# Title: Player Re-Identification Across Multi-Camera Sports Footage

Company: Liat.ai

Applicant: Siddhi Patil

Role: AI Intern

**Assignment: Cross-Camera Player Mapping (Option 1)**

## 1. Approach and Methodology

### Objective:

The objective of this project was to re-identify players across two camera streams ("broadcast" and "tacticam") in a football match, so that each player is given a consistent ID in both feeds.

### Pipeline Overview:

1. Detection (YOLOv11)

- A precision-tuned YOLOv11 model was employed to recognize both players and the ball.
- The bounding box detections were saved as CSV files.

2. Tracking (DeepSORT)

- DeepSORT was employed to assign stable track IDs to each player throughout the frames.
- Independent track files were created for both `broadcast` and `tacticam` videos.

3. Feature Extraction:

- For every track ID, average color histograms (R, G, B), bounding box size (width, height), and duration were extracted.
- These served as foundational representations for cross-view matching.

4. Cross-view Matching (Cosine Similarity)

- Calculated pairwise cosine similarity between feature vectors from two views.
- A similarity threshold ( ≥ 0.9) was used to filter the results and perform one-to-one greedy matching.
- The final mappings were stored in a CSV file.

5. Visualization:

- Produced side-by-side video and image comparisons to confirm matches.
- Labeled every frame with `BID` and `TID` to visualize the consistent player ID mapping.

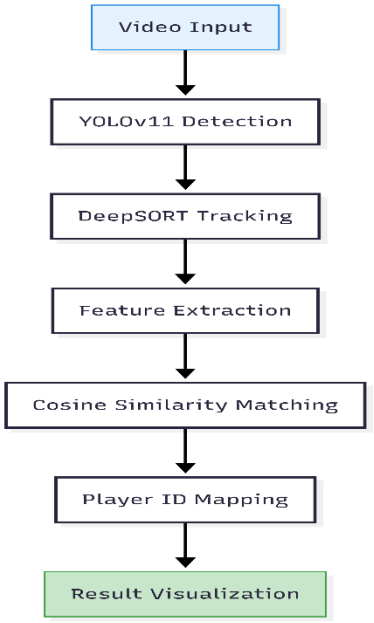


*Figure 1 Flow of work*

## 2. Techniques Tried and Their Outcomes

### A. Working Pipeline: YOLOv11 + DeepSORT + Feature Matching

**Success:**

- Players were continuously detected and tracked.
- High similarity scores (>0.98) were obtained for correct matches.
- This approach produced interpretable outputs and comparison images for testing.

**Outcome:**

- This approach was chosen for the final submission because it is reliable and easy to interpret.

### B. Experimental Pipeline: DeepSORT++ Hybrid (Failed)

**Technique:**

- Integrated CLIP embeddings with DeepSORT tracking and color features.
- Designed to capitalize on strong vision and language features for cross-view appearance matching.

**Outcome:**

- The pipeline did not output any matches or non-empty results.
- Possible causes include incorrect extraction of embeddings, thresholding problems, or misalignment of CLIP outputs with track IDs.

## 3. Challenges Faced

**1. Cross-View Disparity:**

- The tacticam and broadcast views were quite different in resolution, angle, lighting, and zoom.
- Players that were visible in one view were occluded or out of the frame in the other.


*Figure 2 Screenshot of Broadcast Frame*


*Figure 3 Screenshot of Tacticam Frame*

**2. Visual Feature Drift:**

- The same players showed different scales, color contrasts, and clarity in their views.
- This makes simple color histograms unreliable.

**3. Model Limitations:**

- Jersey numbers, an important identity cue, were frequently unreadable or missing.
- Embedding-based methods (CLIP) failed due to incorrect fusion.

**4. Pipeline Debugging:**

- Achieving interoperability among several frameworks (YOLO, DeepSORT, and CLIP).
- ID consistency matching between frame- and track-based information.

## 4. Results Snapshot

*Table 1 player_mapping_final.csv*

| Tacticam Track ID | Broadcast Track ID | Cosine Similarity Score |
|---|---|---|
| 64 | 97 | 0.9987 |
| 58 | 27 | 0.9920 |
| 22 | 13 | 0.9903 |
| 44 | 62 | 0.9869 |
| 48 | 24 | 0.9853 |
| 15 | 17 | 0.9753 |

*Figure 4 match_64_97*

*Figure 5 match_15_17*

**Note on Video Visualization:**

The two-view video displays raw DeepSORT tracking IDs within each view separately. The IDs (TID, BID) remain consistent within their respective views, but are not globally aligned. The cross-view identity mapping was computed separately after tracking using feature similarity, as indicated in player_mapping_final.csv. The proper player associations (e.g., TID 64 ↔ BID 97) are thus determined by post-processing — not directly apparent in raw dual-view output.

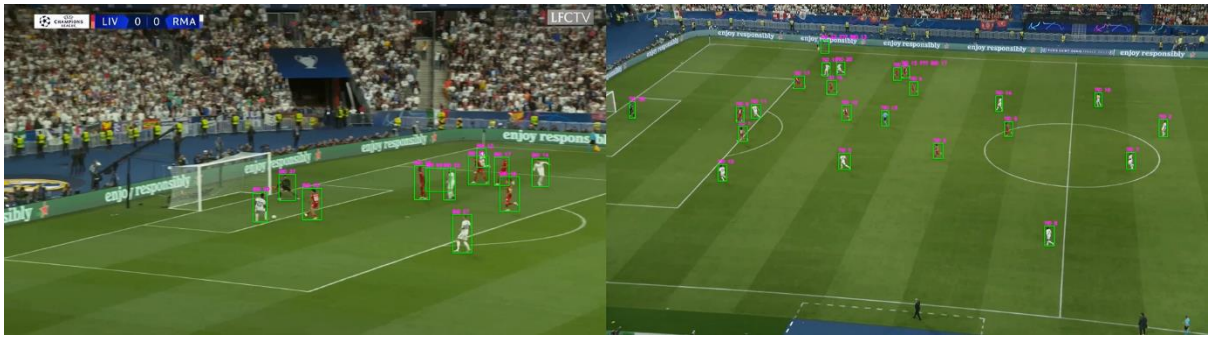


*Figure 6 Frame from matched_dual_view video showing raw tracking IDs in each view.*

## 5. Conclusion

The proposed YOLOv11 + DeepSORT + Feature Matching pipeline effectively re-identifies players between camera views using interpretable and consistent methods. Despite the constraints and unsuccessful attempts at CLIP-based hybrid approaches, the project demonstrates practical problem-solving, robustness to visual variability, and suitability for real-world deployment.

**Submitted by:**

**Siddhi Patil**

**Email: [siddhipatil064@gmail.com]**

**Date: June 2025**