Section A-Research paper

Applications of Computer Vision for Analysing Player Performance in Tennis Sports

[1]Abhilash Manu, [2] Dr. Ganesh D., [3] Dr. Aravind H.S.

[1]Research Scholar, Jain University, Bangalore, Karnataka India
[2] Professor, IT Department, Jain University, Bangalore, Karnataka India
[3] Professor, Dept of Electronics and Communication Engg. Dept.,
JSS Academy of Technical Education, Uttarahalli, Bangalore, Karnataka

[1]manu.jois@gmail.com, [2]d.ganesh@jainuniversity.ac.in, [3]aravindhs1@gmail.com

Abstract: Traditionally subject matter experts' opinion has been the only way for a tennis player to analyse his/her performance. However, advancements and ease of access to complex algorithms have enhanced the usage of Artificial Intelligence and computer vision computing technologies in tennis. In this paper we focus on computer vision and image processing techniques used to analyse player performance in tennis. While there are so many possible methods for such a study it is quite essential for such studies to consider features at a player level such as expertise, age and sex. There are also many hurdles and challenges in achieving such a study since the video feed is heavily dependent on camera, ambient conditions, angle of view, field of depth etc. Multivariate thought process and methods have been applied mostly analysing the player strategy and Spatio-temporal positions. While some studies attempt to use simple cameras, some other use complex camera systems. MPEG4 is a popular format of files used by many. Different approaches have tried using different algorithms - MDL-MDS combination, custom object detection, pose estimation etc. The approach of this article is to provide a detailed introduction to applications based on computer vision in analysing tennis matches with player as the primary focus.

Keywords : Tennis player performance, Pose estimation, Homographies.

1 Introduction

Computer vision abased analytics for the game of tennis and player is comparatively a recent approach that is being used as a primary step for creating player and game analytics. In recent times the advancements in mobile camera technology have enabled easy video recording. such cameras have increased the volume of data being collected. Nowadays most tennis trainers and trainees use camera recordings on a daily basis for analysis. Such videos taken by the players and the trainers could be used to generate deeper analytics by developing computer vision algorithms.

Computer vision algorithms based on CNN architectures have been widely used for field, racket/ ball and player feature extraction. Multiple approaches have been taken by different researchers to generate the temporal context from a given video. Some of the most common 2D based approaches taken by many researchers is to connect convolutional networks and recurrent networks in repetitive blocks and vary the number of such blocks and fine tune the filter sizes in every block. Usual approaches are to use the a VGG16 proposed by Karen

Simonyan & Andrew Zisserman [1], and sequence it with RNN blocks to extract the temporal features.

While 3D analysis is performed using popular approaches based on 3D CNNs published by Paluri et al. [2], where they essentially use 3D CNNs also called 3D ConvNets to extract both special and temporal features. It is also documented to have outperformed 2D CNNs.

However, with Zhe Cao et al. publishing their work on pose estimation [3] using part affinity fields, researchers have tried to implement pose estimation models on tennis players to identify the player pose. Bases on this Kurose et al. [4] have improvised the model for analysing the player pose based on joint estimation. At this point it's important to mention the work done by Elliott et al. [5] who developed an accurate estimation for the true pose of the racket in the sample video of a tennis game by a validation scheme developed to compare pose estimates with data obtained using camera calibration software.

2 History of analytics in tennis

The first system for notational analysis of tennis was carried out by Downey (1973) as observed by Rafael [6]. This was a paper-based system used to make note of the strokes, the position of the player on court, the result of the stroke and effect of each stroke. this being an entirely manual and complex process, both to record information and to analyse, it was almost never used. however, it still instigated thoughts of the next generation tennis analysts.

A platform developed by Infosys was the first notable computer-based analytics/data recording system that was used in professional tennis in the year 1991. Complete statistics of the ATP season of the year 1991 was recorded and analysed. This provided many interesting insights of how the game was being played. Analysis of the data collected resulted in dashboards presenting game-level, player-level and opponent level information, which turned out to be invaluable in the times to come. This was followed by Hughes, M who published the approach to computerised notation of racket sports, in his book Science and racket sports in 1995 [7].

While there were many attempts to decipher the game using computer technologies, Tom Polk et al. published their work on tennis analytics. CourtTime is a comprehensive analysis of the game [8]. Jiang Wu et al [9] convert the game events to a sequence of activities to arrive at a frame work that allows users to set their own weights to the sequences and run a few Minimum description algorithm algorithm to detect sequential tactical patterns. The prediction accuracy of different types of tennis match forecasting models [10] comparing the physical demands and performance characteristics of professional tennis [11] and developing a comprehensive tennis shot taxonomy based on spatiotemporal data [12] are some other generic approaches studied so far.

Overall, the recent advancements in computer vision studies, focus on two major areas:

Match analytics - an area that focuses on generating match level analysis integrating spatial data with contextual data like score, service information and point outcomes. [8] [9]

Player analytics - an area that focuses on details of the individual player movements like speed of the racket, movement of the arm etc [4] [5] [13].

