

Dynamic sports environments: automatic feature players tracking and re-identification with similar appearances

Adil Abdulhur Abushana¹, Riyadh Alsaeedi²

¹Department of Computer Science, Faculty of Education, University of Kufa, Najaf, Iraq

²Department of Statistics, College of Administration and Economics, University of Wasit, Wasit, Iraq

Article Info

Article history:

Received May 24, 2025

Revised Jul 22, 2026

Accepted Aug 12, 2026

Keywords:

3D localization

Deep player re-identification

KSP-ID model

Player tracking

SSD algorithm and object interactions

ABSTRACT

Tracking multiple players in sports videos remains challenging due to frequent occlusions, similar appearances among players, and rapid motion. This paper presents a comprehensive player tracking framework that integrates deep player re-identification (DPRI), an individual probabilistic overlapping bounding box occupancy mapping (IPOM) for robust spatial localization under severe overlap, and an identity-aware k-shortest path optimization algorithm (KSP-ID) for improved trajectory association. The DPRI module combines pose-based embeddings, jersey appearance cues, and facial features to generate discriminative identity representations. Experimental evaluation on the autonomous production of images based on distributed and intelligent sensing (APIDIS) and Shantou University (STU) datasets demonstrates strong performance, achieving 85.5% multiple object tracking accuracy (MOTA), 69.0% generalized multiple object tracking accuracy (GMOTA), and 90.2% rank-1 re-identification accuracy. Furthermore, the proposed framework significantly reduces identity switches (IDS) compared with conventional k-shortest path (KSP) approaches, probabilistic occupancy models, and recent deep learning-based tracking methods. These results demonstrate that integrating semantic identity information with spatial reasoning enables robust and efficient player tracking in highly dynamic sports environments.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Adil Abdulhur Abushana

Department of Computer Science, Faculty of Education, University of Kufa

Najaf, Iraq

Email: adelshanna3000@gmail.com

1. INTRODUCTION

MOT or multi-object tracking in sports is necessary for automatic analytics, tactical performance understanding, and smart sports broadcasting. K-shortest path (KSP) optimization has been extensively used with tracking-by-detection frameworks due to their ability to express the identity association as a constrained flow problem, which allows the global optimization of tracking trajectories to be very efficient [1]. However, even with these advantages, identity switches (IDS) are still a major drawback in the case of sports settings that are very dynamic since the constant interaction between players, motion blur, occlusions, camera angle changes, and similarity in appearance among the players make it very difficult to keep the identity consistent for long periods of time [2], [3].

To help mitigate these problems, some recent works have proposed leveraging both appearance and online learning methods to help differentiate players in crowded scenarios [2], [3]. ID-Net is one such method based on the use of deep representations of the parts of human bodies to improve robustness in identity preservation, using structural cues and localized body parts rather than relying solely on holistic

cues [4]. On the other hand, jersey number detection has become another powerful identity cue, aided by the advancements of coarse-to-fine convolutional neural network (CNN) approaches and sophisticated segmentation architectures that are able to isolate jersey number digits under partial occlusions [5], [6]. Additionally, probabilistic occupancy mapping, multi-view fusion, pose estimation, and camera calibration have improved spatial localization and 3D scene interpretation [7], [8]. While there have been major advancements in sports player tracking in recent years, many of these methods face challenges in preserving identity in cases where there is similarity in visual appearance, partial jersey numbers visibility, or overlapping players. One of the weaknesses of previous methods is the separate consideration of each of the available identity cues—jersey recognition, appearance features or facial information, and spatial localization—which usually results in loss of identity in complex interactions.

In light of the above problem, a comprehensive approach is needed that incorporates appearance and trajectory representation and identity preservation in 3D spatial localization, in one consistent optimization framework. This work presents a multi-player tracking method, which is made up of three tightly connected modules. Firstly, the deep player re-identification (DPRI) module uses pose guided embeddings, jersey semantics, and facial information to generate highly discriminative identity features under occlusion and ambiguous conditions. Secondly, individual probabilistic overlapping bounding box occupancy mapping (IPOM)—the individual probabilistic overlapping bounding box for occupancy mapping—module allows for improved spatial consistency through accurate 3D scene grounding from multiple views, especially in crowded interaction scenarios. Finally, the identity-aware k-shortest path optimization method (KSP-ID) utilizes an identity correlation coefficient in the association graph, which significantly lowers the identity switch rate compared to conventional KSP.

2. RELATED WORK

Over the last decade or so, player detection, identification, and tracking in sports videos have undergone substantial progress toward more accurate solutions. Part and local structure representations have been among the early methods for addressing such tasks as re-identification and coping with appearance changes [4]. Initially, player identification relied on handcrafted feature sets, color, and moment representations invariant to multiple conditions before transitioning to scene perception systems with integrated segmentation, detection, and recognition procedures [9]. In the process of enhancing team identification robustness, there has been a growing interest in utilizing team members' jerseys with unique identifiers based on the technologies ranging from optical character recognition (OCR) systems to those utilizing deep learning, CNNs, region-based convolutional neural network (R-CNN) with human poses, multi-task training, spatial transformation, and visual-text combinations [10]–[13].

Multi-target tracking was a simultaneous process from the beginning, relying on the global optimal graph optimization, probabilistic reasoning, greedy strategies, and KSP solutions among others to keep the temporal identity association stable [14]–[17]. The deep learning revolution then came for the online tracking which, among others, resulted in the deep simple online and real-time tracking (DeepSORT) model that merged deep appearance embeddings with Kalman filtering to increase the identity continuity [15] followed by advanced optimization and graphical modeling across different sports contexts [16], [17].

The deep learning revolution eventually affected the very detection: the convolutional architectures like faster R-CNN, mask R-CNN, you only look once (YOLO), and feature pyramid networks (FPN) took the lead in accuracy, speed, and multi-scale reasoning in the dynamic sports environment by a big margin [18]–[21]. Global contextual modeling along with transformer-based detection boosted the strength in dense and complex scenes indirectly [21], the same way that the benchmarking initiatives taking place did by making performance evaluation a standard across the datasets such as multiple object tracking challenge (MOTChallenge), autonomous production of images based on distributed and intelligent sensing (APIDIS), SoccerNet, and others included in the large-scale protocols [22]. Identity signals helping one another, with face recognition and spatial localization cues being the most prominent, worked together for the reliable identity inference under the most difficult scenarios [23]–[25].

