Cross-Camera Player Mapping using YOLOv11 and DeepSORT

This document details the approach and methodology used for cross-camera player mapping, along with the techniques tried, their outcomes, and the challenges encountered during the process.

s **by shreyansh rao**

Made with GAMMA

Approach and Methodology

1 Objective

To match players from two video sources (Tacticam and Broadcast) of the same gameplay using consistent player IDs, regardless of differing camera views.

2 Detection

Used YOLOv11 object detection model to detect players frame-by-frame in both videos.

3 Tracking

Applied DeepSORT for multi-object tracking within each camera feed to assign local track IDs to players.

4 Feature Extraction

For each track, computed the mean position (center of bounding boxes) as a simple positional feature for cross-view matching.

5 Cross-Camera Matching

Matched players across the two views by comparing their spatial features (mean positions) using the Hungarian Algorithm (linear sum assignment).

Techniques Tried and Their Outcomes

YOLOv11 + DeepSORT

Gave accurate detections and robust tracking within each view. It helped establish consistent local IDs per video.

Positional Feature Matching

Worked reasonably well when both cameras had overlapping field of view. Players in central regions matched more reliably.

Hungarian Algorithm

Provided a cost-optimized one-to-one mapping between Tacticam and Broadcast IDs.

Optional Attempt

Tried testing cosine similarity on average bounding box sizes and player movement directions, but it was less effective than center position matching alone.

Challenges Encountered

Inconsistent Track IDs Across Views

Even with good local tracking, linking players across different perspectives was difficult due to camera angle differences.

Occlusion and Crowding

Players overlapping or being occluded in one camera view affected detection/tracking accuracy.

Appearance Mismatch

No consistent visual descriptor (like jersey number or facial features) was available, which limited matching precision.

Limited Time for Feature Tuning

Did not implement appearance embeddings or re-ID networks, which could improve cross-view matching.