3 Review of Studies so far

In this review, it becomes inevitable to look at the 2 categories of studies - Match analytics and Player Analytics respectively in detail. The parameters that have been chosen to be compared are as follows:

a. Methodology - the approach
b. Data - the physical set up
c. Results – the outcome
d. Discussion - the observations

4. Match analytics

Recent studies on match analytics have made significant progress and are able to generate deep insights about the game enabling analysis of strategies. O'Donoghue [14] in his book defines sports strategy as – 'the plan that is set up prior to competition, to maximize the players´ strengths and reduce their weaknesses, while minimizing the opponents´ strengths and taking advantage of their weaknesses'.

CourtTime [8] Tom Polk ***et.al.*** performs a visual analysis of an entire match based on factual match data like player positions, play events like shots, serve types etc. and provides analytics like point analyser, shot analyser etc.

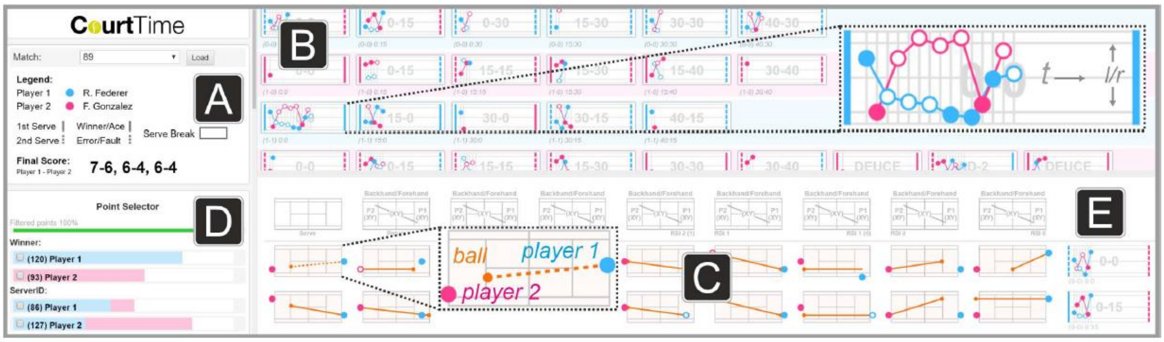

Fig 1. CourtTime [8] visual analytics system.

The tennis court is split into segments using Spatial data discretization process by dividing the continuous space into discrete, non-overlapping regions or cells. This has been done in order to simplify complex spatial data and make it more manageable for analysis. Distance-based dividing of the space into regions based on the distances between points. This can be useful for identifying clusters of points that are close together, but it can be sensitive to the choice of distance threshold. They use a single consumer-level camera, that captures enough data needed to reasonably represent the 2-D shots in a tennis match. Tennis court is divided into 6 sections on each side of the net and is used to represent the spatio-temporal positions. The videos have been manually annotated, taking about 3 hours per hour of video.

Since the focus of this study is more on tactics and strategy of the players and game the focus is on 'win points' that can be attained by making performing a good shot which the opponent misses or forcing the opponent to return a bad shot. Using the visual data, the match sequence is generated by retaining the hit-bounce events. Detailed features such as serve speed, point differential, short shots etc are generated. Parsing the raw data: This involves reading the raw data and extracting the relevant information, such as the players involved, the type of shot, and the location of the shot on the court. Cleaning and pre-processing the data: This involves removing any missing or erroneous data and ensuring that the data is in a format that can be easily analysed.

Aggregating the data: This involves grouping the data into shots and calculating relevant statistics, such as the number of shots, the success rate of shots, and the average shot speed. Adding contextual information: This involves adding additional information to the data, such as the match score and the number of games played, to provide context for the shots. Annotating the shots: This involves adding additional information such as the type of shot, the shot direction, and the shot outcome, to classify the shots in different categories and allow for more specific analysis. Once the data is transformed into shots, it can be used for various types of analysis, such as identifying patterns in shot selection, tracking player performance over time, and comparing the strategies of different players.

5. Methodology:

Data mirroring- Player location data is transformed
Transforming play events into shots - the raw spatial data is converted to hit-bounce event pairs referenced by a sequence number.
Spatial data discretization - nine left-to-right one-dimensional shot patterns and six depth-based one-dimensional shot patterns.
Feature generation - Wide serve, Serve speed, Return of serve stroke side, Point length, Point differential etc.
Visual Analytics application - Point selector, Point Analyzer, and Shot analyser.
Tennis shots are ordered and related ones clustered into sequences. This process of organizing and grouping tennis shots is based on their characteristics.

Some steps the study follows are:

Data set up:
Data from two tennis matches; one from a professional match broadcast on television and another from an amateur match using a single consumer-level camera, getting the minimum amount of data needed to reasonably represent the 2-D shots in a tennis match.

Experiment/Results/Discussion

- It is observed that an average tennis player would not be able to interpret the 1-D space time charts.
- Since only 2 matches were considered, this work will need more data for generalization.

Critical observations

1. Focuses on generating deep tactical information.
2. Spatio-temporal data capturing (3D) ball and player tracking system
3. Transforming the play events into the indexing scheme enables to reconstruct the entire play sequence.
4. Spatial data discretization is also a good approach.

Shot ordering : This involves arranging the shots in a specific order based on the match progression or other criteria, such as the shot type or the player who hit the shot.
Clustering: This involves grouping the shots into clusters based on their characteristics, such as the shot type, the shot location, or the shot outcome.

