *European conference on computer vision*, pages 213–229. Springer, 2020. 2

[5] Alvin Chan, Martin D. Levine, and Mehrsan Javan. Player identification in hockey broadcast videos. *Expert Systems with Applications*, 165:113891, 2021. 2

[6] Hua-Tsung Chen, Kuo-Lian Ma, Jen-Hui Chuang, and Horng-Horng Lin. Recognizing jump patterns with physics-based validation in human moving trajectory. *Journal of visual communication and image representation*, 24(7):1191–1203, 2013. 2, 4

[7] Yen-Jui Chu, Jheng-Wei Su, Kai-Wen Hsiao, Chi-Yu Lien, Shu-Ho Fan, Min-Chun Hu, Ruen-Rone Lee, Chih-Yuan Yao, and Hung-Kuo Chu. Sports field registration via keypoints-aware label condition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 1

[8] Anthony Cioppa, Adrien Deliège, Silvio Giancola, Bernard Ghanem, and Marc Van Droogenbroeck. Scaling up soccernet with multi-view spatial localization and re-identification. *Scientific Data*, 9, 2022. 5

[9] Yutao Cui, Chenkai Zeng, Xiaoyu Zhao, Yichun Yang, Gangshan Wu, and Limin Wang. Sportsmot: A large multi-object tracking dataset in multiple sports scenes. *arXiv preprint arXiv:2304.05170*, 2023. 2, 5

[10] Zheng Ge, Songtao Liu, Feng Wang, Zeming Li, and Jian Sun. Yolox: Exceeding yolo series in 2021. *arXiv preprint arXiv:2107.08430*, 2021. 2, 3

[11] Konrad Habel, Fabian Deuser, and Norbert Oswald. Clip-reident: Contrastive training for player re-identification. In *Proceedings of the 5th International ACM Workshop on Multimedia Content Analysis in Sports*, pages 129–135, 2022. 2

[12] James Hong, Haotian Zhang, Michaël Gharbi, Matthew Fisher, and Kayvon Fatahalian. Spotting temporally precise, fine-grained events in video. In *European Conference on Computer Vision*, pages 33–51. Springer, 2022. 2, 6

[13] Chun-Chieh Hsu, Hua-Tsung Chen, Chien-Li Chou, Chien-Peng Ho, and Suh-Yin Lee. Trajectory based jump pattern recognition in broadcast volleyball videos. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 1117–1120, 2014. 3, 4

[14] Bingkun Huang, Zhiyu Zhao, Guozhen Zhang, Yu Qiao, and Limin Wang. Mgmae: Motion guided masking for video masked autoencoding. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13493–13504, 2023. 4

[15] Hsiang-Wei Huang, Cheng-Yen Yang, Samartha Ramkumar, Chung-I Huang, Jenq-Neng Hwang, Pyong-Kun Kim, Kyoungoh Lee, and Kwangju Kim. Observation centric and central distance recovery for athlete tracking. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 454–460, 2023. 2

[16] Hsiang-Wei Huang, Cheng-Yen Yang, Jiacheng Sun, Pyong-Kun Kim, Kwang-Ju Kim, Kyoungoh Lee, Chung-I Huang, and Jenq-Neng Hwang. 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*, pages 163–172, 2024. 2, 3, 5

[17] Mostafa S. Ibrahim, Srikanth Muralidharan, Zhiwei Deng, Arash Vahdat, and Greg Mori. A hierarchical deep temporal model for group activity recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1971–1980, 2016. 5

[18] Wei Jiang, Juan Higuera, Baptiste Angles, Weiwei Sun, Mehrsan Javan Roshtkhari, and Kwang Yi. Optimizing through learned errors for accurate sports field registration. In *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 201–210, 2020. 1

[19] Rudolf Emil Kalman et al. Contributions to the theory of optimal control. *Bol. soc. mat. mexicana*, 5(2):102–119, 1960. 2

[20] Maria Koshkina and James H Elder. A general framework for jersey number recognition in sports video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3235–3244, 2024. 2

[21] Xiaohan Nie, Shixing Chen, and Raffay Hamid. A robust and efficient framework for sports-field registration. In *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1935–1943. IEEE, 2021. 1

[22] Yin May Oo, Ankhzaya Jamsrandorj, Vanyi Chao, Kyung-Ryoul Mun, and Jinwook Kim. A residual attention-based efficientnet homography estimation model for sports field registration. In *IECON 2023-49th Annual Conference of the IEEE Industrial Electronics Society*, pages 1–7. IEEE, 2023. 1, 4

[23] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015. 2

[24] Vladimir Somers, Victor Joos, Anthony Cioppa, Silvio Giancola, Seyed Abolfazl Ghasemzadeh, Floriane Magera, Baptiste Standaert, Amir M Mansourian, Xin Zhou, Shohreh Kasaei, et al. Soccernet game state reconstruction: End-to-end athlete tracking and identification on a minimap. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3293–3305, 2024. 2

[25] Jonas Theiner, Wolfgang Gritz, Eric Müller-Budack, Robert Rein, Daniel Memmert, and Ralph Ewerth. Extraction of positional player data from broadcast soccer videos. In *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1463–1473, 2022. 2

[26] Kanav Vats, William McNally, Pascale Walters, David A Clausi, and John S Zelek. Ice hockey player identification via transformers and weakly supervised learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3451–3460, 2022. 2

[27] Nicolai Wojke, Alex Bewley, and Dietrich Paulus. Simple online and realtime tracking with a deep association metric. In *2017 IEEE international conference on image processing (ICIP)*, pages 3645–3649. IEEE, 2017. 2

[28] Artur Xarles, Sergio Escalera, Thomas B Moeslund, and Albert Clapés. T-deed: Temporal-discriminability enhancer encoder-decoder for precise event spotting in sports videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3410–3419, 2024. 2, 6

[29] Bin Xiao, Haiping Wu, Weijian Xu, Xiyang Dai, Houdong Hu, Yumao Lu, Michael Zeng, Ce Liu, and Lu Yuan. Florence-2: Advancing a unified representation for a variety of vision tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4818–4829, 2024. 4

[30] Rui Yan, Lingxi Xie, Jinhui Tang, Xiangbo Shu, and Qi Tian. Social adaptive module for weakly-supervised group activity recognition, 2020. 5

[31] Yifu Zhang, Peize Sun, Yi Jiang, Dongdong Yu, Fucheng Weng, Zehuan Yuan, Ping Luo, Wenyu Liu, and Xinggang Wang. Bytetrack: Multi-object tracking by associating every detection box. In *European Conference on Computer Vision*, pages 1–21. Springer, 2022. 2

[32] Kaiyang Zhou, Yongxin Yang, Andrea Cavallaro, and Tao Xiang. Omni-scale feature learning for person re-identification. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3702–3712, 2019. 2, 3

[33] Kaiyang Zhou, Yongxin Yang, Andrea Cavallaro, and Tao Xiang. Learning generalisable omni-scale representations for person re-identification. *IEEE transactions on pattern analysis and machine intelligence*, 44(9):5056–5069, 2021. 2

[34] Gal Ziv and Ronnie Lidor. Vertical jump in female and male basketball players—a review of observational and experimental studies. *Journal of science and medicine in sport*, 13(3):332–339, 2010. 4