An AI Referee of Badminton Matches

Jing-En Huang (黃靖恩), Department of Mathematics, National Taiwan Normal University

40840210s@gapps.ntnu.edu.tw, Advisor: Prof. Mei-Heng Yueh (樂美亨)

Abstract

In the past, gathering information for professional badminton matches required coaches to manually record the tactics of both players and opponents during the game. However, due to the high speed of shuttlecock and player movement, collecting such information posed difficulties. In recent years, although the use of computer vision with multiple cameras based on deep learning for tracking the shuttlecock and players has become relatively mature, it is not allowed to set up multiple cameras for information gathering in professional game settings. Therefore, inspired by "Teaching Computers to Watch Badminton - Taiwan's First Competition Combining AI and Sports," this study focuses on analyzing single-camera perspective broadcasts of badminton matches.

Problem Description

Figure 1. One frame from the video under analysis.

(The badminton video dataset used in this study was entirely provided by the organizers of the competition.)

The competition provides a training set consisting of 800 annotated videos, where each round is defined as a single scoring opportunity. The annotations include information such as the player, shot type (e.g., smash, slice), and the specific shot sequence (e.g., which shot occurs on which beat). Participants are required to analyze an additional set of 400 test videos, which include the shot count per round, three-dimensional positions of the shuttlecock during each shot, player positions, shot types, forehand/backhand shots, and the winning player. The implementation of this task poses the following challenges:

1. Accurately tracking a shuttlecock is challenging due to its diminutive size and rapid movement. (Figure 2.)

Figure 2.

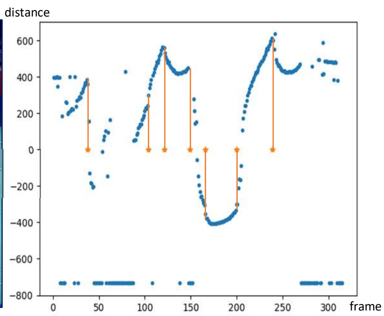

Figure 3.

2. In 720*1280 resolution videos, there are numerous color patches that resemble the shape of a shuttlecock, leading to imperfect tracking results. Consequently, relying solely on the trajectory of the shuttlecock to determine the shot count per round is not feasible. (Figure 3. The blue dots are the results of the shuttlecock detections, where the orange stars are the Ground-Truth of the shot frames.)
3. Estimating the three-dimensional coordinates of the shuttlecock in a single-camera view theoretically requires at least two different camera perspectives for calculation.

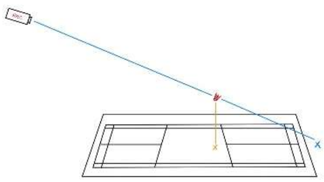

Figure 4. Without the use of projection backtracking estimation, the obtained coordinates of the ball (blue crosses) will deviate significantly from the correct coordinates (orange crosses).

4. Recognition of player's shot types presents a challenge.
5. Player's body skeletal tracking is susceptible to interference from nearby referees and spectators. (Figure 5.)

Player being occluded.

Figure 5.

Results and Discussion

Firstly, accurately determining the shot count per round is crucial. PoseC3D action recognition was used to estimate shot count by detecting player's swing actions, improving accuracy from 43% (TrackNetV2) to 67%. Referee's head orientation (Figure 6.) near the court aligns with the shuttlecock direction. Utilizing Faster-RCNN for human detection and HRNet-w32 for pose estimation, shot count accuracy improved to 83%. Shot count estimation in videos without referees relied on player's swing actions.

Figure 6.

Figure 7. (Words detected by EasyOCR)

Figure 8. (The RGB information of the court after removing the players.)

Secondly, EasyOCR (Figure 7.) recognized venues to capture referee position. Venues were classified for masking and detection of referee head orientations. RGB information (Figure 8.) of venues was used for pairing. Frames were averaged and compared using SSIM to effectively group the remaining unclassified videos.

Thirdly, the referee's line of sight was calculated from skeletal keypoints formed by the eyes and mouth to determine shot count.

Fourthly, estimating the shuttlecock's three-dimensional coordinates considers only the y-axis as the camera is positioned behind. Disregarding the z-axis, the shuttlecock is expected to lie on the line segment AB connecting the players. TrackNetV2's tracking results determine its position along AB based on velocity and players' dominant foot positions.

Finally, despite PoseC3D being a state-of-the-art model, its accuracy remains around 50%. Data augmentation techniques will be experimented with to enhance accuracy.

Figure 9. Final Result

Reference

[1] N. E. Sun, Y. C. Lin, S. P. Chuang, T. H. Hsu, D. R. Yu, H. Y. Chung, T. U. Ik - TrackNetV2: Efficient Shuttlecock Tracking Network

[2] Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun - Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks

[3] Zigang Geng, Ke Sun, Bin Xiao, Zhaoxiang Zhang, Jingdong Wang - Bottom-Up Human Pose Estimation Via Disentangled Keypoint Regression

[4] Haodong Duan, Yue Zhao, Kai Chen, Dahua Lin, Bo Dai - Revisiting Skeleton-based Action Recognition

[5] M. H. Liao, Z. H. Zou, Z. Y. Wan, C. Yao, X. Bai - Real-Time Scene Text Detection with Differentiable Binarization and Adaptive Scale Fusion

[6] Amir Shahroudy, Jun Liu, T. T. Ng, G. Wang - NTU RGB+D: A Large Scale Dataset for 3D Human Activity Analysis