The distance between 2 events A and B in a pair is computed using the Gower distance (its simply the distances between observations). Metric Multidimensional Scaling calculates the distances between data points and generates the closest record compared to Gower metric.
This study focuses on generating tactical information and does not focus on player body position, player body movements etc.

Jiang Wu et al [9] study the analysis of multivariate event sequence data in tennis following an approach of generating a sequence leading to the win points. In this approach the sequence of hits is considered as a multivariate event set. Such multivariate events and the sub-sequences embedded in the larger sets are analysed. Each set of event sequence is represented by S = (e1, e2, e3, ..., en) in which each event is denoted as ei = {a1 = vi1, a2 = vi2, ..., an = vin}. Once the sequencing is generated, user can set his own weights to the events and run MDL algorithm (minimum description length –

A pattern mining algorithm) to detect sequential tactical patterns. The basic idea behind MDL is to encode the data using a set of patterns (or features) that are chosen to minimize the total length of the encoded data. To do this, MDL starts by assuming that the data is random and assigns a high description length to it. Then, it iteratively adds patterns to the description and updates the length until the description length reaches a minimum. MDL can be applied to various types of data such as time series, sequences and images. One of the applications of MDL is to find regularities in sequences such as sequences of letters or numbers.

Methodology

Tennis match is modelled as a sequence of multivariate events. Each sequence S has a list of multiple events that is represented as S = (e1, e2, e3,...,en) and where each event ei = {a1 = vi1,a2 = vi2,...,an = vin}. Allows users to set their own weights to the sequences, run a few MDL algorithm to detect sequential tactical patterns.

Data Setup: Tennis datasets have been procured by two different teams of tennis experts. Interns have been used to record the proceeding of the match manually. The framework consists of a database based on MongoDB, a backend based on Flask, and a web application based on React.

Experiments/Results/Discussion : The outcome is a software tool that can analyse the on-ground sequence of play and arrive at the strategy used, strength of the player etc.

Critical Observations:

- Analyses multivariate event sequences.
- The sequence generation and pattern mining could be useful.
- Ignores important information like the position of the ball landing, shot selection, player position etc

This method does not deploy computer vision algorithms rather it depends on domain experts, hence manual and not scalable. The framework runs on a simple MongoDb, React, and flask application that runs MDS algorithm. The basic idea behind MDS is to find a set of low-dimensional coordinates for the data points such that the pairwise distances between the points in the low-dimensional space are as close as possible to the pairwise distances between the points in the high-dimensional space. This is done by minimizing a cost function that measures the difference between the pairwise distances in the high-dimensional space and the pairwise distances in the low-dimensional space.

It is important to note that MDS is sensitive to outliers, missing data and high-dimensional noise, so the data is pre-processed before using MDS. Another interesting study to use the existing closed-circuit cameras (CCTV cameras - making some minor positional adjustments) has been proposed in a study by **Rafael Martín et al. [15]**. The approach here is to use feeds from multiple CCTV cameras positioned at vantage to cover the field to generate statistics like player position and tennis match analytics. The video input devices which are simple CCTV cameras placed at nine different locations on either side of the court. Videos were ASF files with MPEG-4 codec formats were used for the processing.

Methodology

- The Tracking block receives the videos from 2 sets of cameras placed on each side of the field and masks noisy data.
- The Homographies are used to change the perspectives as per the reference points, generates the top view.
- The Fusion block joins the processed tracking's from each camera and generates the overall trajectory along the entire field.

Data set up: The approach here is to use feeds from multiple CCTV cameras positioned at vantage to cover the field to generate statistics like player position and tennis match analytics. Videos are in ASF files encoded in MPEG-4. Seven cameras used a microphone and a resolution capacity of 640x480. 2 other cameras have a resolution of 704x576 pixels Videos are synchronized at the start of each sequence.

Experiment/Results/Discussion: An over lapping effect occurs between the frames due to different perspectives generated by different lenses, precision in homographies, video tracking errors, etc.

This effect is further magnified while using this data for processing using statistics. However, the system is able to detect and track the player in each side of the field.

Critical observations:

It captures videos from CCTVs, merges them to create a single vision. Player tracking is done using the highlights in the video frame. Based entirely on video filters and video fusion. The entire analysis needs to be done synchronously - camera operations, tracking and fusion. Hence creating some level of complexity.

The set of mono cameras feed live videos to a 'foreground segmentation module' that masks stationary objects and noisy background. The next set of processing is to change the camera perspectives (angle of observation) by using 'homographies', normalized liner transformation algorithms are used for this purpose. These videos are the fused to form a single view the field.