Classical description methods have already demonstrated high efficiency: even with the moderate robustness of human detection by histogram of oriented gradient (HOG), sparse histogram encoding, correlation filters, and multiple instance learning, it was easy to fool such systems with occlusion, motion blurring, and significant scale change [26], [27]. The jersey recognition and identity modeling processes were going through the architectural advancements of CNNs, visual-textual data fusion, and robust recognition techniques suitable for solving the problem of variability in matches [28]. Meanwhile, new tracking approaches were being designed to achieve global optimality in terms of association, probabilistic graphical models, and efficient identity stability [29]–[34]; hence, they benefited from the structured temporal reasoning advances. Pose-based player descriptors along with their structural and part-based characteristics have been one of the key re-identification improvements in almost identical sports domains [35]–[40].

Such further improvements have helped in reaching high performance levels in the following areas: high-order optimization in graphs, heuristic solutions for better association, and resilient jersey identity persistence in challenging scenes [41], [42]. Standard datasets, benchmarks, and evaluation protocols helped in providing unbiased comparison between competing algorithms [43]. Part-based re-identification approaches were created to use multi-scale structural features and local discriminative patches; therefore, jersey digits have become less sensitive in case of masking or distortion [44]. The jersey recognition methods were gradually expanding, and visual patterns and numerical identifiers became increasingly related semantically; hence, the reliability of gameplay analysis became higher [45]–[47]. Besides, benchmarking and structured evaluations have finally created a solid base for objective assessment in sports analytics research [48]. With the development of multi-camera tracking in densely populated sports scenarios, advanced KSP optimization, probabilistic formulations, and identity persistence mechanisms came into play [49], [50]. The use of probabilistic occupancy mapping, volumetric reasoning, depth cues, and marked point processes brought to localization a remarkable resilience even under severe occlusion [51]–[56], and all of these advanced techniques resulted in the creation of state-of-the-art multi-camera tracking pipelines in which spatial re-identification, calibration, and identity fusion strategies together provided accurate multi-player tracking performance [52], [57].

However, notwithstanding this impressive advancement, most of the previous methods still consider jersey recognition, facial cues, spatial localization, and identity association as loosely related modules, which makes identity very weak in situations when players frequently overlap, have very similar appearances, or move through the view of multiple cameras. As a result of these deficiencies, our research integrates pose-guided jersey recognition, deep re-identification, probabilistic multi-view localization, and identity-aware KSP optimization into one coherent framework that is specifically designed for the reliable identity preservation in highly dynamic sports environments.

3. METHOD

3.1. Dataset

The deep player model is trained in three stages following the image-centric sampling strategy [21] with the stochastic gradient descent (SGD) solver [22]. The residual network 50 layers (ResNet-50) backbone is pre-trained, while the remaining layers are randomly initialized with a Gaussian distribution. The loss is defined as (1).

$$L = \lambda^1 L_{face} + \lambda^2 L_{jers} \quad (1)$$

Weakly labeled player bounding boxes contrast with fully labeled jersey number boxes. APIDIS offers 10,000 weakly and 2,000 fully labeled samples, with empirical weights $\lambda_1 = 0.55$ and $\lambda_2 = 0.55$. IPOM and KSP-ID parameters are dependent on data sets: grids of size 128×72 for APIDIS, and 112×60 for Shantou University (STU). APIDIS [23] contains 1,500 frames per second at 25 fps for seven cameras (800×600) with 12 players, while STU features 2,500 frames per second at 24 fps for eight cameras (1280×720) with 16 players, and has an expanded 11-period version in development. Hue, saturation, and value (HSV) is possible for analyzing jersey numbers and face recognition in both datasets. Figure 1 is illustration of the dictionary creation process involving three complementary components that use color and texture features obtained in HSV.

3.2. Robust segmentation

This study considers a step-by-step approach to image segmentation beginning with an initial image from the sequence. Several human detection methods, which involve baseline CNN models, Faster R-CNN, and HOG algorithm, were compared for their ability to recognize players [24]–[26]. Although scale-invariant feature transform (SIFT) descriptor has been used mostly for face recognition, other mentioned detection methods also involve such cues as jersey numbers and overlay texts [27]. The segmented regions are represented as (2).

$$R_i = (l_i, r_i, b_i, t_i, w_i, h_i, c_i, z_i) \quad (2)$$

Where l_i, r_i, b_i, t_i denote the left, right, bottom, and top coordinates of bounding rectangle; w_i, h_i represent its width and height; c_i is the class label; and z_i contains additional attributes for player identity.

The segmented region captures the pipe-like structure of jersey numbers using a morphological open operation, leveraging the observation that digital characters consist of pipe-like components. To manage occlusions, merge-split ambiguities, and spatial overlaps, a MOT system is applied using nearest-neighbor

assignment, considering blob size, spatial relationships, and reliable feature descriptors [28]–[30]. Thresholds and feature weights are adaptively tuned for optimal performance in dynamic sports environments [30]. Algorithm 1 summarizes this segmentation procedure, from background-based candidate detection through color-based region merging.

Algorithm 1. Robust segmentation of jerseys

- i) Initialize parameters set thresholds: color similarity threshold th_{color} and size threshold for candidates th_{size} .
- ii) Capture video frames: open the video file and initialize a list to store frames, while the video is still open read each frame F_n and append it to the list of frames.
- iii) Create background model by convert the list of frames into an array F . Calculate the median background model using: $B = \text{median}(F)$ (along the temporal axis).
- iv) Process each frame for each frame F_n in the captured frames
 - a) Segment players compute the absolute difference: $D_n = |F_n - B|$ convert to grayscale:
 - $G_n = \text{cvtColor}(D_n, \text{COLOR_BGR2GRAY})$
 - Apply binary thresholding: $T_n = \begin{cases} 255 & \text{if } G_n > 300 \\ 0 & \text{otherwise} \end{cases}$
 - Apply morphological opening $O_n = \text{morphologyEx}(T_n, \text{MORPH_OPEN}, K)$ where K is a structuring element.
 - b) Find identify contours C in O_n :
 - $C = \text{findContours}(O_n, \text{RETR_EXTERNAL}, \text{CHAIN_APPROX_SIMPLE})$.
 - Initialize an empty list for candidate blobs.
 - c) Filter candidates for each contour $c \in C$ as follows compute area A_c c: $A_c = \text{contourArea}(c)$:
 - If $A_c \geq th_{size}$: compute the bounding rectangle $(x, y, w, h) = \text{boundingRect}(c)$.
 - Add (x, y, w, h) to the candidates list.
 - 4.5 merge adjacent regions.
 - Initialize a list for merged candidates for each pair of candidates (R_i, R_j) .
 - Check color similarity: $C_i = \text{mean}(R_i)$ and $C_j = \text{mean}(R_j)$.
 - Compute the color distance: $d = \|C_i - C_j\| < t_c$.
 - If d is true, merge the bounding rectangles: $M = (\min(x_i, x_j), \min(y_i, y_j), \max(x_i + w_i, x_j + w_j) - \min(x_i, x_j), \max(y_i + h_i, y_j + h_j) - \min(y_i, y_j))$. Add M to the merged candidates list.
- v) Draw rectangles for each merged candidate $(F_n, (x, y), (x + w, y + h), (0, 255, 0), 2)$.

