Project Report: Multi-Camera Multi-Person 3D Pose Estimation in Cricket

Proposed by: Aditya Kushwaha, Aksh Neelesh Shah, Aditya Melinkari

1. Introduction

In cricket, tracking the precise movements of players in 3D can provide valuable insights into performance analysis, injury prevention, and game strategy. Traditional tracking methods rely on wearable sensors or single-camera setups, which have limitations such as occlusions, inaccurate depth estimation, and dependency on controlled environments.

This project aims to develop a **multi-camera, multi-person 3D pose estimation system** that accurately reconstructs players' body movements from multiple synchronized video feeds. Our approach leverages **computer vision, deep learning, and geometry-based reconstruction** techniques to estimate 3D poses with high accuracy.

2. Problem Statement

The primary challenges in achieving multi-camera 3D pose estimation for cricket include:

1. **Occlusions** – Players frequently block each other, leading to missing body parts in some camera views.
2. **Camera Synchronization** – Different cameras must be precisely time-synchronized to align frames correctly.
3. **Player Identification and Tracking** – Ensuring that each player is consistently tracked across multiple frames and different viewpoints.
4. **3D Pose Reconstruction** – Estimating accurate joint positions by triangulating 2D keypoints from multiple camera perspectives.
5. **Real-Time Processing** – The system should process data efficiently to be useful in live analysis.

3. Approach

Our solution follows a **five-step pipeline** for **multi-camera 3D pose estimation**:

Step 1: Camera Setup and Calibration

Before capturing data, we set up multiple cameras around the field. Each camera must be **calibrated** to obtain its **intrinsic (lens properties) and extrinsic (position and orientation) parameters**.

- **Calibration Method**: We use **Zhang's calibration method** with a checkerboard pattern to estimate the camera parameters.
- **Synchronized Recording**: Cameras are synchronized to capture frames at the same timestamp using **hardware sync** or **frame alignment algorithms**.
- **Multi-View Geometry**: Camera projections are modeled using **epipolar geometry** to enable 3D triangulation.

Step 2: 2D Pose Estimation Using Mediapipe & OpenPose

Each camera independently detects **2D human keypoints** in the captured frames.

- **Algorithm Choice**: We use **Mediapipe Pose** and **OpenPose** for detecting **landmark joints (shoulders, elbows, knees, ankles, etc.)**.
- **Output**: A **set of 2D keypoints** for each player in each camera view.
- **Occlusion Handling**: If a joint is missing in one camera due to occlusion, we rely on other camera views for estimation.

Step 3: Multi-Camera Association & Player Tracking (DeepSORT + Appearance Matching)

Since multiple players exist in the field, we need to:

- **Assign the correct identity to each player across frames**.
- **Ensure the same player is tracked across different camera views**.

We achieve this by:

1. **DeepSORT-based Tracking**: Assigning unique IDs to each player using a **Kalman Filter + CNN-based Re-ID model**.
2. **Homography-Based Matching**: Projecting players detected in one view to another using a **homography transformation**.
3. **Feature Matching**: Extracting **appearance features (e.g., jersey color, player number)** to match players across cameras.

Step 4: 3D Pose Reconstruction Using Multi-View Triangulation

Once 2D joint positions are detected and player identities are assigned across multiple cameras, we compute the **3D joint positions** using **triangulation**.

- **Direct Triangulation**: Using the **Direct Linear Transform (DLT)** method to compute 3D positions.
- **Bundle Adjustment**: Refining 3D positions by minimizing **projection errors** using **Levenberg–Marquardt optimization**.
- **Filtering Outlier Joints**: If a joint estimate is inconsistent across different camera views, we apply a **RANSAC-based outlier rejection**.

Step 5: Post-Processing & Visualization (Open3D + Blender Rendering)

After obtaining **accurate 3D poses**, we refine and visualize the results:

1. **Smoothing & Temporal Consistency**
   - Using **Savitzky-Golay filtering** to reduce jitter.
   - Applying **Kalman filters** for smooth transitions.
2. **3D Visualization**
   - Displaying the **animated skeletons** of players in **Open3D**.
   - Rendering high-quality **3D cricket player models** in **Blender** for realistic visualization.

4. System Workflow

1. **Input Video Feeds**
   - Multiple synchronized cameras capture cricket players from different angles.
   - Each camera's position is calibrated in the global coordinate system.
2. **2D Pose Detection**
   - Mediapipe and OpenPose extract **body landmarks** from each frame.
3. **Multi-View Association & Tracking**
   - DeepSORT + feature matching ensures each player has a unique ID across views.
4. **3D Reconstruction**
   - The **triangulation process** estimates each player's 3D joint locations.
5. **Visualization & Analysis**
   - Final 3D poses are rendered and analyzed in Open3D and Blender.

5. Challenges & Limitations

- **Lighting & Camera Quality**: Poor lighting conditions can affect 2D pose detection accuracy.
- **Occlusions & Player Overlaps**: Heavy occlusions require robust tracking and interpolation techniques.
- **Processing Time**: Real-time performance demands efficient implementation using **GPU acceleration** (e.g., TensorRT).

6. Deliverables

- **Fully Functional Multi-Camera 3D Pose Estimation System**
- **Codebase with Documentation**
- **Demo Video Showing System Performance**
- **Technical Report & Evaluation Metrics**

10. Conclusion

This project presents an innovative approach to **multi-camera, multi-person 3D pose estimation in cricket**, combining **deep learning, geometric vision, and tracking algorithms**. The developed system has applications in **sports analytics, performance evaluation, and biomechanics research**.

By leveraging **computer vision** techniques and **multi-view geometry**, our approach ensures accurate **real-time tracking of players** on the field, opening doors for **advanced AI-powered sports analysis**.