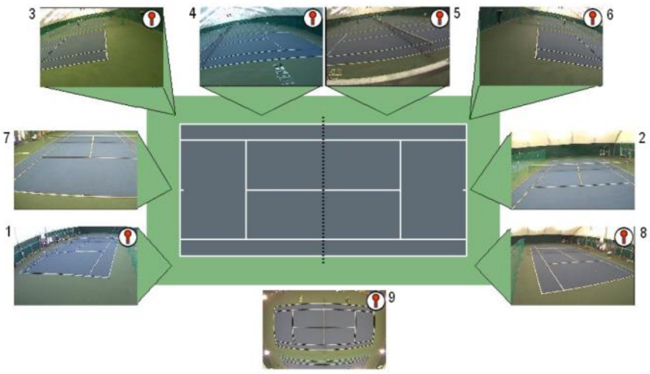

Fig 2. Rafael Martín et al. [15] - multi-camera tennis videos

Another study that analyses the return strokes and services of singles tennis games was conducted by **Ryota Mukai et al. [16]** the approach was to use 2 hi-definition cameras (one on each side) to generate a 2-dimentional player position and a 3-dimentional ball position. The ball position is identified by subtraction method applied on cameras from either side on a set of 3 images namely Ti-1 (pervious), Ti (current), and T(i+1) (next). This is necessary to eliminate the shadow of the ball being confused with the ball itself. Player extraction is simple since a direct subtraction of images from 2 cameras yield the required output.

However, the player position is identified by the direction of feet and the distance between the feet. Though this study was path breaking when it was published, developments in algorithms like pose estimation and object detection provide easier and more efficient option to track and generate match analytics. Methodology: The ball position in the right and left images are detected by the subtraction method by using three successive images to eliminate misdetections caused by ball shadows on the tennis court. The three images are previous, current and following frame images.Players can be extracted using the background subtraction method. The background image has no players. The player position is defined as the middle point between the toes of the player's shoes.

Data Setup: Two high-definition TV cameras (right and left) are used to detect 3-D ball position and 2-D player position.

Experiments/Results/Discussion: The strokes in the tennis games were analysed. The 3-D trajectories of ball in a game and the relative positions of players were detected by TV cameras. The six skill parameters were defined as skill factors. It was found that ball speed and ball scattering are strongly related to return stroke skill, and that the skill factors differ between players.

Critical Observations: This is an application of object detection in videos using computer vision. Something that was path breaking in 2011, when the paper was published. Not so important today, as most of this is assumed to be std these days.

Player Analytics

A tennis player performance depends on the capabilities such as agility, body balance, reaction speed, situational awareness and the ability to implement the strategies [17]. And an efficient study on these aspects of the player necessitates the requirement of a powerful tool that can provide required levels of detailing on player pose, movement, positions etc. The study published by M. Paluri et al. [2] and Zhe Cao et al. [3] on pose estimation can be considered as path breaking since researchers were now able to generate human pose estimates using a simple RGB snapshot or video.

Inspired by these pose estimation algorithms, **Ryunosuke Kurose et al [4]** worked upon pose estimation of tennis players. The method proposed in the paper is based on the use of a deep learning-based pose estimation algorithm to detect and track the body keypoints of players in the video. The authors use the popular OpenPose algorithm, which is a multi-person version of the single-person COCO (Common Objects in Context) architecture, that can detect body keypoints, such as the head, shoulders, elbows, wrists, hips, knees, and ankles, in real-time.

Ryunosuke Kurose et al [4] apply player pose estimation method using PAFs (Part Affinity Fields) as a process to derive analytics on pose of the tennis player. The pose estimation algorithm creates feature vectors with 'joint position coordinates' derived from the posture estimation [3]. The authors evaluate the performance of their method on a dataset of tennis videos, which contains a total of 1,280 frames and was collected from YouTube. The dataset contains footage of both professional and amateur players, and the players are shown in various playing positions, such as serving, volleying, and groundstrokes. They report that the method was able to accurately detect and track player poses, with an average keypoint detection rate of 94.5% across all frames.

Methodology:

This approach estimates the pose of the player using Part Affinity Fields. 17 points are chosen to be identified like eye, shoulder ear, nose, elbow, hip, knee, wrist, ankle, neck. Then the joint feature vectors are clustered for each frame by GMM. Predict the shot success probability using SVM.

Data Set up: Video - the match video is 1920 X 1080 and the frame rate is 30 fps. There are images in which multiple players are performing serve, forehand, backhand, and standby state in this dataset.
When classifying the joint position coordinates by GMM, they have decided experimentally the number of classes as 20 classes.

Experiments/ Results/Discussion:

Posture characteristics of shots and the specific posture have been identified. The probability of failure in the forehand and backhand is high beyond the lower 13% and higher 8% respectively. Error rate of forehand is high, there are about two posture classes that appear with high probability in every shot.

Critical observations:
- In Pose estimation: the pelvic position is assumed to be the centre point of the hip positions
- assumes that the height of all players is the same.
- In Feature vector extraction, they don't distinguish shot positions in the front and back direction.

To eliminate the spatial dependencies on the co-ordinates of the tennis player, the co-ordinates of the joints identified uses relative coordinates of the pelvis as the origin. The created feature vectors are clustered into 20 clusters frame by frame using Gaussian mixture model (GMM) generating a bag of words on postural data set. Using support vector machine algorithm, the possible result of the shot is predicted. Video feed of a tennis match with 1920 X 1080 pixels and the frame rate is 30 fps is taken. This implies that the number of frames generated was about 44000 frames.