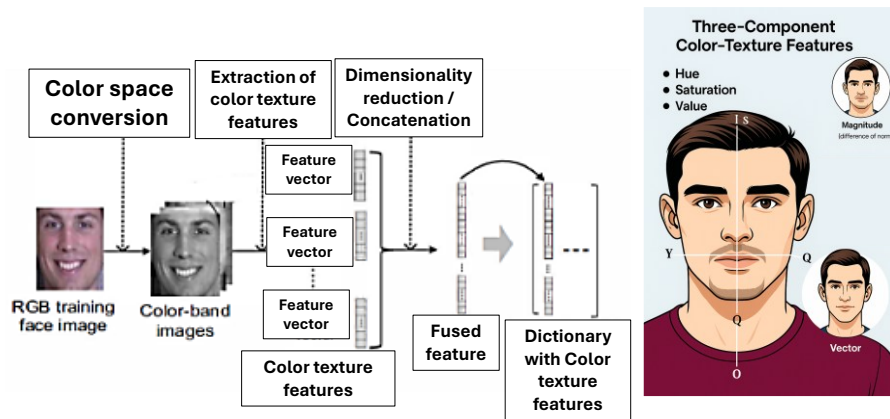


Figure 1. Creating dictionary with three components color texture features HSV

3.3. Color texture extraction for facial expression

The proposed dictionary system uses color–texture features from RGB face images. Histogram-based features are extracted across multiple color bands to enhance sparse representation face recognition [31], [32], while novel color vector difference patterns improve skin texture discrimination, computed using a circular neighborhood of radius R and P pixels [33].

3.4. Automatic face and jersey finder

Building on prior work [34], [35] the system integrates an automatic face and jersey finder using a hierarchical top–down parsing tree to detect jersey numbers and facial features. Frames are recursively

segmented to reduce computation, while kernelized correlation filters and multiple instance learning [37] enable fast tracking and adapt to appearance variations. These mechanisms together improve accuracy and real-time player monitoring.

3.5. KSP-ID model with deep learning techniques

Subsection 3.2 outputs feed into the KSP-ID model, where efficient segmentation ensures precise player identification. Recognition is done through multiple camera and close-up views, using both face detection and jersey number recognition [38], [39]. Some approaches use direct jersey number recognition through alternative descriptors [41]. The MOT technique used is based on track-to-detection association based on KSP optimization [42], but switching identities takes place in the absence of strong appearance models [43]; online MOT based on MOT can only happen when cues are present in motion and appearance [44]. In terms of robustness, the deep learning techniques have been useful: ID-Net uses body attributes while deep regression trackers [45] use residual connections and hard-sample mining on CNNs. KSP-ID fuses face and jersey tracking by using spatial color features and local binary pattern (LBP) with deep players' identification [46]. Appearance similarity scores $S_a(d_k, t_j)$ assesses visual characteristics using metrics like color histograms and texture descriptors. Motion similarity $S_m(d_k, t_j)$ is derived from the predicted position $\hat{t}_j$ using a motion model, calculating the distance between the detection and predicted position. Proximity similarity $S_p(d_k, t_j)$ measures the closeness of the detection to the predicted position. Total similarity score combines these metrics as in (3).

$$S(d_k, t_j) = \omega_a \cdot S_a(d_k, t_j) + \omega_m \cdot S_m(d_k, t_j) + \omega_p \cdot S_p(d_k, t_j) \quad (3)$$

A cost matrix C is formed where $C_{kj} = 1 - S(d_k, t_j)$ defined as assignment problem. The Hungarian algorithm optimally assigns detections to tracks, confirming a track as normal if it maintains consistent associations across frames. In cases of absence, tracks may transition to "Lost" or "Occluded," influenced by the duration of the absence, while identity recovery utilizes appearance models for real-time surveillance and effective multi-target monitoring through probabilistic data association. To reduce drifting and mismatching after occlusion, thresholds T_{lost} and $T_{occluded}$ are applied to normal tracks without assigned measurements. Normal tracks without assigned measurements may transition to "Lost" or "Occluded" based on absence duration, while identity recovery uses appearance models and probabilistic data association to enable real-time surveillance of low-resolution color targets in dynamic environments, referred to as the re-identification mechanism [42], [44], [45]. Algorithm 2 details this procedure, covering similarity-score computation, the Hungarian assignment step, and the track-state transition rules described above.

Algorithm 2. Similarity scores computation and track management

- i) Similarity scores computation for each detection d_k and existing track t_k , compute the similarity scores:
 - Appearance similarity $S_a(d_k, t_j)$: this function can incorporate metrics like color histograms, texture descriptors, and shape features.
 - Motion similarity $S_m(d_k, t_j)$: predicted position $\hat{t}_j$ using a motion model (e.g., Kalman filter):
 $\hat{t}_j = f(t_j, u_j)$ where u_j is the control input.
 - Compute the distance from the detection: $S_m(d_k, t_j) = \exp(-\frac{\|d_k - \hat{t}_j\|^2}{2\sigma^2})$.
 - Proximity similarity: $S_p(d_k, t_j) = \frac{1}{1 + \|d_k - \hat{t}_j\|}$.
 - Total similarity score $S(d_k, t_j)$: combine the scores using weights $\omega_a, \omega_m, \omega_p$:
 $S(d_k, t_j) = \omega_a \cdot S_a(d_k, t_j) + \omega_m \cdot S_m(d_k, t_j) + \omega_p \cdot S_p(d_k, t_j)$
- ii) Assignment problem: given similarity scores $S(d_k, t_j)$, construct a cost matrix C :
 - Total similarity score: $C_{kj} = 1 - S(d_k, t_j)$.
 - Use the Hungarian algorithm to find the optimal assignment that minimizes total cost: minimize $\sum_{j=1}^M C_{kj} X_{kj}$, where X_{kj} is a binary variable indicating whether detection k is assigned to track j .
 - Track state transition conditions.
 - Tentative track confirm as normal track if associated across multiple frames:
 If $\sum_{t=1}^T \text{association}(d_k, t_j, t) \geq T_{confirm} \Rightarrow \text{normal track}$.
- iii) Transition to lost track:
 - Transition to lost track if no association for T_{lost} frames:
 If $\text{association}(d_k, t_j, t) = 0$ for T_{lost} frames $\Rightarrow$ lost track.
 - Occluded track handling: remain in occluded track if partial detection is associated: if $\text{partial_association}(d_k, t_j, t) \Rightarrow$ occluded track.
 - Termination of tracks: terminate if in lost track for T_{lost} frames
 If not detected for T_{lost} frames $\Rightarrow$ terminate track.

