BEYOND THE COACH: EXPLORING THE EFFICACY OF A MACHINE LEARNING APPLICATION FOR IMPROVING TENNIS PLAYERS' PERFORMANCE

Jiawen Hao[1], Huijun Hu[2]

[1]Rye High School, 1 Parsons Street Hudson Valley Rye, Westchester, New York 10580
[2]Computer Science Department, California State Polytechnic University, Pomona, CA 91768

ABSTRACT

At some point in their lives, most tennis players are likely to encounter the dilemma in which they can't constantly receive guidance from a coach. Are there any alternatives that can provide players with the means to improve their game?

For a tennis player, it is difficult to always rely on a coach for tips and suggestions [4]. Mundane constraints often prevent players from consistently improving their techniques and strokes through the guidance of a professional figure [5]. In this paper, we discuss the prospect of using an application to act in the stead of a coach to help aspiring tennis players improve. Using machine learning, the application analyzes and compares two videos of corresponding strokes inputted by users [6]. The AI aligns the frames by clustering the motions of the strokes and then outputs appropriate tips according to the results [7]. This application will allow tennis players to work on and improve their performance on occasions where their coaches aren't available, improving both efficiency and consistency.

KEYWORDS

Tennis players, Coach alternatives, Machine learning

1. INTRODUCTION

Suited for all ages, tennis is a popular sport enjoyed by many throughout the world [8]. Amongst those who enjoy the sport, there are many who strive toward becoming a professional player, as well as many others who play the sport for the sake of enjoying it. Yet despite the level of commitment, most players at some point in their lives had sought professional help to improve their performance in this sport. A lesson taught by a professional often averages around $60, and clients often have to adhere to the professional's schedule. Due to the nature of these lessons, complications and inconveniences frequently emerge. And while players can alternatively practice with a friend or partner at a local club or park, they lack the means to discern their own mistakes and improve upon them. As a result, many players who seek to make a breakthrough are often perplexed by this dilemma.

David C. Wyld et al. (Eds): SIPM, ITCA, CMIT, FCST, CoNeCo, SAIM, ICITE, ACSIT, SNLP - 2023
pp. 65-71, 2023. CS & IT - CSCP 2023 DOI: 10.5121/csit.2023.130907

One of the approaches to improving the skills of aspiring players is tennis academies. These academies are dedicated to raising and molding kids and teenagers into competitive tennis players. Alumni spend around 6 hours everyday improving their fitness and practicing tennis [9]. Without a doubt, players are able to improve their performance drastically in such an environment. But one evident problem this method presents is the cost of the enrollment. An individual has to pay up to $40,000 per year to train at tennis academies [10]. And more often than not, players have to sacrifice proper education to train in these academies. Additionally, another method was proposed by Head, a company that develops and sells tennis gears and apparel. The company developed an attachable sensor that can detect the strokes of the players. Using the app they created alongside the product, players are able to review their game statistics, the placement of the balls, and the velocity of their shots. The downside of this method is the cost of the product. The sensor costs around $190, and most casual players may find the product to be very expensive. Our proposed application utilizes machine learning to analyze inputted videos [11]. Users can compare their strokes and those of their coaches or professional players and visualize what they need to improve on based on the videos. The idea behind this application was inspired by a myriad of other trainer apps on the App Store and Playstore [12]. Although the application cannot fully serve as a tennis coach or professional, it can act as a very helpful alternative during their absence. It is not very costly, as compared to many of the methods I stated before. It is efficient in that the App can be applied in practically all circumstances, be it a practice at a local park, or a private lesson with a coach.

To prove our results, corresponding sets of different strokes were inputted into the program, as well as videos of different lengths. Using this method, we were able to find out to what extent our program can process before it fails to function. First of all, the videos have to be taken in similar orientations and angles, this ensures that the AI can pick up on the landmarks properly. Second, the videos should only contain one stroke and cannot exceed 1 minute or so for optimization.