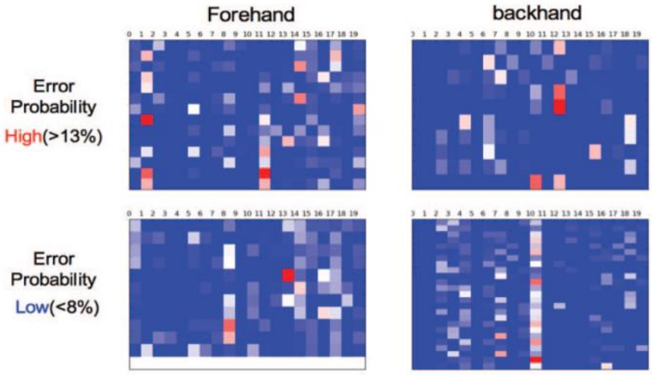

Fig 3 Ryunosuke Kurose et al [4]. Prediction results

As a result, the tennis player pose is identified accurately with results matching Zhe Cao's work. Also, this study is able to compare the failure rates between the fore arm and back arm shots as shown in Fig 3. The authors also demonstrate the applicability of their method by analysing the poses of players in the dataset and comparing them to the ideal poses recommended by professional coaches. They found that many players had poor posture and technique, such as a lack of knee flexion during the serve, and that their method could provide detailed information about these issues to coaches and players.

In conclusion, the paper presents a new approach for analysing the poses of tennis players in video footage, which can provide detailed information about player posture, movement, and technique. The authors have used a deep learning-based approach that can detect and track the body key points of players in real-time, and evaluate their method on a dataset of tennis videos. They show that the method is able to accurately detect and track player poses and it can be used to identify and rectify poor posture and technique.

Rajdeep Chatterjee et al. [18] propose the use of a framework called 'Detectron 2' for tennis pose estimation. The goal of the research is to develop a robust and accurate system for recognizing different sports activities using only pose information, which can be captured using a single camera or depth sensor. The method proposed in the paper is based on a combination of deep learning and traditional computer vision techniques. The authors first use a pre-trained deep learning model, OpenPose, to extract human body keypoints from the input images or videos. They then use these keypoints as features to train a support vector machine (SVM) classifier for each sport activity. Essentially, it's a ResNet 50 based model as shown in fig 4.

This model is trained on masked or classified feature set consisting of 50 features out of which 14 are newly created for this study (that is, 12 key points ×3 = 36 most important features+14 newly created features). To predict the action and movement of the athlete, they have classified the player poses into 3 classes namely forehand motion, backhand motion, and the reset position or the base position of the player. Cameras are positioned at court level to maintain uniformity in camera angles and avoid the need for homographies to correct perspective errors. A total of 3K images have been collected from YouTube as an input data set.

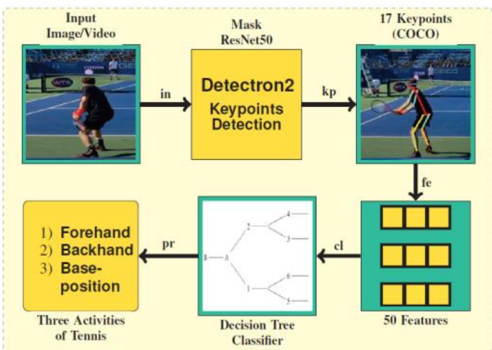

Fig 4. Rajdeep Chatterjee et al. [18] Detectron2

The study compares the results with many popular convolution neural network based algorithms like AlexNet, VGG16, ResNet50, MobileNetV2, and EfficientNetb71 and stands out by achieving better results of about 98.60% accuracy. The authors evaluate their method on a dataset of sports activities, which contains videos of different sports activities, such as basketball, soccer, and volleyball, and the dataset was collected from the internet. They report that their method was able to achieve high accuracy in classifying the different sports activities, with an average accuracy of 97% across all activities. The authors also compare the performance of their method to other existing methods for sports activity classification, such as using only RGB images or using both RGB and depth images, and showed that their method performed better. They also demonstrate the robustness of their method by testing it on different types of cameras and lighting conditions and showed that it still achieved good results.

Methodology:

- Detectron
- model pose and key-point identification.
- Random forest for classification.

Data Set up:

- Google Colab used as online coding editor.
- 12GB of RAM. Pytorch 1.7.0+CUDA 10.1 framework and Python 3.7 have been used for coding.
- The Detectron2 model for key-point detection is Mask RCNN R 50 FPN 3x.

Experiments/ Results/Discussion:

Higher accuracy of 98.60% achieved in comparison with: AlexNet, VGG16, ResNet50, MobileNetV2, and EfficientNetb71.

Critical observations: Cameras are placed at a nominal height of about 5fts player to maintain uniformity. The tennis action motion is classified into the following subcategories.

1) Initialization
2) Racket back
3) Swing stage
4) Contact point
5) Follow through

In conclusion, the paper presents a new approach for classifying sports activities based on human pose information. The authors have used a combination of deep learning and traditional computer vision techniques to extract human body keypoints from the input images or videos and used these keypoints as features to train a classifier for each sport activity. They evaluate their method on a dataset of sports activities and show that it can achieve high accuracy in classifying different sports activities. This approach can be applied to other domains, such as action recognition, human-computer interaction, and surveillance systems.

While studies focus on generating player level analytics using a simple mobile / handheld / ubiquitous camera using pose estimation, focussed research with lab set up using high end cameras and 3D analytics have also been done. One of the most recent works on tennis was published in 2021 by **Maria et al. [19].** In this study sophisticated motion capture systems are used in a lab environment to study the player movement during forehand and backhand.