iv) Re-identification mechanism use appearance models for identity recovery:

- $S_{reID}(d_k, t_j, t) = \text{similarity}(\text{appearance embedding}(d_k), \text{appearance embedding}(t_j))$.
A real time adaptive visual surveillance system for tracking low-resolution color targets in dynamically changing scenes.

3.6. Evaluation metrics

To improve annotation quality, the initial labels are refined by suppressing false positives and incorporating missed detections, producing corrected labels y_{corr} . These are used to fine-tune the teacher network, yielding the final teacher network (FTN) with improved annotations FTN(X). Performance is measured using the multiple object tracking accuracy (MOTA) metric, as defined in (4) [47].

$$MOTA = 1 - \frac{(\sum_t (c_m m_t + c_{fp} p_t + c_s mme_t))}{(\sum_t g_t)} \quad (4)$$

Where m_t , fp_t , and mme_t represent false negatives (FN), false positives (FP), and IDS, respectively, with g_t as ground truth. To address sensitivity to IDS, the generalized multiple object tracking accuracy (GMOTA) [2] replaces mme with gmme, accounting for correctly labeled trajectories, as defined in (5).

$$GMOTA = 1 - \frac{(\sum_t (c_m m_t + c_{fp} p_t + c_s gmme_t))}{(\sum_t g_t)} \quad (5)$$

Alongside MOTA and GMOTA, true positives (TP), FP, FN, and IDS were reported. Knowledge distillation for real-time detection to ensure real-time deployment, knowledge is distilled from the ResNet-50 teacher network into a faster ResNet-18 student network, both enhanced with an FPN [48]. The pipeline is trained on unlabeled SoccerNet broadcast videos [49], improve detection of faces, jerseys, and player bodies, balancing accuracy with computational efficiency for real-time sports analytics.

4. RESULTS AND DISCUSSION

4.1. Result space, evaluation possibilities, and algorithmic behavior

The suggested model was thoroughly tested in real-world operations to assess its durability and stability. The cases used involved the scattering of players in an uneven manner, interaction at medium levels, and densely packed playing sessions, where the key problem was often overlapping of the players' images. In addition, the test was performed in static and highly dynamic motion environments involving short and long-term activity monitoring. This evaluation was carried out in both single- and multiple-camera configurations allowing for the comparison of single-view spatial perception and multiple-view redundancy. Some of the factors involved in the evaluation include jersey distortion from quick body movements, occlusions owing to players overlapping each other, resemblance between players, and image blur because of varying light conditions.

In fact, the DPRI block succeeded in maintaining an excellent discriminative power among visually similar players across all conditions mentioned above, while the individual probabilistic overlapping bounding box contributed significantly to 3D groundings in the presence of noise. With the introduction of an identity-aware KSP-ID method, a continuous identity flow for tracking-by-detection could be achieved without facing the issue of fragmentation and identity drift, a typical feature of KSP approaches [49], [50]. Generally speaking, the model was highly resilient in all scenarios, generating consistent tracks with minimal identity loss.

4.2. Quantitative evaluation

Quantitative evaluation on APIDIS and comparative analysis with literature. Evaluation results on the APIDIS dataset are shown in Table 1. The traditional KSP-based algorithms KSP [49] and constrained k-shortest paths (C-KSP) perform well in terms of moderate tracking accuracy (MOTA = 49.0% for KSP and 58.9% for C-KSP) but experience high identity fragmentation (IDS = 6.5 for KSP and 4.6 for C-KSP), mainly due to weak appearance reasoning capability. Probabilistic and occupancy-based solutions like dynamic programming (DP) [51] and volumetric mass densities (VMD)/volumetric mass densities constrained (VMDc) [52] can enhance spatial grounding and lower IDS, albeit in dense scenes.

More advanced systems such as temporal motion and spatial association (TMOSA) [53], trajectory matrix approximation via expectation-maximization (TMAP-EM) [54], and multi-sensor pedestrian tracking (MSPT) [55] incorporate richer temporal and structural constraints, achieving higher MOTA (79.9–85.5%) and substantially reduced IDS. Meanwhile, multi-camera deep frameworks such as multi-camera multi-person tracking and identification (MCMPTI) [58] exploit spatial redundancy to improve global tracking performance, obtaining high GMOTA (72.3%) and the lowest reported IDS (0.6). The given framework, in direct comparison, scored MOTA 84.5%, GMOTA 67%, and IDS 1.0, thus showing accuracy

that is either on par or better than most recent top-notch systems and at the same time ensuring identity stability that is as good as or better than the best probabilistic and hybrid association models [53]–[57]. It should be noted that the proposed method is not like TMOSA [53] that requires the imposition of pre-structured constraints or MCMPTI [57] that can only work under certain multi-camera conditions, but rather it is a strong performer in the area of fully automated and generalizable pipelines.