The rest of the paper is organized in the following way: Section 2 describes the challenges that we met during when we were experimenting and designing the sample; Section 3 details the solutions of the problems we met during section 2; Section 4 further elaborates on the project by giving additional details. Section 5 presents related work to our project. And lastly, Section 6 concludes the paper and gives details regarding future works.

2. CHALLENGES

In order to build the project, a few challenges have been identified as follows.

2.1. To Process and Compare the Two Videos

To process and compare the two videos inputted by the users, specific libraries have to be utilized to achieve that end. Some examples are mediapipe, cv2, and etc [13]. These libraries, however, are exclusive to Python and cannot be imported to Flutter, the language we are using to construct the UI. As such, the videos inputted through Flutter must be uploaded to a server with the proper libraries. This way, the videos can be properly analyzed and compared. In essence, two videos inputted by the user through the Flutter application are uploaded to a platform containing all the necessary libraries, such as a server, to be processed and compared. The results are then sent back to the Flutter app from the server and displayed to the user.

2.2. Different Length of Videos

The videos inputted through the Application may differ in lengths. Some may be only a few seconds, while some might be half a minute or so, but none should be able to exceed 1 minute for the sake of consistency and performance. The frames containing the tennis strokes may not perfectly align and correspond to each other according to the time elapsed in these videos. To calibrate these differences and align the frames of the videos, machine learning must be deployed to achieve this purpose. Machine learning can analyze the motions of the strokes and assign them into categories. Through this means, the videos inputted can be aligned regardless of their lengths.

2.3. The Output of the Data

After the videos are inputted, the output of the data processing needs to be able to help players improve their strokes. However, it was unsure what kind of output would most effectively help the players realize what they need to improve on. Since each player possesses their own style, it is unnecessary to output a rating based on the similarity between the coach stroke and the student stroke. It is generally inefficient to output tips and grades according to the similarity, as some users might simply use two different videos of him or her practicing the same stroke to observe the difference in details. Therefore, we determined that outputting corresponding frames of the two videos side by side would most effectively let the players realize how they could improve.

3. SOLUTION

The aforementioned application serves the purpose of analyzing and comparing tennis videos with the intent of improving the performance of tennis players. There are three key components that make up the application: A frontend user interface, and a backend server for machine learning. The frontend user interface is written in Dart in Flutter, and is capable of running on both Android and IOS devices. Its flexibility saves us from the hassle of creating multiple versions of the app in order to upload it on both the IOS and Android devices. The backend server written in Python contains the codes necessary for machine learning. The server connected to the frontend UI would analyze the videos inputted by the user and make comparisons between them. To train the AI, we assigned numbers to the landmarks captured in the videos and utilized K-means clustering to align them. From there, the frames are printed side to side to let the user compare and make improvements to their game.

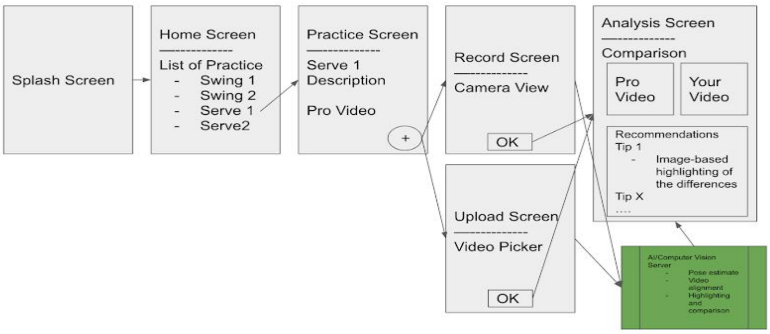

Figure 1. Overview of the solution

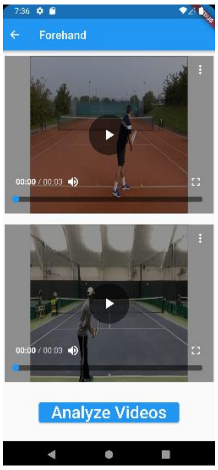

Figure 2. Video page

To implement the frontend UI, we utilized Flutter, a mobile development framework that employs Dart as the language and is eligible for both IOS and Android [15]. We imported plugins such as video_play, video_item, and image_picker. These libraries allow us to make code that requests and processes the videos from the users. To put it into context, as seen in Figure 1 above, the videos displayed in the application are first obtained through the getVideo function. The videos are stored and subsequently prepared to be displayed in containers on the appropriate screen. After the videos are stored and properly displayed in the application, users can press “Analyze Videos” which would send the videos stored to the backend server.