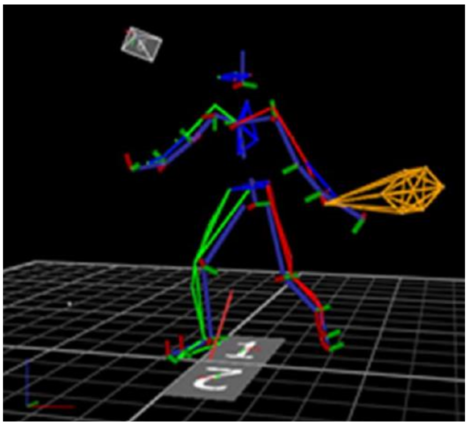

Fig 5. Maria et al. [19]. 3D model with racket.

Fuzzy C-Means (FCM) clustering aims to partition a set of n data points into c clusters, where each data point belongs to every cluster to a certain degree. This degree of membership is represented by a membership matrix, where each element represents the degree of membership of a data point in a specific cluster. The dynamic time warping (DTW)-based distance is a method used to calculate the similarity between two time series. It is particularly useful when the time series have different lengths and when there is a non-linear alignment between the time series. In the context of clustering tennis shots, DTW can be used to measure the similarity between different shots, allowing for the clustering of similar shots.con

The data from the markers on the tennis racket positions and participant's body are used to build the 3D model of a tennis player holding a tennis racket. Subsequently two, three, and four clusters are formed and Fuzzy Spatial-Temporal Graph Convolutional Neural network classifier is applied. Then the frames of each of the following tennis moves are sequenced. Thus, obtaining trajectories of the body parts and racket which were analysed using fuzzy clustering - Fuzzy C-Means (FCM) clustering to group of forty tennis shots. When using the FCM algorithm with DTW-based distance for clustering analysis of tennis shots, the algorithm first assigns each shot a degree of membership in each cluster based on the DTW distance between the shot and the cluster's centroid. The centroids are then recalculated based on the membership matrix and the process is repeated until the membership matrix stops changing or a stopping criterion is reached.

Methodology:

- The movements of the tennis player was identified with the help of the optical motion capture system.
- The data points thus obtained are re-processed to arrive at 3D data for both the player and the ball.
- In the entire game data continuous shots are paired. The trajectories formed by the ball were analysed using fuzzy clustering - Fuzzy C-Means (FCM) clustering

Data Set up:

- High end camera systems from Vicon was used in this study.
- 'retroreflective markers' were attached to tennis player's body. About 40 labelled as forehand and back hand served as the data set for analysis.

Experiments/ Results/Discussion:

- Clusters were classified -Forehand and Backhand (with ball and with-out ball) 2.
- However, when the clusters were compared with a ball hit and without a ball hit, the clusters failed to group them correctly.

Critical observations:

- The study was performed in the 'Laboratory of Motion Analysis and Interface Ergonomics at the Lublin University of Technology'.
- A software called NEXUS was used for post-processing.
- Data was exported to c3d format.

The resulting clusters represent groups of similar tennis shots, and the degree of membership of each shot in each cluster can be used to infer the level of similarity between shots and the clusters. This can provide valuable insights into the different types of shots used in tennis, such as identifying patterns in shot selection and technique, and can aid in understanding the strategies used by players.

A much simpler approach is taken by **Jhen-Min et al. [20]** in their study on tennis player pose classification using the pre-trained Yolo and MLP. it uses a single convolutional neural network (CNN) to predict bounding boxes and class probabilities directly from full images in one evaluation. Multi-layer perceptron (MLP) is a feedforward neural network architecture composed of multiple layers of perceptrons, which are simple neural network models. The study describes a method to classify the pose of tennis players using these two technologies using YOLO to detect the player in an image and then using an MLP to classify the pose of the player.

It is also using YOLO to detect key points on the player's body and then using an MLP to classify the pose based on those key points. A general television broadcast video of 2 tennis matches with resolution is 1280 × 720 and 26 Fps are used. Using OpenCV, the clip is resampled. YoloV5 pre-trained models are used to identify the player and the ball. A MLP layer with TensorFlow 2.0 and tf.keras - 117 input layers, 3 hidden layers with 50 nodes each, ReLU activation has been used to identify the pose. However, the model is only able to identify the players and the ball. Details of tennis player analytics - tennis specific poses etc. have not been mentioned in the published study.

Methodology:

- A video of 1280 × 720 resolution is used. Its later resampled and edited using OpenCV.
  For detecting multiple players, they use YOLOv5. With a precision avg of 0.977 after training.
  NN based MLP models are trained using existing data sets with TensorFlow 2.0 and tf.keras.
- Data Set up: TV telecast video of a tennis match was used where the video clip resolution is 1280 × 720. The frame rate is 26fps. Then, the clip images are resampled and cropped with OpenCV. The resolution and frame rate of the half-court cropped clips are 1200 × 450 and 26fps

Experiments/ Results/Discussion

- The outcome of this study is that the model is able to identify the location of a player and a ball as MLP inputs.
- Identification of a player is only a primary solution that have already been achieved quite easily by multiple approaches and studies earlier.