Table 1. Quantitative comparison of proposed method against state-of-the-art trackers on the APIDIS dataset

Method	MOTA↑	GMOTA↑	TP↑	FP↓	FN↓	IDS↓
KSP [49]	49.0	38.0	86.2	13.8	31.3	6.5
C-KSP [50]	58.9	43.3	88.1	11.9	22.5	4.6
DP [51]	49.5	39.1	—	—	—	—
VMD [52]	59.7	—	88.7	17.2	28.6	1.4
VMDc [52]	67.5	—	93.0	12.5	24.4	1.3
TMOSA [53]	85.5	68.7	104.5*	2.6	13.5	1.2
TMAP-EM [54]	79.9	—	101.6*	6.6	15.4	1.6
MSPT [55]	78.6	—	98.7	10.2	22.2	4.3
Hybrid full [56]	62.6	—	—	—	—	—
MCMPTI [57]	81.1	72.3	103.2*	6.2	14.6	0.6
This work	84.5	67.0	101.0*	4.2	12.9	1.0

4.3. Ablation analysis on APIDIS

The study conducted an ablation analysis to determine the impact of each new component with the help of Table 2. The basic probabilistic occupancy mapping combined with KSP-ID shows that high rates of performance are impossible (MOTA 49.0%, GMOTA 38.0%, and IDS 6.5), the results are pretty much the same as those previously reported for graph-optimization frameworks that are classical [49], [50]. Adding cascade mask R-CNN to the process not only greatly increases the integrity of detection but also lowers the IDS to 4.7, which indicates the significance of reliable segmentation priors found earlier in player thick environments [18], [24].

Table 2. Ablation experiment results on the APIDIS dataset

Method	MOTA↑	GMOTA↑	TP↑	FP↓	FN↓	IDS↓
POM + KSP	49.0	38.0	86.2	22.1	31.3	6.5
Cascade mask RCNN-P + POM +KSP	61.3	51.4	92.4	14.5	26.9	4.7
Ours DPRI	72.0	60.0	97.8	11.3	21.4	3.6
Ours IPOBB	75.0	62.0	99.2	10.6	20.0	2.8
Ours KSP-ID	78.0	65.0	101.3*	8.5	18.6	2.1
Ours (full)	85.5	69.0	104.8*	4.2	12.2	0.7

The addition of the DPRI module has a drastic effect on the tracking consistency that now reaches (MOTA 72.0%) and further brings the IDS down to 3.6. This shows the need for strong appearance and identity embeddings which are already corroborated by [35], [44] to be the case with deep identity learning systems. Then came the addition of IPOM and KSP-ID which lowered IDS further to 2.1 by means of better 3D localization and identity-aware association, which is similar to the results earlier obtained in [14], [51] with multi-view occupancy tracking frameworks and [41], [52] with identity-correlated graph optimization. The complete framework at last achieves MOTA 85.5%, GMOTA 69.0%, IDS 0.7, thus evidencing the interdependence of jersey recognition, facial cues, deep embeddings, and KSP-based identity reinforcement through their complementary interaction.

4.4. Evaluation on the STU dataset

Evaluating the system with the STU dataset once more is an assurance of its generalization as shown in Table 3. The very first baseline results (MOTA 41.4%, GMOTA 34.9%, and IDS 13.6) indicate the extreme difficulty of the dataset which is marked by very high player density and frequent severe occlusions. Though MOTA is escalated to 55.1% by cascade mask R- as for CNN, the degree of identity fragmentation remains rather high since the intersection of all scenarios leads to a lot of ambiguities. At that, the inclusion of DPRI, IPOM, and KSP-ID modules in sequence contributes to a consistent increase in MOTA up to 80.0% from 70.0% and in GMOTA up to 65.0% from 55.0%, while at the same time leading to a reduction in IDS as well. Such a tendency in performance can be regarded as an example of what was stated previously

regarding the problem of identity preservation in dense sports tracking [52], [54]. At the same time, CNN is proved more stable than KSP-based pipelines and approaches.

Table 3. Ablation experiment results on the STU dataset

Method	MOTA ↑	GMOTA ↑	TP ↑	FP ↓	FN ↓	IDS ↓
POM + KSP	41.4	34.9	77.9	25.6	36.2	13.6
Cascade mask RCNN-P + POM +KSP	55.1	42.5	86.8	22.3	31.4	9.9
Ours DPRI	67.0	54.0	97.3	12.9	25.7	–
Ours IPOM	69.0	56.0	100.3*	12.1	12.1	24.3
Ours KSP-ID	72.0	58.0	103.2*	10.0	10.0	23.0
Ours (full)	80.0	65.0	107.6*	5.7	5.7	17.2

4.5. Re-identification performance

Player re-identification performance is summarized in Table 4. Baseline performance remains weak due to limited appearance discrimination, while the cascade mask R-CNN model considerably improves match consistency. The deep player module surpasses both, achieving rank-1 = 90.2%, top-3 = 93.5%, and mean average precision (mAP) = 80.0%, consistent with prior evidence that deep structured identity representation substantially enhances player recognition robustness in sports analytics [35], [52].

Table 4. Ablation studies on player re-identification on the STU dataset

Method	top-1 ↑	top-3 ↑	mAP ↑
POM + KSP	58.4	68.1	9.8
Cascade mask RCNN-P + POM +KSP	85.9	90.9	75.0
Deep player	90.2	93.5	80.0

4.6. Discussion

The experimental results clearly show that combining the deep re-identification approach, the probabilistic multi-view localization, and the identity-aware graph optimization has a tremendous effect on the reliability of multiple players' tracking in dynamic sports environments. The classical KSP trackers [49], [50], on the other hand, which mainly rely on keeping track of the trajectories and use very little appearance cues, do not have the proposed framework that introduces discriminative identity semantics and spatial grounding, whose role is to set up stabilizing constraints in the times of ambiguous association states. This is especially true for the high-density scenes where there are frequent inter-player interactions and overlapping bounding regions that have been leading to fragmentation and identity drift traditionally, phenomena that have been extensively reported in the previous tracking literature [14], [51], [52], [54]. The evaluation results that show a marked reduction of IDS in our case confirm that deep feature embeddings and jersey-guided recognition reinforcement of identity provide a corrective influence over the conventional motion-only reasoning.

One of the most significant results is the DPRI module's support of long-term identity consistency in a significant way. The model, like previous identity-centric systems [35], [44], is aided by a hybrid appearance representation that comprises global player structure, jersey semantics, and localized discriminative embeddings, but it is not as previous works where jersey recognition was an isolated recognition task [10]–[13], [38], [39] and recognition done separately, thus, the current framework connects jersey cues to the association process and turning jersey digits from a passive feature into an active identity constraint. This accounts for the strong correlation between improved re-identification metrics and reduced IDS values, which further supports the recent conclusions of multi-camera and sports tracking research [13], [52], [57].

