Sports Coaching Using Computer Vision

This project utilizes computer vision to evaluate a player's performance when playing squash and aims to help players improve their game by providing them with statistics that indicate the percentage of time of the game they spend in a commanding position from where they can attack vs. percentage of time they spend in non-commanding positions from where they can only defend.

Squash as a sport is played in enclosed court where a 640x975 cm floor space, front wall, two side walls and a back wall make up the playing surface. Once a server initiates a rally, the opponent aims to strike the ball before is bounces twice and aims to return it to the front wall without any bounces. Since squash is played in an enclosed space, a player standing in the center of the court has a commanding position as he/she has the shortest path to the ball at any corner of the court. Thus, the more time a player spends in the center of the court, the better. This project uses computer vision to calculate the percentage of time a player stays in the center vs. off-center. The workflow involves obtaining video frames, detecting players, tagging players, tracing players, plotting their tracks on a 2D grid, creating heatmaps of their positions and, finally, computing a score.

Video frames showing (A)two players playing, (B)players with PERSON identification from YOLO & (C)players tagged using OpenCV similarity features

Obtaining Video Frames
Video frames may be obtained through a video file or through a webcam. Some courts in colleges and universities now have camera systems mounted behind the back-wall of the court that can be used to obtain video.

Detecting Players
The Darknet YOLO models, specifically YoloV3 and YoloV4 were used in this project to detect objects. Once detection was complete, only the *person* class was retained since the aim was to identify players in a given frame. Note that the typical model inference is fairly slow. On a CPU, it runs at a rate of 0.5 frames per second. The model can be made faster by training it further to detect the person class only and ignoring all other classes.

Tagging Players
In this application, in addition to identifying persons, we also need to keep a track of players. Player tagging was done using similarity features available from within OpenCV framework. In OpenCV, a bounded box for a frame is compared to bounded boxes from previous frames to figure out whether the new box represents player 1 or player 2. Numerous other solutions are available for tagging, including

the DeepSort algorithm, averaging colors within a bounded box or to use Random Forest Regressor to identify most common colors.

Tracing Players

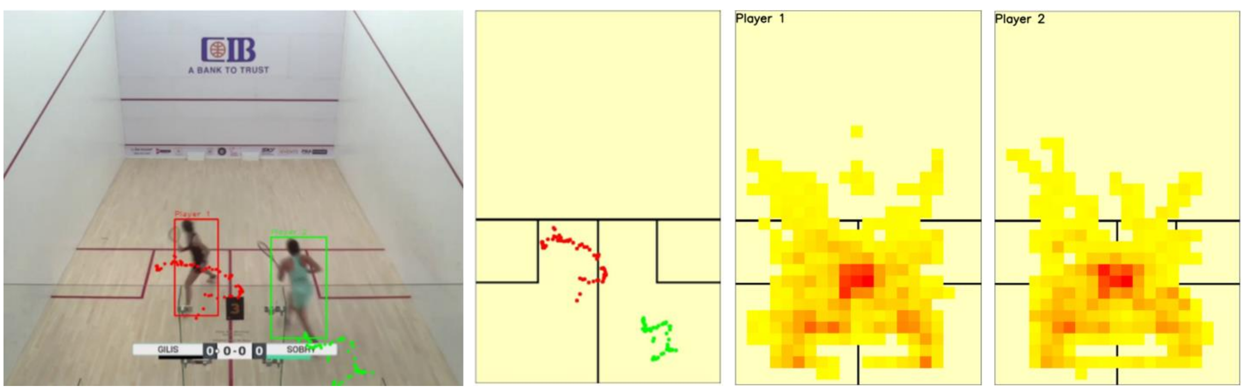
Once tagging is complete and bounding boxes have been drawn for the players, the bottom center position was taken as the position for the player. While this approach is adequate, it may be further refined by using the pose-estimator pre-trained model available online as well. The pose-estimator model will identify specific joints in a body and the center point between two ankles can be taken as the player location.

3D to 2D mapping

Squash is typically filmed from over the backwall, so the court floor is always viewed at an angle. In order to visualize the floor from the top, the camera perspective needs to be transformed. OpenCV provides built-in functions, specifically getPerspectiveTransform and warpPerspectiveTransform that facilitate this transformation. This resultant transformation matrix can be used with other points to visualize the tracks from a 3D camera view to a 2D surface.

Heat Maps

Once 2D tracking information is available, heatmaps can be created using the Seaborn package in Python. The OpenCV package also provides features to overlay shapes on an image to create the heat maps. Once a heat map is available, a previously determined scoring methodology can be used to compute a score. For the purposes of this project, staying in the center got a player 100 points and staying in the corners got 0 points. The result is a weighted average based on the video stream processed.



Series of figures showing (A) two players playing with players tagged and traced, (B) player tracks plotted on a 2D transformation of the court, (C) Player 1 heatmap after the game is finished, (D) Player 2 heatmap after the game is finished.

Summary

Overall, this type of information would be useful for players learning to play squash and interested in improving their game. Computer vision can be used to further enhance coaching by plotting a person's path to the ball or by assessing racquet preparation before each shot.