Critical observations

The study was based on sample clips from the broadcast video of Jason Jung vs. Federico Coria and Marin Cilic vs. Rafael Nadal games are from the US Open Tennis Championships 2020 and 2019. The model is able to identify the players and the ball. Details of tennis player analytics - tennis specific poses etc. have not been mentioned in the paper.

5 Results and discussions

Tennis match analysis involves evaluating various aspects of a player's performance during a match, such as their shot selection, strategy, and ball placement etc. Most studies achieve this by watching the footage of the match and taking notes, or by using data and statistics from the match, such as the number of aces, unforced errors, and winners. The analysis can help identify areas of improvement for a player and provide insight into why a player may have won or lost a match. It is often used by coaches, players, and commentators to gain a deeper understanding of the game and to make predictions about future matches.

Tennis player analysis involves evaluating various aspects of a player's game, such as their strengths and weaknesses, playing style, and overall performance. This can be done by watching footage of the player's matches, studying statistics and data, and assessing the player's physical and mental attributes. Some of the key areas that analysts focus on when evaluating players include:

- Shot selection: what types of shots a player hits, how often they hit them, and how successful they are with those shots.
- Movement and footwork: how well a player moves around the court, and how effectively they are able to reach balls and set up for shots.
- Serve: how well a player serves, including the speed and accuracy of their serves, and how often they are able to win points on their serve.
- Return of serve: how well a player receives and returns serves, and how often they are able to break their opponent's serve.
- Mental toughness: how well a player handles pressure, maintains focus, and manages emotions during matches.

This analysis can be helpful for coaches, players, and commentators to understand the players better and to predict the outcome of future matches.

	Results of the observations			
	Study	Methodology	Algorithm	Outcome
Match Analysis	Courtime, Polk *et.al*	Spatial data discretization	MVS algorithm	1-D Space-Time Charts
Match Analysis	Multivariate event sequence, Jiang Wu et al.	Multivaria sequence clustering	MDL and MDS algorithms	Tennis Event Sequence Mining
Match Analysis	Multi camera approach Rafeal et al.	Homographies and fusion	Video Sync algorithm	Player tracking
Match Analysis	Evaluation of Tennis, Royota et al.	Object identification and tracking	Object detection	Tennis ball detection
Player Analysis	Player pose analysis, Ryunosuke et al.	17 pose point identification	Pose estimation and SVM	Player pose identification - Error < 15%
Player Analysis	Pose based activity classification, Rajdeep Chatterjee et al.	Pose identification and classification	Detectron-2	Player pose identification - 98% accuracy
Player Analysis	Player pose classification, Jhen-Min Hung et al.	Pose and shot identification	OpenCV and YoloV5	Player shot identification - 0.97 Precision
Player Analysis	Multivariate Time Series Clustering, Maria et al.	Marker based movement identification	Fuzzy C-Means (FCM) clustering	Forehand backhand identification - 75%+ Mean SD

Table 1 – Results and Observations

5 Conclusion

The studies so far have been able to provide an estimation of various player poses and are able to classify the stance. However, based on the current studies we cannot conclude on any methodology for performance analysis of a player either at a given point in time or over a time period. In video analytics, the studies so far have been limited to what can be referred to as the "pixel presence". Analysis is based on the spatial pixel identification with respect to a primary axis. That is, the position of the player on the field or the position of the ball in time or the position of the player body etc. They do not study Player stance, Player movement and Time series analysis. Marker based studies provide a detailed 3D analysis. The studies are extremely focussed on the body part and the analytics generated. These studies are True only in lab conditions, need 3D cameras and elaborate markers (sensors) and these systems are expensive and hence out of reach of the common man.

References

[1] Karen Simonyan_ & Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. Published as a conference paper at ICLR 2015. arXiv:1409.1556v6 [cs.CV] 10 Apr 2015

[2] M. Paluri, D. Tran, L. Bourdev, R. Fergus, and L. Torresani, "Learning spatiotemporal features with 3d convolutional networks", Computer Vision and Pattern Recognition, 2015. arXiv:1412.0767v4 [cs.CV].

[3] Zhe Cao, Tomas Simon, Shih-En Wei, Yaser Sheikh. Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 7291-7299

[4] Ryunosuke Kurose, Masaki Hayashi, Takeo Ishii "Player Pose Analysis in Tennis Video based on Pose Estimation", International Workshop on Advanced Image Technology (IWAIT) 2018

[5] N. Elliott, S. Choppin, S. R. Goodwill, T. Allen. "Markerless tracking of tennis racket motion using a camera," Procedia Engineering, Volume 72, 2014, pp. 344-3

[6] Rafael Martínez, ITF Coaching and Sport Science Review 2018; 74 (26) https://www.researchgate.net/publication/324601508

[7] Hughes, M Computerised notation of racket sports. In: Reilly, T, Hughes, M, Lees, A (eds). Science and racket sports, London: E & FN Spon, 1995, pp. 249–256.

[8] Tom Polk, Dominik Jackle, Johannes H ¨ außler, and Jing Yang "CourtTime: Generating Actionable Insights into Tennis Matches Using Visual Analytics", IEEE Transactions on visualization and computer graphics, vol. 26, no. 1, January 2020.