The probabilistic occupancy and IPOM methods give an extra structural benefit in particular to gameplay situations that are heavily blocked. Early works using probabilistic and volumetric methods [14], [51], [53], [56] have shown that 3D reasoning with occupancy restrictions can substantially lower localization ambiguity. The current system goes one step further by integrating occupancy into identity-aware graph optimization instead of using it as a separate localization stage. By doing so, the system makes sure that the spatial and identity decisions are made concurrently throughout the tracking process, which avoids the common "identity snap" phenomenon happening when a system relies only on geometric alignment or a visual similarity. The identical performance of APIDIS and STU indicates that the new model has the capability of maintaining its robustness over datasets with very different levels of congestion and visual complexity. This has been a major problem in sports analytics historically.

The evaluation of the proposed system was against TMOSA [53], TMAP-EM [54], MSPT [55], hybrid stochastic–deterministic trackers [17], and MCMPTI [57] and showed that the proposed system has a well-balanced performance profile. MCMPTI, with its extensive multi-camera redundancy, and TMOSA, with its strong structured constraints, have achieved better performance than the proposed model. However, the proposed architecture is able to perform at a similar level and even show some competitive traits in terms of identity stability without aggressive scene assumptions or complex constraint tuning. This indicates that the proposed architecture has a more adaptable solution for different types of sporting situations, such as those with inconsistent camera calibration or without any infrastructure at all, where it may not be always guaranteed that the camera calibration is the same or that the infrastructure is present.

In spite of these benefits, there are still a few limitations that need to be pointed out. Over extreme and prolonged occlusion, or when jersey digits stay completely invisible for long periods of time, the system performance still gives way. This is consistent with past research on jersey-driven systems [10]–[13], [38]. Similarly, it could be that visual embeddings used for identity rely too much on the domain, and transferring across significantly different visual environments, lighting conditions, or team uniform designs may become an issue. The system also suffers from high computational cost due to the multi-stage processing and high-dimensional embedding extraction, which is a characteristic shared by advanced deep tracking architectures [21], [35]. Besides, even though the stability is enhanced by multi-camera conditions, the performance in challenging monocular environments could still be improved by stronger temporal reasoning or learned motion priors.

In summary, the presented results bolster the argument that the combination of jersey semantics, deep re-identification, probabilistic spatial reasoning, and KSP-based identity-aware optimization is the most effective approach to sports multi-player tracking from a structural point of view. The showcased advantages over basic KSP frameworks [49], [50], probabilistic occupancy models [14], [51], hybrid optimization trackers [17], [41], and contemporary deep learning and multi-camera solutions [52], [54], [55], [57] indicate that the proposed method is a significant and scalable step forward in the direction of dependable automated sports analytics.

5. CONCLUSION

In this paper, a comprehensive player tracking method is suggested which combines DPRI, multi-view occupancy modeling, and identity-aware KSP optimization for handling identity changes and the instability of paths in dynamic sports scenes. The tests done on APIDIS and STU datasets showed that the proposed method provides considerable benefits in terms of tracking accuracy, global stability, and identity consistency when compared with traditional KSP systems, probabilistic occupancy methods, and cutting-edge deep and hybrid trackers. The ablation study confirmed that the robustness achieved is the result of the collaboration between jersey-guided semantics, deep embeddings, and spatial grounding. The approach that is being presented still reveals a strong level of generalization and scalability even though extreme occlusions and domain shifts are still issues impacting performance. The future research will then point at improving computational efficiency, carrying out evaluations in a wider variety of sports scenarios, and integrating the latest temporal and self-supervised learning techniques to further confirm long-term identity preservation as the most critical problem to be solved.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Adil Abdulhur Abushana	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓
Riyadh Alsaeedi		✓				✓	✓	✓	✓	✓	✓	✓		✓

C : **C**onceptualization

M : **M**ethodology

So : **S**oftware

Va : **V**alidation

Fo : **F**ormal analysis

I : **I**nterpretation

R : **R**esources

D : **D**ata Curation

O : **O**riginal Draft

E : **E**xperimentation

Vi : **V**isualization

Su : **S**upervision

P : **P**roject administration

Fu : **F**unding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

This section not applicable. This study did not involve human participants, human data, or any personally identifiable information. All data used were either publicly available, fully anonymized, or derived from non-human sources, and therefore no informed consent was required from individuals.

ETHICAL APPROVAL

This section not applicable. This research did not involve human subjects, human biological materials, or experimental procedures on animals. The work was conducted solely on computational models, publicly available datasets, or non-sensitive data that did not require intervention with living organisms. Therefore, ethical approval from an institutional review board or animal ethics committee was not necessary for this study.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [AA], upon reasonable request.

