International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2020): 7.803

Virtual Tennis Coach using Pose Estimation

Taha Ali[1], Yashwant Yadav[2], Sunny Jayant[3], Jasraj Meena[4]

[1, 2, 3, 4]Department of Information Technology, Delhi Technological University
taslash74[at]gmail.com, yyashwant245[at]gmail.com, jayantsunny08[at]gmail.com, jasrajengg[at]gmail.com

Abstract: ***Tennis enjoys a considerable following in India, although it is limited to urban areas but still it is counted among the most popular sports in India. India has produced a number of tennis players, who have achieved international recognition. However, Coaching for this sport is expensive making it unaffordable for many. A method for creating virtual coaches is presented which uses tenser flow posenet algorithm. A database is developed which contains the images of top tennis players. After applying machine learning algorithms on it, the data is able to segregate good posture from the bad ones. The player is provided with an option to compare his game with the game of top ATA players and the end result is shown in visual format to make it more user friendly. The system works on real time data.***

Keywords: tenser flow, pose net algorithm, machine learning, ATA players, virtual coach, tennis

1. Introduction

Sports include different forms of competitive physical activities and games which help us to maintain and improve our physical fitness, along with that it provides enjoyment and entertainment to the spectators. There are hundreds of sports including football, tennis, golf, cricket, table tennis etc. In order to play these sports professionally, it requires proper training and coaching. This can only be done if there is availability of professionals or coaches. In INDIA people are willing to play sports at professional level but due to lack of resources and coaching facilities they aren't able to do that. It is an area of concern. In order to deal with this situation, a model is trained using the data of professional players like how they play shots, what should be the correct body position to play a particular shot. It provides the user to upload his Images to check his body posture, then model performs calculation and provide result in visual format on a human skeleton. As there are a lot of videos on the internet that provide the teaching for any sport, but these videos are only one side interaction. Video is just a static learning experience but modern user requires a one on one Interaction. The main idea is to develop a model that will help participants to learn and improve there playing skills in tennis by comparing their playing pattern with a professional player like Roger Federer, Rafael Nadal, Novak Djokovic etc. While many alternate pose detection systems have been open sourced, all require specialized hardware, as well as quite a bit of system setup. With Pose net running on Tensor flow anyone with a decent webcam-equipped desktop or phone can experience this technology right from within a web browser.

2. Related Work

Several researches have been done in the area of human pose estimation. The publication [1] uses human pose estimation library to detect fall using a Kinect based camera. It uses 2D human pose estimation to detect a drastic change in the position of head coordinates of human being. This is useful in case of fall detection of older people. There have been several publications that aim to improve the accuracy of publication [2]. It uses part affinity fields and part confidence maps for body part association. Also uses bipartite matching to find association between body parts. This paper provides detailed information about the 2D pose estimation algorithm but does not talk about any of the applications for this library. The publication [3]aims to find out size, orientation or position of human body parts in sports videos. It uses salient region detection and foreground segmentation via skin tone detection on sequence of input images. Then body parts initialization is used to obtain image Silhouette. It uses body parts model for body parts estimation. It extracts 5 basic key points and generates 7 sub key points from them. This paper aims to provide accurate tracking for movement of body parts in sports, however it does involve processing this information for improvisation. Publication [4] focuses on the analysis of player movement in tennis video. It finds out the probability of success of the shot using unsupervised machine learning model, fetches joint position coordinates from input RGB images, combines this information with player position to create feature vectors. These are then used to estimate the probability of shot success this is whether the shot will provide point or not. However, it does not provide information to the athlete to improve his shot. This paper focuses on the probability of failure rather than the correctness of the pose in shots like forehand and backhand.

Conventional research on sport image analysis have been performed in [5]. In this research, it aims to detects shots such as backhand and forehand by analyzing the positional information of the ball and the players in the video using inter frame difference. However this does not focus on improvisation of pose position by comparison with accurate form. As a research related to posture state of the athlete, research has been done in [6] to identify characteristics of different body part movements while hitting a serve. The paper uses marker less motion capture technique to record motion and velocity of wrist and elbow to classify the type of serve, as flat, kick and slice serves. It requires eight cameras and controlled environment to identify 3D posture state. It takes time and effort to arrange and requires special calibration depending on the environment.

3. Proposed Method

Virtual tennis coach methodology is shown in Fig.1. It involves data creation, data preprocessing, removal of

outliers, image analysis and video analysis. The output is shown as Human figure where joints are highlighted with colours (green, orange, red) which represent the correctness of the posture.