[9] Jiang Wu* Ziyang Guo* Zuobin Wang* Qingyang Xu* Yingcai Wu "Visual Analytics of Multivariate Event Sequence Datain Racquet Sports", 2020 IEEE Conference on Visual Analytics Science and Technology (VAST)

[10] Kovalchik, SA . Searching for the GOAT of tennis win prediction. J Quant Anal Sports 2016; 12: 127–138.

[11] Kovalchik, SA, Reid, M. Comparing matchplay characteristics and physical demands of junior and professional tennis athletes in the era of big data. J Sport Sci Med 2017; 16: 489–497.

[12] Kovalchik, SA, Reid, M. A shot taxonomy in the era of tracking data in professional tennis. J Sport Sci 2018; 36: 2096–2104.
[13] Rajdeep Chatterjee, Soham Roy, SK Hafizul Islam, Debabrata Samanta, "An AI Approach to Pose-based Sports Activity Classification" 2021 8th International Conference on Signal Processing and Integrated Networks (SPIN) | 978-1-6654-3564-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/SPIN52536.2021.9565996.
[14] O'Donoghue, Research Methods for Sports Performance Analysis, ISBN 9780415496230, Published December 14, 2009 by Routledge
[15] Rafael Martín Nieto, José María Martínez Sánchez, "An automatic system for sports analytics in multi-camera tennis videos", Workshop on Activity Monitoring by Multiple Distributed Sensing (AMMDS) in conjunction with 2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance.
[16] Ryota Mukai, Toshio Asano and Hajime Hara, "Analysis and Evaluation of Tennis Plays by Computer Vision", Proceedings of the 2011 IEEE International Conference on Mechatronics and Automation August 7 - 10, Beijing, China
[17] Satoshi Ochi and Mary Jo Campbell. The Progressive Physical Development of a High-Performance Tennis Player Strength and Conditioning Journal 2009. www.nsca-lift.org. pages 59-68.
[18] Rajdeep Chatterjee, Soham Roy, SK Hafizul Islam and Debabrata Samanta. An AI Approach to Pose-based Sports Activity Classification. 2021 8th International Conference on Signal Processing and Integrated Networks (SPIN) | 978-1-6654-3564-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/SPIN52536.2021.9565996
[19] Maria Skublewska-Paszkowska, Paweł Karczmarek, Edyta Lukasik, "Tennis Multivariate Time Series Clustering", 2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) | 978-1-6654-4407-1/21/$31.00 ©2021 IEEE | DOI: 0.1109/FUZZ45933.2021.9494420
[20] Jhen-Min Hung, Jinn-Yen Chiang and Kerwin Wang. Tennis Player Pose Classification using YOLO and MLP Neural Networks, International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS) | 978-1-6654-1951-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISPACS51563.2021.9650925
[21] Blobel T, Pfab F, Wanner P, et al. Healthy Reference Patterns (HRP) supporting prevention and rehabilitation process in professional football. In: World conference on science and soccer, Reneese: Universite Reneese, France, 31 May–2 June 2017, pp.183–184.
[22] Hughes, M Computerised notation of racket sports. In: Reilly, T, Hughes, M, Lees, A (eds). Science and racket sports, London: E & FN Spon, 1995, pp. 249–256.
[23] Unierzyski, P, Wieczorek, A Comparison of tactical solutions and game patterns in the finals of two grand slam tournaments in tennis. In: Lees, A, Kahn, J-F, Maynard, I (eds). Science and racket sports III, Oxon: Routledge, 2004, pp. 169–174.
[24] Kovalchik, SA, Reid, M. A shot taxonomy in the era of tracking data in professional tennis. J Sport Sci 2018; 36: 2096–2104.
[25] Kovalchik, SA . Searching for the GOAT of tennis win prediction. J Quant Anal Sports 2016; 12: 127–138.
[26] Kovalchik, SA, Reid, M. Comparing matchplay characteristics and physical demands of junior and professional tennis athletes in the era of big data. J Sport Sci Med 2017; 16: 489–497.

[27] Hall SJ. What Is Biomechanics. In: Hall SJ. eds. Basic Biomechanics, 8e New York, NY: McGraw-Hill; 2019..
[28] Brukner P. Brukner and Khan's Clinical Sports Medicine. North Ryde: McGraw-Hill; 2012.
[29] The British Association of Sport and Exercise Sciences. More About Biomechanics. http://www.bases.org.uk/Biomechanics. 2 May 2016.
[30] Biomechanics and tennis - B Elliott. https://bjsm.bmj.com/content/40/5/392
[31] The biomechanics of tennis elbow. An integrated approach. Roetert EP1, Brody H, Dillman CJ, Groppel JL, Schultheis JM 2020
[32] Biomechanics of the Tennis Groundstrokes: Implications for Strength Training Roetert, E Paul PhD1; Kovacs, Mark PhD, CSCS1; Knudson, Duane PhD2; Groppel, Jack L PhD3 2021
[33] Influence of the forehand stance on knee biomechanics: Implications for potential injury risks in tennis players Caroline Martin,Anthony Sorel,Pierre Touzard,Benoit Bideau,Ronan Gaborit,Hugo DeGroot &Richard Kulpa.2021.
[34] The accuracy of several pose estimation methods for 3D joint centre localisation. Laurie Needham, Murray Evans, Darren P. Cosker, Logan Wade, Polly M. McGuigan, James L. Bilzon & Steffi L. Colyer Oct 2021.