REFERENCES

- [1] H. Pirsiavash, D. Ramanan, and C. C. Fowlkes, "Globally-optimal greedy algorithms for tracking a variable number of objects," in *IEEE Computer Vision and Pattern Recognition (CVPR) 2011*, Jun. 2011, pp. 1201–1208, doi: 10.1109/CVPR.2011.5995604.
- [2] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *2017 IEEE International Conference on Image Processing*, Sep. 2017, pp. 3645–3649, doi: 10.1109/ICIP.2017.8296962.
- [3] A. Milan, L. L. -Taixe, I. Reid, S. Roth, and K. Schindler, "MOT16: a benchmark for multi-object tracking," 2016, *arXiv:1603.00831*.
- [4] H. Wu, Y. Hu, K. Wang, H. Li, L. Nie, and H. Cheng, "Instance-aware representation learning and association for online multi-person tracking," *Pattern Recognition*, vol. 94, pp. 25–34, Oct. 2019, doi: 10.1016/j.patcog.2019.04.018.
- [5] S. Gerke, K. Muller, and R. Schafer, "Soccer jersey number recognition using convolutional neural networks," in *2015 IEEE International Conference on Computer Vision Workshop*, Dec. 2015, pp. 734–741, doi: 10.1109/ICCVW.2015.100.
- [6] Z. Cai and N. Vasconcelos, "Cascade R-CNN: delving into high quality object detection," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 6154–6162, doi: 10.1109/CVPR.2018.00644.
- [7] A. Khotanzad and Y. H. Hong, "Invariant image recognition by Zernike moments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 5, pp. 489–497, May 1990, doi: 10.1109/34.55109.
- [8] S. H. Lee, J. Y. Choi, Y. M. Ro, and K. N. Plataniotis, "Local color vector binary patterns from multichannel face images for face recognition," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2347–2353, Apr. 2012, doi: 10.1109/TIP.2011.2181526.
- [9] Z. Tu, X. Chen, Yuille, and Zhu, "Image parsing: unifying segmentation, detection, and recognition," in *Ninth IEEE International Conference on Computer Vision*, 2003, pp. 18–25, doi: 10.1109/ICCV.2003.1238309.
- [10] H. Li, H. Zhang, and Y. Chen, "Dual-TSST: a dual-branch temporal-spectral-spatial transformer model for EEG decoding," *IEEE Journal of Biomedical and Health Informatics*, vol. 29, no. 9, pp. 6524–6537, Sep. 2025, doi: 10.1109/JBHI.2025.3577611.
- [11] H. Liu and B. Bhanu, "Pose-guided R-CNN for jersey number recognition in sports," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Jun. 2019, pp. 2457–2466, doi: 10.1109/CVPRW.2019.00301.
- [12] K. Vats, M. Fani, D. A. Clausi, and J. Zelek, "Multitask learning for jersey number recognition in ice hockey," in *4th International Workshop on Multimedia Content Analysis in Sports*, Oct. 2021, pp. 11–15, doi: 10.1145/3475722.3482794.
- [13] A. Cioppa, A. Delière, S. Giancola, B. Ghanem, and M. V. Droogenbroeck, "Scaling up SoccerNet with multi-view spatial localization and re-identification," *Scientific Data*, vol. 9, no. 1, Jun. 2022, doi: 10.1038/s41597-022-01469-1.
- [14] F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua, "Multicamera people tracking with a probabilistic occupancy map," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 267–282, Feb. 2008, doi: 10.1109/TPAMI.2007.1174.
- [15] Q. Wang, Z. Liu, Z. Wu, J. Wang, C. Qu, and J. Li, "Drift suppression method based on signal stability detection and adaptive Kalman filter for NMR sensor," *Digital Signal Processing*, vol. 156, Jan. 2025, doi: 10.1016/j.dsp.2024.104812.
- [16] W. Lu, X. Wang, W. An, C. Xiao, Q. Yin, and G. Zhang, "OITrack: multi-object tracking for small targets in satellite video via online trajectory completion and iterative expansion over union," *Remote Sensing*, vol. 17, no. 12, Jun. 2025, doi: 10.3390/rs17122042.
- [17] R. T. Collins and P. Carr, "Hybrid stochastic/deterministic optimization for tracking sports players and pedestrians," in *Computer Vision – ECCV 2014*, Cham, Switzerland: Springer, 2014, pp. 298–313, doi: 10.1007/978-3-319-10605-2_20.
- [18] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 386–397, Feb. 2020, doi: 10.1109/TPAMI.2018.2844175.
- [19] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.
- [20] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition*, Jul. 2017, pp. 936–944, doi: 10.1109/CVPR.2017.106.

- [21] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Computer Vision – ECCV 2020*, Cham, Switzerland: Springer, 2020, pp. 213–229, doi: 10.1007/978-3-030-58452-8_13.
- [22] R. Kasturi *et al.*, "Framework for performance evaluation of face, text, and vehicle detection and tracking in video: data, metrics, and protocol," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 319–336, Feb. 2009, doi: 10.1109/TPAMI.2008.57.
- [23] C. D. Vleeschouwer and D. Delannay, "APIDIS dataset," *ispgroup.gitlab.io*. 2011. Accessed: Oct. 01, 2025. [Online]. Available: <https://ispgroup.gitlab.io/code/apidis/>
- [24] H. Qin, J. Yan, X. Li, and X. Hu, "Joint training of cascaded CNN for face detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2016, pp. 3456–3465, doi: 10.1109/CVPR.2016.376.
- [25] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/TPAMI.2016.2577031.
- [26] Y. Pang, Y. Yuan, X. Li, and J. Pan, "Efficient HOG human detection," *Signal Processing*, vol. 91, no. 4, pp. 773–781, Apr. 2011, doi: 10.1016/j.sigpro.2010.08.010.
- [27] D. Ozkan and P. Duygulu, "A graph-based approach for naming faces in news photos," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, 2006, pp. 1–6, doi: 10.1109/CVPR.2006.29.
- [28] K. Nayak and D. P. Dogra, "A comprehensive review on group re-identification in surveillance videos," *ACM Computing Surveys*, vol. 57, no. 6, pp. 1–35, Jun. 2025, doi: 10.1145/3711126.
- [29] M. M. Azad, Y. Cheon, I. Raouf, S. Khalid, and H. S. Kim, "Intelligent computational methods for damage detection of laminated composite structures for mobility applications: a comprehensive review," *Archives of Computational Methods in Engineering*, vol. 32, no. 1, pp. 441–469, Jan. 2025, doi: 10.1007/s11831-024-10146-y.
- [30] D. Yang and X. Liu, "Research on large-scale data processing and dynamic content optimization algorithm based on reinforcement learning," *Procedia Computer Science*, vol. 261, pp. 458–466, 2025, doi: 10.1016/j.procs.2025.04.234.
- [31] T. Zhao *et al.*, "A foundation model for joint segmentation, detection, and recognition of biomedical objects," *Nature Methods*, vol. 22, no. 1, pp. 166–176, Jan. 2025, doi: 10.1038/s41592-024-02499-w.
- [32] B. Babenko, M.-H. Yang, and S. Belongie, "Robust object tracking with online multiple instance learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1619–1632, Aug. 2011, doi: 10.1109/TPAMI.2010.226.
- [33] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 583–596, Mar. 2015, doi: 10.1109/TPAMI.2014.2345390.
- [34] S. Heidari *et al.*, "Exploring the application of knowledge transfer to sports video data," *Frontiers in Sports and Active Living*, vol. 6, Feb. 2025, doi: 10.3389/fspor.2024.1460429.
- [35] A. Senocak, T.-H. Oh, J. Kim, and I. S. Kweon, "Part-based player identification using deep convolutional representation and multi-scale pooling," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Jun. 2018, pp. 1813–1817, doi: 10.1109/CVPRW.2018.00225.
- [36] Y. M. Oo *et al.*, "Jump-aware: player position rectification and identification in dynamic sports using jump event spotting," in *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Jun. 2025, pp. 5925–5934, doi: 10.1109/CVPRW67362.2025.00591.
- [37] W. Li, T. Wang, Y. Liu, and K. Liu, "Perceptual image hashing fusing Zernike moments and saliency-based local binary patterns," *Computers*, vol. 14, no. 9, Sep. 2025, doi: 10.3390/computers14090401.
- [38] Y. Lai, "Cross-modal pedestrian re-identification technique based on multi-scale feature attention and strategy balancing," *Engineering Research Express*, vol. 7, no. 1, Mar. 2025, doi: 10.1088/2631-8695/adb93c.
- [39] M. Koshkina and J. H. Elder, "Towards long-term player tracking with graph hierarchies and domain-specific features," in *2025 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops*, Feb. 2025, pp. 1175–1185, doi: 10.1109/WACVW65960.2025.00140.
- [40] C. H. Chan and J. Kittler, "Sparse representation of (multiscale) histograms for face recognition robust to registration and illumination problems," in *2010 IEEE International Conference on Image Processing*, Sep. 2010, pp. 2441–2444, doi: 10.1109/ICIP.2010.5651933.
- [41] F. R. Sadikin, E. Hikmawati, and S. Aulia, "TrashTrack: a smart waste monitoring system with constraint-based greedy algorithm for optimized collection routes," in *2025 IEEE International Conference on Artificial Intelligence for Learning and Optimization*, Aug. 2025, pp. 262–268, doi: 10.1109/ICoAILO66760.2025.11155998.
- [42] L. Di Bella, Y. Lyu, B. Cornelis, and A. Munteanu, "HybridTrack: a hybrid approach for robust multi-object tracking," *IEEE Robotics and Automation Letters*, vol. 10, no. 7, pp. 7238–7245, Jul. 2025, doi: 10.1109/LRA.2025.3575311.
- [43] S. Li, H. Ren, X. Xie, and Y. Cao, "A review of multi-object tracking in recent times," *IET Computer Vision*, vol. 19, no. 1, Jan. 2025, doi: 10.1049/cvi2.70010.
- [44] R. K. Nath and D. Mitra, "Learning part-based features for vehicle re-identification with global context," *Applied Sciences*, vol. 15, no. 13, Jun. 2025, doi: 10.3390/app15137041.
- [45] X. Lu, C. Ma, B. Ni, X. Yang, I. Reid, and M.-H. Yang, "Deep regression tracking with shrinkage loss," in *Computer Vision – ECCV 2018*, Cham, Switzerland: Springer, 2018, pp. 369–386, doi: 10.1007/978-3-030-01264-9_22.
- [46] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, "Performance measures and a dataset for multi-target, multi-camera tracking," in *Computer Vision – ECCV 2016 Workshops*, Cham, Switzerland: Springer, 2016, pp. 17–35, doi: 10.1007/978-3-319-48881-3_2.
- [47] P. Ma, X. He, Y. Chen, and Y. Liu, "ISOD: improved small object detection based on extended scale feature pyramid network," *The Visual Computer*, vol. 41, no. 1, pp. 465–479, Jan. 2025, doi: 10.1007/s00371-024-03341-2.
- [48] S. Giancola, M. Amine, T. Dghaily, and B. Ghanem, "SoccerNet: a scalable dataset for action spotting in soccer videos," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Jun. 2018, pp. 1792–179210, doi: 10.1109/CVPRW.2018.00223.
- [49] J. Berclaz, F. Fleuret, E. Turetken, and P. Fua, "Multiple object tracking using k-shortest paths optimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 9, pp. 1806–1819, Sep. 2011, doi: 10.1109/TPAMI.2011.21.
- [50] A. Andriyenko and K. Schindler, "Globally optimal multi-target tracking on a hexagonal lattice," in *Computer Vision – ECCV 2010*, Berlin, Germany: Springer, 2010, pp. 466–479, doi: 10.1007/978-3-642-15549-9_34.
- [51] A. Vaghela, D. Lu, A. A. Verma, B. Chakravarthi, H. Wei, and Y. Yang, "MC-BEVRO: multi-camera bird eye view road occupancy detection for traffic monitoring," in *2025 IEEE 28th International Conference on Intelligent Transportation Systems*, Nov. 2025, pp. 4143–4149, doi: 10.1109/ITSC60802.2025.11423511.