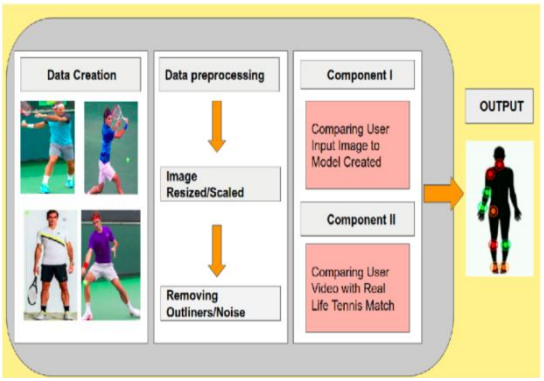

Figure 1: Overview of Steps involved in the algorithm

Fig.2 shows the result of posenet algorithm. Itgives 18coordinates numbered from 0 to 17 which represent Nose, Neck, Rshoulder, Relbow, Rwrist, Lshoulder, Lelbow, Lwrist, Rhip, Rknee, Rankle, Lhip, Lknee, Lankle, Reye, Leye, Rear, Lear respectively.

A. Dataset Creation

In order to make the research relevant, real timedata is captured from various sources. Videos are analyzed and then frames are chosen out of them. Frames are cropped, zoomed to normalize the dataset. Choosing a right frame is very crucial and important, as the accuracy of the whole system will be based on it.

Figure 2: Output of posenet algorithm

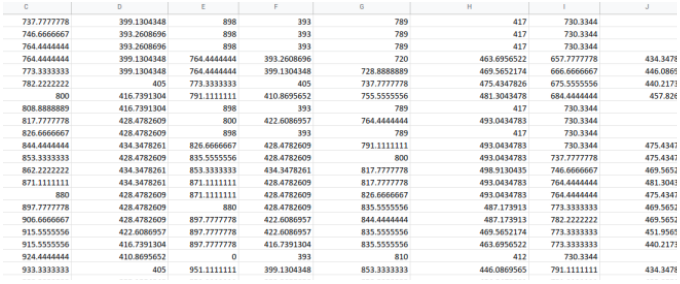

Figure 3: Extraction of joint coordinates using Posenet

B. Data Preprocessing

Outliers points were identified and removed from the dat using mean calculation technique. A sample has been shown in Fig.4 and Fig.5.

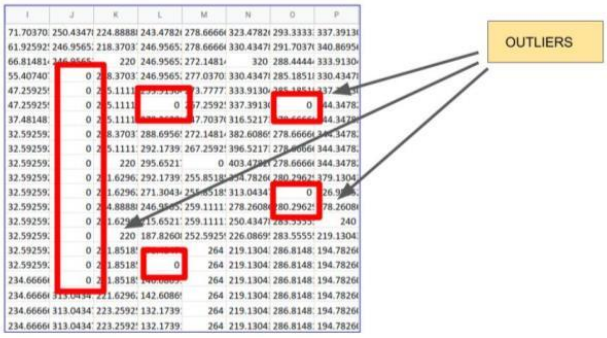

Figure 4: Identification of Outliers

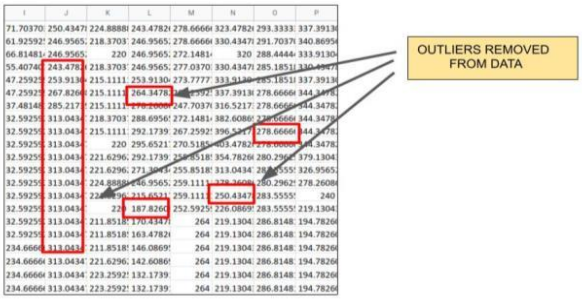

Figure 5: Removal of Outliers

C. Algorithm Component 1-Image Analysis

Algorithm 1 VIRTUAL TENNIS COACH COMPONENT I

1: *Set S* ← (coordinate Nose,coordinate RElbow, coordinate RWrist ,..... ,coordinate REar)
2: *DataSet D* ← (S1,S2,S3,.....,S200)
3: **function** ML_MODEL(D,k=3)
4: *ClusterGood* ← KMeans(D,k)
5: *ClusterAverage* ← KMeans(D,k)
6: *ClusterBad* ← KMeans(D,k)
7: *Set INPUT_IMAGE* ← (coordinate Nose,coordinate RElbow,..... ,coordinate REar)
8: **for** $k \leftarrow 1$ to 18 **do**
If (INPUT_IMAGE[k] lies under ClusterGood)**then** $Good + 1$
If (INPUT_IMAGE[k] lies under ClusterAvg)**then** $Avg + 1$
If (INPUT_IMAGE[k] lies under ClusterBad)**then** $Bad + 1$
9: return **Max**(Good, Avg, Bad)

The first part of the algorithm includes training of the model using the dataset and generation of three cluster groups i. e. good, average and bad cluster. Set S contains the location of the 18 body joints under consideration. D denotes the dataset used to train the model. For each joint position we extract its location from all of the 200 images. For example we extract the location (i. e. coordinates) of left elbow from all the images. These 200 left elbow coordinates are then supplied to density based k-means clustering algorithm.

