

# Player Re-Identification in Sports Footage

## Overview

This is a project based on player detection, tracking, and team classification using video footage of a football match. I've tried combining detection models with team recognition using jersey colors, and also added a basic action recognition layer.

## 📁 Repository Contents

```
.
├── models/
│   └── yolo_players.pt          # YOLO model trained for football player
and ball detection
├── sort/
│   └── sort.py                 # SORT tracking code
├── videos/
│   └── 15sec_input_720p.mp4    # Test video
├── results/
│   └── final_clip_output.mp4   # Output video after tracking and labeling
├── detect_players.ipynb        # Main Python code
└── Player Re.docx             # Readme and Report
```

## ⚙️ Setup Instructions

### Requirements

- Python 3.8 or above

### Install Dependencies

```
pip install torch torchvision opencv-python numpy pillow
git+https://github.com/openai/CLIP.git ultralytics
```

Make sure the YOLO model is placed inside `models/yolo_players.pt`

### To Run

```
python vlm_action_tracking.py
```

This will process the video and show the player IDs, team colors, and basic action labels.

---

# Project Report

## 🎯 Objective

To detect players and the ball in a football match, identify which team they belong to using jersey color, and also label simple actions like movement using a short sequence of frames.

## 🔍 What I Did

1. **Detection:** Used YOLO model for detecting players and the ball.
2. **Tracking:** SORT tracker for keeping ID consistent.
3. **Team Classification:**
  - Took the upper part of the player's body (jersey area).
  - Used CLIP with prompt sentences like “red jersey” or “blue jersey” to decide team.
4. **Action Detection:**
  - Kept a short buffer of 16 cropped frames per player.
  - Ran it through a pretrained R3D model to label basic actions.

## ✅ What Worked and What Didn't

Method	Description	Result
HSV Histogram	Color-based detection (basic)	Too sensitive to lighting
KMeans Clustering	Color clustering	Not consistent in complex scenes
CLIP (VLM)	Using natural language + jersey image	Much more reliable
R3D Action Model	Recognizing actions in 16-frame clip	Worked decently for demo level

## ⚠️ Problems Faced

- CLIP was a bit slow on CPU, had to move to GPU.
- Needed exactly 16 frames for action recognition.
- Jersey detection didn't always work well if the player was turned away.

## 🔧 Things I Would Add

- Replace R3D with more advanced models like SlowFast.
- Try recognizing not just team but player roles (like goalkeeper).
- Connect it with known players using face ID or jersey number.

---

## My Experience

I have tried working with OpenCV and YOLO before for some small projects. Recently, I've been learning about CLIP and its potential to connect vision and text, and this was my first time using it for team classification. For action recognition, I explored pretrained models like R3D and tried using it on player crops.

Though not perfect, the overall results gave a good demonstration of combining tracking, language, and visual models for sports analysis.

If there's more time, I'd continue exploring deeper temporal models and maybe even add commentary tagging or tactic detection.

---

Feel free to reach out if anything needs clarification.