The backend server is coded using Python, a programming language commonly used for data analysis. We imported Mediapipe and OpenCV, two libraries necessary for the tracking of motions displayed in the videos, to use their functions and record the landmarks of the subjects in the videos inputted. The data recorded from identifying the landmarks of the players can then be processed through Machine Learning. From there on, we imported K-means from scikit-learn, a library containing various methods for machine learning, in order to perform clustering. By using machine learning, we can align the two videos inputted and print out keyframes on the application.

4. Experiment

4.1. Experiment 1

In our first experiment, we sought to determine whether the angle that the videos are taken from impacts the accuracy of the output. We have only tested the application with videos that are taken from the back of the players. While the results were mostly accurate, occasionally, wrong frames would end up aligning with each other. Therefore, by determining if angles affect the accuracy of the output, we can notify the users of the most optimal angle to take the videos from. After the experiment was conducted, the results showed that the accuracy of the output is largely the same regardless of the angle the videos are taken. The videos taken from the front and back yielded slightly more accurate output, but the improvement in the accuracy from changing angles is largely negligible.

4.2. Experiment 2

In our second experiment, we sought to find whether the presence of additional players in the background of the videos taken hinders the accuracy of the output. A sizable portion of the users we wish to attract with this application are recreational players. More often than not, these recreational players would practice tennis in local parks due to their convenient nature. When videos are taken in such settings, it is inevitable that players in the background would be caught in the videos. We want to find out if videos taken with players in the background would yield less accurate results than ones taken without players in the background. After the second experiment, we found that the landmarks of the individuals in the background would not be recognized if the camera is focused on the main subject of the video. This means that having other individuals in the background of the videos would only minimally impact the accuracy of the output.

When it comes to video taking, there are many extraneous variables that can significantly impact the accuracy of the program. The camera angle and moving figures in the background were two factors that were especially of concern to us. Surprisingly, however, neither factor was able to make a strong impact on the accuracy of the results. Therefore, the experiments showed that there is no need to include specific restrictions on the camera angles and the locations the videos should be taken.

5. RELATED WORK

Edelmann-Nusser, A., et al evaluated the validity of the tennis sensor applications such as the HEAD Tennis Sensor and Babolat Pure Drive Play [1]. In order to operate the applications developed by these companies, users have to attach special sensors that track their swings and strokes to their tennis rackets. The data recorded from the sensors is then sent to the application where users can view the motions of their serves along with other statistics such as racket head velocity. Unlike my work, the sensor-based applications focus more on the statistics of the strokes and less on the visualization. These applications also provide more constructive tips and training videos alongside the statistics, which my application does not possess. However, the accuracy of their statistics may sometimes be in question.

Seggerman, Ryan created an application that explores shot probability through simulation [2]. The application Seggerman proposed sets up a simulation that allows two players to choose the location of their shots and respond accordingly to the incoming shots. This application seeks to improve tennis performance by enhancing decision making skills rather than improving the motion of the strokes, which is what my application seeks to achieve. While knowledge of better shot selection can be enhanced using Seggerman’s application, it would be difficult to execute the shots without proper motion.

Neal, Bradley S., et al. investigated the validity and reliability of an application known as Hudl Technique [3]. The application serves the purpose of letting users compare their performance in a sport to that of a professional in slow motion. It also allows users to make comments and tips over their videos. As of now, Hudl Technique may be better than my application in every facet. My application only displays corresponding frames side by side while Hudl Technique is capable of displaying the two videos simultaneously while aligned.

