Cross Camera Player Mapping

## 1. Introduction:

This project focuses on identifying and assigning consistent IDs to football players across two short 720p videos recorded from different camera angles. The challenge lies in matching the same player in both videos despite differences in camera zoom, perspective, lighting, and partial occlusions.  
  
The goal is to build a system that tracks and re-identifies each player accurately across both video streams without relying on facial features or manual annotation.

## 2. Objective

- To detect and track football players in two different videos.  
- To assign each player a consistent global ID across both videos.  
- To ensure ID consistency even if a player temporarily moves out of the frame.  
- To maintain robustness despite challenges like different zoom levels, blurry visuals, and poor player visibility.

## 3. System Overview

The system processes two 15-second football videos:  
- Video A is used to extract player features and assign base global IDs.  
- Video B players are matched with Video A using appearance-based embedding similarity.  
  
The complete pipeline includes:  
- Object detection and tracking using fine tuned YOLOv11 + BoT-SORT.  
- Embedding extraction using a ReID model (osnet\_x1\_0).  
- Cosine similarity-based matching to assign global player IDs.

## 4. Methodology

5.1 Detection and Tracking  
- A fine-tuned YOLOv11 model detects objects in both videos.  
- BoT-SORT tracker assigns short-term track IDs per video.  
- Only players is considered.  
  
5.2 Embedding Extraction  
- For each detected player in Video A, cropped frames are saved.  
- These are passed through the osnet\_x1\_0 ReID model (from TorchReID) to extract 512-dim embeddings.  
- The embeddings are averaged to form a robust representation per player.  
  
5.3 ID Matching in Video B  
- For each new detected player in Video B:  
 - A single ReID embedding is extracted.  
 - Cosine similarity is computed with all averaged embeddings from Video A.  
 - If the highest similarity is above a threshold (0.75), the corresponding global ID is assigned.  
 - Otherwise, a new global ID is generated.

## 5. Output

- Two annotated videos are saved:  
 - output\_A/video\_A\_annotated.avi  
 - output\_B/video\_B\_annotated.avi  
- Each video has player bounding boxes with global IDs overlaid.  
- IDs remain consistent for players detected in both videos.

## 6. Approaches, Techniques, and Models Tried

This section highlights various experiments performed to improve accuracy.  
  
7.1 ReID Models Tested  
- osnet\_x1\_0 (final model): Lightweight yet accurate ReID network designed for real-time tracking.  
- resnet50: Initially tried but produced weaker embeddings due to poor resolution.  
- Attempted agw\_resnet50 but was unavailable via TorchReID default registry.  
  
7.2 Embedding Improvements  
- Initially used single-frame embeddings → very unstable.  
- Switched to temporal averaging across multiple frames per player, significantly improving consistency.  
  
7.3 Similarity Techniques  
- Cosine Similarity (Final Approach): Most reliable and lightweight.  
- Histogram Comparison (Removed): Added noise in low-resolution videos.  
- Dominant Color Matching (Removed): Failed in real-world videos due to jersey overlap and lighting.

## 7. Challenges Faced

- Videos had very low visual quality with barely visible faces.  
- Zoom-level differences caused players to appear at drastically different scales.  
- Different fields of view meant that some players only appeared in one video.  
- ReID on small blurry crops led to unreliable identity embeddings.  
- Ball tracking was irrelevant for the task and was ignored.

## 8. Results and Evaluation

- Final method achieves moderate accuracy in maintaining consistent IDs across videos.  
- Performance degrades when:  
 - Players are occluded  
 - Appear briefly  
 - Drastically change size or angle

## 9. Conclusion

This project demonstrates a working system for multi-camera player re-identification using deep learning. It combines object detection, deep feature extraction, and cosine similarity matching to handle real-world football video challenges.  
  
The pipeline is clean, modular, and extensible, with room for further improvement via better models or longer video inputs.

## 10. Related Studies

Some studies have explored **motion-based** tracking, where each player's movement across frames is stored as a temporal history. These motion patterns are then used to help maintain consistent identities, even when appearance features are weak. While effective in longer or synchronized videos, such methods depend heavily on reliable tracking and continuous player visibility. Recent approaches also combine appearance and motion cues to improve re-identification in sports settings

[Paper](https://arxiv.org/abs/2211.02125)

## 11. Potential Improvements with More Resources

With additional computational and data resources, the system's accuracy can be significantly improved:

* **Use of Advanced ReID Models**: Integrating transformer-based models like **TransReID** or **MGN** could produce more robust embeddings, especially in low-quality video.
* **Fine-Tuning on Domain-Specific Data**: Training or fine-tuning the ReID model on **football-specific datasets** would help it learn jersey patterns, player posture, and field context.
* **Temporal Modeling**: Incorporating **motion history** or **trajectory prediction models** would allow for better identity consistency when players leave and re-enter frames.
* **Multi-View Fusion**: Synchronizing and aligning both video feeds could enable **multi-camera feature fusion**, improving matching across angles.
* **Higher-Resolution Input**: Using **1080p or 4K videos** would allow for clearer crops and better embedding quality, especially for distant or partially visible players.

These enhancements would likely result in more stable and accurate player identification across varying conditions.