- [52] R. Zhang, L. Wu, Y. Yang, W. Wu, Y. Chen, and M. Xu, "Multi-camera multi-player tracking with deep player identification in sports video," *Pattern Recognition*, vol. 102, Jun. 2020, doi: 10.1016/j.patcog.2020.107260.
- [53] H. Possegger, S. Sternig, T. Mauthner, P. M. Roth, and H. Bischof, "Robust real-time tracking of multiple objects by volumetric mass densities," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2013, pp. 2395–2402, doi: 10.1109/CVPR.2013.310.
- [54] T. Sekii, "Robust, real-time 3D tracking of multiple objects with similar appearances," in *2016 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2016, pp. 4275–4283, doi: 10.1109/CVPR.2016.463.
- [55] A. Crivellaro, M. Rad, Y. Verdie, K. M. Yi, P. Fua, and V. Lepetit, "A novel representation of parts for accurate 3D object detection and tracking in monocular images," in *2015 IEEE International Conference on Computer Vision*, Dec. 2015, pp. 4391–4399, doi: 10.1109/ICCV.2015.499.
- [56] A. Utasi and C. Benedek, "A 3-D marked point process model for multi-view people detection," in *CVPR 2011*, Jun. 2011, pp. 3385–3392, doi: 10.1109/CVPR.2011.5995699.
- [57] Y. Nishikawa, H. Sato, and J. Ozawa, "Multiple sports player tracking system based on graph optimization using low-cost cameras," in *2018 IEEE International Conference on Consumer Electronics (ICCE)*, Jan. 2018, pp. 1–4, doi: 10.1109/ICCE.2018.8326126.

BIOGRAPHIES OF AUTHORS



Adil Abdulhur Abushana    is a lecturer at the Department of Computer Science, University of Kufa, he received his B.Sc. degree in Mathematics from Salahaddin University, Iraq, in 1989, and his M.Sc. degree from the Department of Information Technology, University of Utara Malaysia, in 2009, respectively. He is currently pursuing a Ph.D. in the Department of Information Systems at Universiti Teknologi Malaysia. His research interests include pattern recognition and computer vision, with applications to biometrics. He was one of the participants who received a paper award at the IEEE conference in 2015. He can be contacted at email: adelshanna3000@gmail.com.



Riyadh Alsaeedi    holds a master's degree in Computer Science from Belarusian State University of Informatics and Radioelectronics, obtained in 2012. He is currently working in the Department of Statistics at the College of Administration and Economics at University of Wasit. He has published four research papers in the field of artificial intelligence. He can be contacted at email: reyad.tarik@gmail.com or ralsaeedi@uowasit.edu.iq.