Multi-View Football Player Tracking with Consistent Global IDs

Kirtiraj Tilakdhari Jamnotiya
B.Tech Mechatronics Engineering, IIT Bhilai
kjamnotiya@gmail.com — +91 9270865999

Abstract

This report outlines the implementation of a system that performs consistent multi-player tracking across two different video perspectives of a football match. The task involved detecting players, tracking them over time in each view, extracting motion features, and matching player identities across views based on movement. Given the absence of facial or jersey-number visibility, motion and temporal cues were used as the sole basis for identity matching.

1. Introduction

In multi-camera sports analytics, maintaining consistent identities of players across camera angles is essential for robust performance analysis. In this assignment, we are provided with two video feeds of the same football match:

- **broadcast.mp4**: a standard TV-like view
- **tacticam.mp4**: a top-down tactical view

The primary objective was to track each player in both videos and assign a global, consistent ID, despite differing camera angles and the lack of visible features such as jersey numbers or faces.

2. Approach Overview

The tracking pipeline was divided into the following main components:

1. Player Detection using a fine-tuned YOLOv11 model.
2. Per-view tracking using a custom centroid-based tracker.
3. Motion feature extraction from bounding box trajectories.
4. Track matching across views using cosine similarity and the Hungarian algorithm.
5. Visualization and video annotation with global player IDs.

3. Methodology

3.1. Player Detection

A pretrained YOLOv11 model (`best.pt`) was used to detect player bounding boxes in both videos. The detections were filtered by confidence and stored per frame.

3.2. Tracking

Each view was processed independently using a custom *centroid tracker*. This tracker maintained identities based on proximity of bounding box centers frame-by-frame and handled ID assignment robustly over time.

3.3. Motion Feature Extraction

For each track ID, a motion signature was generated:

- The trajectory of the bounding box center over time was recorded.
- Delta vectors $(\Delta x, \Delta y)$ between successive frames were calculated.
- The final motion vector was padded or truncated to a fixed length (60 values).

3.4. Cross-View Matching

To match players between the two videos:

1. Cosine similarity was computed between each pair of motion vectors.
2. The Hungarian algorithm was applied to determine optimal one-to-one assignments.
3. A global ID was assigned to each matched pair.

3.5. Annotation

Each video was re-processed to draw bounding boxes and display the assigned `global_player_id`. The output videos demonstrate visually that player identities are maintained across both views.

4. Challenges Faced

- **No visible facial or jersey-number features**: Identification had to rely entirely on temporal movement.
- **Camera angle distortion**: Perspective distortion caused large differences in player positioning between views.
- **Occlusions and overlapping players**: This often broke tracking temporarily or caused track switching.
- **Synchronization mismatch**: Frame alignment across videos had to be assumed or estimated.

5. Possible Improvements

To enhance the system's robustness and accuracy, the following improvements can be made:

- **Pose-based action recognition**: Use OpenPose or MediaPipe to extract player pose and compare based on movement style.
- **Ball interaction modeling**: Track the ball and include interaction patterns as additional identity features.
- **Homography estimation**: Project bounding boxes to a normalized field coordinate space for easier matching.
- **Advanced tracking models**: Incorporate ByteTrack or DeepSORT for more reliable long-term identity tracking.

- **Temporal synchronization**: Estimate and correct for frame misalignment using cross-correlation or manual sync.

6. Conclusion

This project successfully demonstrates the feasibility of tracking football players across distinct camera views using motion-based signatures. Despite the absence of appearance features, the use of temporal dynamics and optimal matching techniques allowed consistent global IDs to be assigned and visualized. The pipeline serves as a foundational approach that can be extended with richer visual and semantic features in future deployments.