Figure 6: Output of Component 1 Image Analysis

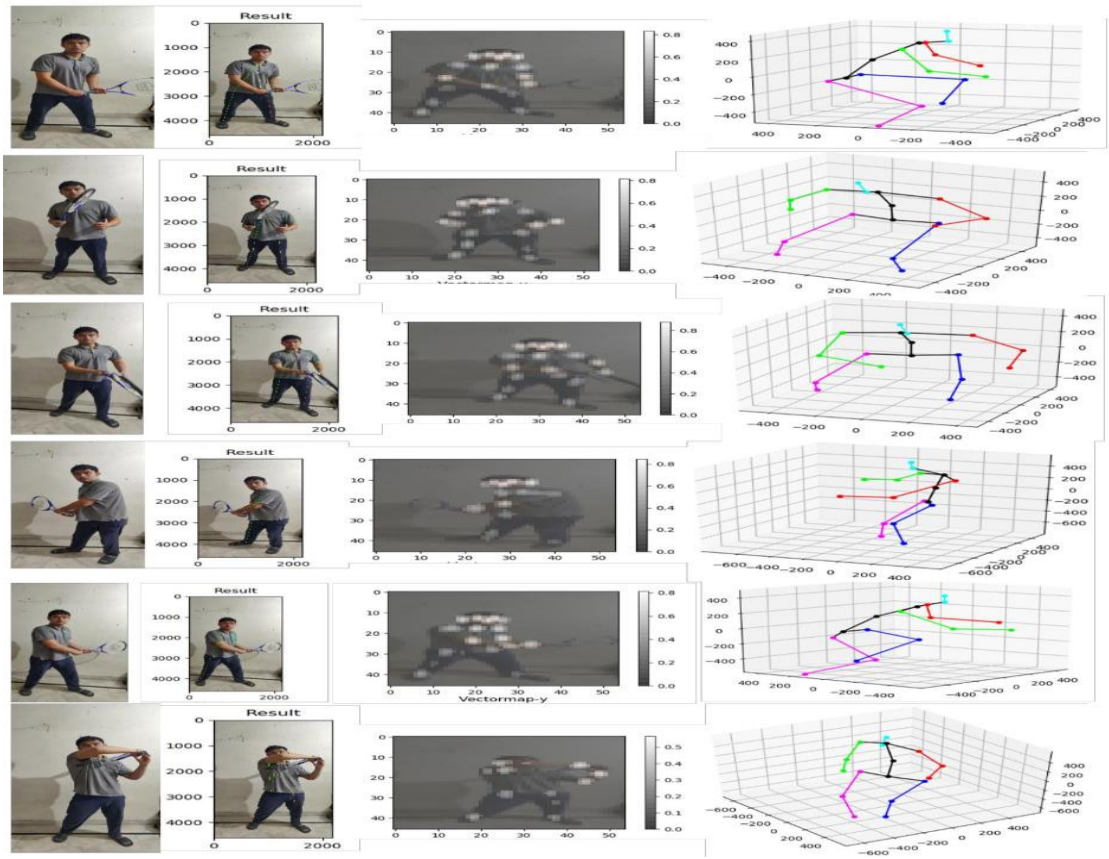

Figure 7: Generating human skeleton, heat-map and 3D posture from input images

The value of hyper parameter k is set to 3. The cluster with the highest density is classified as good cluster, the cluster with the lowest density as bad cluster and the middle one as average cluster. This process is repeated for all the 18 body joints. Now we have obtained a way to classify the position of each joint as good, average or bad. Step 8 of the algorithm shows the set INPUT_IMAGE.

This set is generated by applying posenet algorithm on image given by the user. For example this image could be of the ready pose i.e. the standing position to receive a serve. Posenet gives us the location of the 18 joint coordinates of the user. Three counters good, bad and avg are maintained. For each point, we find out the nearest cluster. If the point belongs to cluster 1 i. e. the good cluster value of good is incremented by 1. Similarly if it belongs to cluster 2and 3 value of avg and bad is incremented by 1 respectively. In the end the maximum of good, bad and avg. is returned. Thus the overall posture of the user is classified as good, average or bad. In the final output Fig.6, the classification of each joint is shown separately to provide more detailed information to the user. For example the posture of legs might have been correct but the elbow position might be wrong.

