Assignment: Player Detection and Tracking in Sports Video

This project implements a computer vision pipeline for **player detection and tracking** in sports footage. The main objective was to detect players and the ball in each frame, assign consistent IDs, and visualize movement and game events over time.

The pipeline includes:

- **Object Detection:**
  A fine-tuned YOLOv11 model was used to detect players and the ball frame by frame.
- **Tracking:**
  ByteTrack was applied to link detections across consecutive frames and assign unique IDs.
- **Camera Motion Compensation:**
  Optical flow-based estimation was used to adjust player positions for camera movement.
- **Perspective Transformation:**
  Positions were transformed into a normalized top-down court coordinate system.
- **Post-Processing and Visualization:**
  - Player speed and distance traveled were estimated.
  - Team assignment was performed using color clustering.
  - Ball possession was assigned to the nearest player.
  - An annotated output video was generated showing bounding boxes, player IDs, team colors, camera movement, and possession statistics.

Note:
While the pipeline was functional end-to-end, some challenges remained with **player re-identification**—particularly maintaining consistent IDs when players left and re-entered the frame. Improvements would include embedding-based re-ID models and more robust temporal smoothing.

Methodology and Tools

Approach Overview

The project was implemented in **Python**, structured as a modular pipeline combining detection, tracking, camera motion estimation, and visualization. The overall flow:

1. **Video Loading:**
    - Frames were read using OpenCV.
    - Preprocessing was applied as needed.
2. **Object Detection:**
    - A fine-tuned YOLOv11 model was used to detect players and the ball.
    - Batched inference was performed to improve efficiency.
3. **Tracking:**
    - ByteTrack associated detections across frames to assign IDs.
    - Tracking data was stored for reproducibility using pickle stubs.
4. **Camera Motion Estimation:**
    - Optical flow (Lucas-Kanade) estimated camera movement between frames.
    - Player positions were compensated for this motion.
5. **Perspective Transformation:**
    - Detected positions were mapped to a standardized court coordinate system via a homography matrix.
6. **Post-Processing:**
    - **Team assignment:** KMeans clustering of jersey colors.
    - **Ball possession:** Nearest-player assignment.
    - **Speed & distance:** Estimated using frame-to-frame displacements.
7. **Visualization & Output:**
    - Annotations were drawn onto frames (bounding boxes, IDs, speed, possession, camera movement).
    - The final video was saved as MP4.

Challenges and Next Steps

Challenges Encountered

While the pipeline performed the main detection and visualization tasks successfully, several challenges were observed:

- **Player Re-Identification Consistency**
  - Player IDs occasionally changed when players exited and re-entered the frame.
  - No embedding-based appearance re-identification was implemented, which limited temporal consistency.
- **Detection Variability**
  - The YOLOv11 model sometimes missed detections in crowded frames or when players were partially occluded.
  - Occasional bounding box jitter and flickering.
- **Team Assignment Accuracy**
  - KMeans color clustering worked but was sensitive to lighting and viewpoint changes.
  - Some players were incorrectly assigned team colors due to jersey color similarities.
- **Camera Motion Estimation Noise**
  - Optical flow sometimes produced noisy estimates, especially in frames with fast pans or limited features.

What Could Be Improved

If more time and resources were available, I would focus on:

- **Embedding-based Re-Identification**
  - Integrate a re-ID model to extract appearance features and maintain player identity over occlusions.
- **Tracking Smoothing**
  - Implement Kalman filtering or trajectory smoothing to reduce ID switching and jitter.
- **Detection Enhancements**
  - Apply test-time augmentation and confidence threshold tuning.
  - Explore combining multiple detection models for improved robustness.
- **Better Color Segmentation**
  - Incorporate learned color histograms or fine-tuned classifiers for jersey identification.
- **3D Perspective Calibration**
  - Use multiple known court landmarks to improve homography accuracy.

Summary

What Was Implemented

- **Detection and Tracking**
  - Loaded videos and detected players and the ball using YOLOv11.
  - Tracked detections with ByteTrack to assign IDs frame by frame.
- **Camera Motion Compensation**
  - Applied optical flow estimation to adjust positions relative to camera movement.
- **Perspective Transformation**
  - Mapped positions to a standardized court coordinate system via homography.
- **Team Assignment and Ball Possession**
  - Used KMeans clustering for team color classification.
  - Assigned ball possession to the nearest player in each frame.
- **Speed and Distance Estimation**
  - Calculated approximate speed and distance covered over time windows.
- **Visualization and Export**
  - Drew bounding boxes, IDs, team colors, speed, and ball control stats.
  - Generated and saved a final annotated video.

What Remains

- **Robust Re-Identification**
    - Re-ID consistency was limited; IDs could switch when players left/re-entered.
    - No deep appearance embeddings were implemented.
- **Improved Stability**
    - Detection flicker and bounding box jitter occasionally occurred.
    - Kalman smoothing and trajectory filtering could improve this.
- **Advanced Team Classification**
    - Simple KMeans color clustering was sensitive to lighting and viewpoint.

Reflections

This project provided hands-on experience with:

- Real-world sports video analysis challenges.
- Combining detection, tracking, and post-processing modules.
- Working with noisy detections and partial occlusions.