6. CONCLUSIONS

To provide a less costly and more efficient alternative for tennis players to improve their performance, we proposed the creation of a mobile application. By combining mobile computing and machine learning, we created an application that requests two videos containing the same tennis strokes from the user. Once the videos have been inputted via the frontend UI, a backend server containing codes necessary for machine learning will align the two videos and send back key corresponding frames to be displayed on the application [14]. This approach allows users to precisely visualize the difference between their strokes and that of a professional. In our experiments, we evaluated whether the camera angle and the presence of individuals in motion in the background can affect aligning of the two videos and the accuracy of the results. It was found that the two factors largely do not impact the end results, meaning there is no need to add extra precautionary instruction for the users.

As of the moment, some limitations to the application include a crude UI, long processing time, and some inaccuracies in the output. While simplicity is not necessarily bad, the current frontend user interface appears incomplete and to some degree, unprofessional. The long processing time of the videos may give the users impression that the application is broken or unusable. And inaccurate outputs will simply result in dissatisfaction from the users. These are issues that need to be quickly addressed.

In the near future, to solve the current limitations, we will first address the long processing time of the videos by temporarily adding a loading screen of sorts. Then we will focus on reworking the UI to look more aesthetically pleasing. We will try to increase the accuracy of the outputs by modifying the algorithms and also expedite the time it takes to process the videos. Lastly, we can also add constructive tips according to the difference in motion between the two videos inputted.

REFERENCES

[1] Edelmann-Nusser, A., et al. "Validation of sensor-based game analysis tools in tennis." International Journal of Computer Science in Sport 18.2 (2019): 49-59.

[2] Seggerman, Ryan. "Virtual Tennis: Serving Up A Simulator." (2022).

[3] Neal, Bradley S., et al. "Is markerless, smart phone recorded two-dimensional video a clinically useful measure of relevant lower limb kinematics in runners with patellofemoral pain? A validity and reliability study." Physical Therapy in Sport 43 (2020): 36-42.

[4] Abrams, Geoffrey D., Per A. Renstrom, and Marc R. Safran. "Epidemiology of musculoskeletal injury in the tennis player." British journal of sports medicine 46.7 (2012): 492-498.

[5] Pennycook, Alastair, and Emi Otsuji. "Mundane metrolingualism." International Journal of Multilingualism 16.2 (2019): 175-186.

[6] Jordan, Michael I., and Tom M. Mitchell. "Machine learning: Trends, perspectives, and prospects." Science 349.6245 (2015): 255-260.

[7] Brynjolfsson, Erik, and A. N. D. R. E. W. Mcafee. "Artificial intelligence, for real." Harvard business review 1 (2017): 1-31.

[8] Kovacs, M. S. "Applied physiology of tennis performance." British journal of sports medicine 40.5 (2006): 381-386.

[9] Baade, Robert A., and Jeffrey O. Sundberg. "What determines alumni generosity?." Economics of Education Review 15.1 (1996): 75-81.

[10] Borisova, O. V. "Tennis: history and the present." Pedagogics, psychology, medical-biological problems of physical training and sports 11 (2012): 119-124.

[11] Mahesh, Batta. "Machine learning algorithms-a review." International Journal of Science and Research (IJSR).[Internet] 9 (2020): 381-386.

[12] Martin, William, et al. "A survey of app store analysis for software engineering." IEEE transactions on software engineering 43.9 (2016): 817-847.

[13] Lugaresi, Camillo, et al. "Mediapipe: A framework for perceiving and processing reality." Third Workshop on Computer Vision for AR/VR at IEEE Computer Vision and Pattern Recognition (CVPR). Vol. 2019. 2019.
[14] Borgman, Christine L., Donald O. Case, and Charles T. Meadow. "The design and evaluation of a front-end user interface for energy researchers." Journal of the American Society for Information Science 40.2 (1989): 99-109.
[15] Goadrich, Mark H., and Michael P. Rogers. "Smart smartphone development: iOS versus Android." Proceedings of the 42nd ACM technical symposium on Computer science education. 2011.