D. Component 2-Video Analysis

Algorithm 1 VIRTUAL TENNIS COACH COMPONENT II

1: S_INPUT_VIDEO S_PLAYER_VIDEO
2: **for** *Each Frame In* S_INPUT_VIDEO ,S_PLAYER_VIDEO **do**
Cosine_Similarity ← S_INPUT_VIDEO_FRAME,S_PLAYER_VIDEO_FRAME
if (Cosine_Similarity >80)**then** *Good* + 1
else if(Cosine_Similarity >65)**then** *Avg* + 1
else *Bad* + 1
return Max (Good,Avg,Bad)

In the second component we compare user's input video with professional tennis player videos. S_INPUT_VIDEO is the sample video of the user, S_PLAYER_VIDEO denotes the video of professional player. Both the videos are divided into equal number of frames. Now one frame is picked from user video and the corresponding frame is picked from professional tennis player video. Posenet algorithm is applied on each frame to generate the location coordinates of all the18 body joints. These 18 coordinates are combined to form a vector. Point vector is generated for both the frames. Then, the cosine similarity between the two vectors is calculated. If the value of cosine similarity is greater than 80, Good is incremented by 1. If the value of cosine similarity is greaterthan65 and less than 80, Avg is incremented by 1, else Bad is incremented by 1. In the end the maximum of good, bad and avgis returned. Thus the overall game play of the user is classified as good, average or bad in comparison to a professional player. Fig.8shows the output in a frame by frame format.

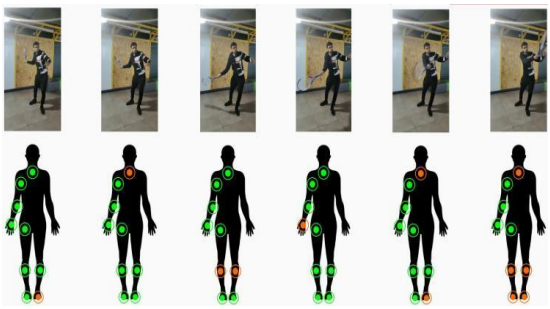

Figure 8: Output of Component 2 Video Analysis

4. Conclusion

Through this paper we successfully devised a model to reliably and accurately assist people with their tennis coach needs without a real coach at all. The system is highly reliable and robust because it works on real time data, instead of some old data from database. Future prospects is to expand the dataset size and to further increase the accuracy of the model. IOT can be used to make the video camera smart. The system should be able to handle big Input file smoothly and should not lag. More graphical and Visual forms of output can be added to make it more user friendly.

References

[1] Angal, Y., &Jagtap, A. (2016, December). Fall detection system for older adults. In *2016 IEEE International Conference on Advances in Electronics, Communication and Computer Technology (ICAECCT)* (pp.262-266). IEEE.

[2] Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp.7291-7299).

[3] Jalal, A., Nadeem, A., &Bobasu, S. (2019, March). Human body parts estimation and detection for physical sports movements. In *2019 2nd International Conference on Communication, Computing and Digital systems (C-CODE)* (pp.104-109). IEEE.

[4] Kurose, R., Hayashi, M., Ishii, T., & Aoki, Y. (2018, January). Player pose analysis in tennis video based on pose estimation. In *2018 International Workshop on Advanced Image Technology (IWAIT)* (pp.1-4). IEEE.

[5] ChihiroAntoku, Masayuki Kashima, Kiminori Sato, Mutsumi Watanabe, "Research for automatic tennis play recognition and recording based on motion analysis", IPSJ SIG Technical Report, 2013.

[6] Alison L. Sheets, Geoffrey D. Abrams, Stefano Corazza, Marc R. Safran, Thomas p. Andrriacchi, "Kinematics Differences Between the Flat, Kick, and Slice Serves Measured Using Markerless Motion Capture Method", Annals of biomedical engineering 39.12 (2011): 3011-3020.

[7] Varun Ramakrishna, Daniel Munoz, Martial Hebert, J. Andrew Bagnell, Yaser Sheikh, "Pose Machine: Articulated Pose Estimation via Inference Machines", European Conference on Computer Vision. Springer International Publishing (2014).

[8] Shih-En Wei, Varun Ramakrishna, Takeo Kanade, Yaser Sheikh, "Convolutional Pose Machines", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016).

[9] Wang, C., Wang, Y., Lin, Z., Yuille, A. L., &Gao, W. (2014). Robust estimation of 3d human poses from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp.2361-2368).

[10] Du, Y., Wong, Y., Liu, Y., Han, F., Gui, Y., Wang, Z.,. . . &Geng, W. (2016, October). Marker-less 3d human motion capture with monocular image sequence and height-maps. In *European Conference on Computer Vision* (pp.20-36). Springer, Cham.

[11] Chow, J. W., L. G. Carlton, Y. T. Lim, W. S. Chae, J. H. Shim, A. F. Kuenster, and K. Kokubun. Comparing the pre-and post-impact ball and racquet kinematics of elite tennis players' first and second serves: a preliminary study. J Sports Sci 21 (7): 529–537, 2003.

[12] E. Simo-Serra, A. Quattoni, C. Torras, and F. Moreno-Noguer, "A Joint Model for 2D and 3D Pose Estimation from a Single Image, " CVPR, 2013.