Toward the Perfect Stroke: A Multimodal Approach for Table Tennis Stroke Evaluation

Panyawut Sri-iesaranusorn*, Felan Carlo Garcia*, Francis Tiausas*,
Supatsara Wattanakriengkrai*, Kazushi Ikeda*, and Junichiro Yoshimoto*
* Nara Institute of Science and Technology, Nara, Japan
Email: {sri-iesaranusorn.panyawut.sg0, juniti-y}@is.naist.jp

***Abstract*—In table tennis, developing a consistent and proper stroke is quite challenging, perhaps even more so for non-professional players. To build such consistency in beginner players, there is a need to understand how the stroke differs between beginner and standard players. So far, prior works have used video, accelerometers embedded in the table tennis rackets themselves, or infrared (IR) depth sensors for capturing and evaluating the stroke. However, these methods face certain challenges such as having insufficient data to analyse complete strokes, time-consuming and costly data collection, and use of non-prevalent equipment. Hence, to improve the beginner player's performance, an ubiquitous method using readily accessible commercial devices is essential for stroke evaluation. To achieve the goal of this study, we (i) recorded video and accelerometer data from standard and beginner players using consumer-grade products, and (ii) analysed the stroke consistency between both groups. The results of both video-based and accelerometer-based data show the differences in the strokes between both kinds of players. These findings motivate us to further examine methods to help beginner players improve by providing guidance through procedural knowledge of a standard player's stroke, and implement applications for motor-skill instruction.**

***Index Terms*—Table tennis, Stroke detection, Motion Capture, Joint Kinematics, Sports Analytics**

I. Introduction

Table tennis is one of the fastest ball games in the world, requiring precise reaction time and strokes to perform different kinds of spins and tactics [1], [2]. The nature of this game requires players to consistently build good habits in their forms as they develop muscle memory for their strokes. Given this nature, proper form building and stroke mechanics are vital not only for the competitive aspects of the game but also for its enjoyment – especially for casual players.

Building a consistent stroke is a challenge, especially for beginner players. Although professional coaching can usually provide assessments for stroke technique and quality, this also requires time and monetary investment on the part of the casual player [3]. Hence, developing an ubiquitous way of evaluating table tennis stroke through readily accessible commercial devices would greatly help beginner players improve their performance through self-practice and a little effort.

Recent advancements in mobile computing technology have given birth to a plethora of wearable devices marketed towards physically active groups of people that aim to track their fitness through activities such as running [4]. Despite the success of fitness trackers for daily physical activities, the applications of such devices specifically targeted towards particular sports are still in their infancy. For example, several researches and applications have been done for tennis such as sensor-based swing and stroke classification as a tool for personal feedback and coaching [5]. Similar research has also been done with regards to the baseball pitching [6], and the golf [7].

In terms of table tennis applications, prior studies used video-based methods and ball tracking as a way to give feedback to players [8]. A recent study used an accelerometer embedded in a table tennis racket to classify strokes made by beginner players [9]. However, there methods still face certain challenges such as the data being insufficient to provide an analysis of a complete stroke and have it replicated by players within a given time. Other prior works used infrared (IR) depth sensors to detect the in-corrected played strokes accurately, but the method is time-consuming and costly, requiring a complicated sensor set up and custom proprietary software [10]. Hence, to improve the beginner players' skills, an ubiquitous solution using prevalent equipment for stroke evaluation is essential. Furthermore, to evaluate the stroke consistency of both players' groups, multimodal approaches have shown to perform better as opposed to unimodal approaches specifically in the field of human activity sensing and recognition [11].

In this study, we aim to find the factor for improving beginner players' strokes by first investigating the difference in stroke consistency between the standard and the beginner players. We recorded videos and accelerometer signals for two groups of players using consumer-grade products like the Apple Watch. For video-based data, we use DeepLabCut to extract arm positions [12]. Subsequently, we removed the noise signal from both extracted arm positions and accelerometer signals using Kalman filtering. Lastly, we evaluated the stroke consistency of both beginner and standard players through dynamic time wrapping with root mean square error.

II. Methodology

A. Dataset

To investigate the stroke based on the player's skill, we selected two groups of table tennis players for this study. The first group is composed of standard players who have playing experience of more than 5 years, or have participated in prior table tennis competitions according to the criteria proposed by [13]. The second group is composed of beginner

978-4-907626-48-8/21/$31.00 ©2021 IPSJ

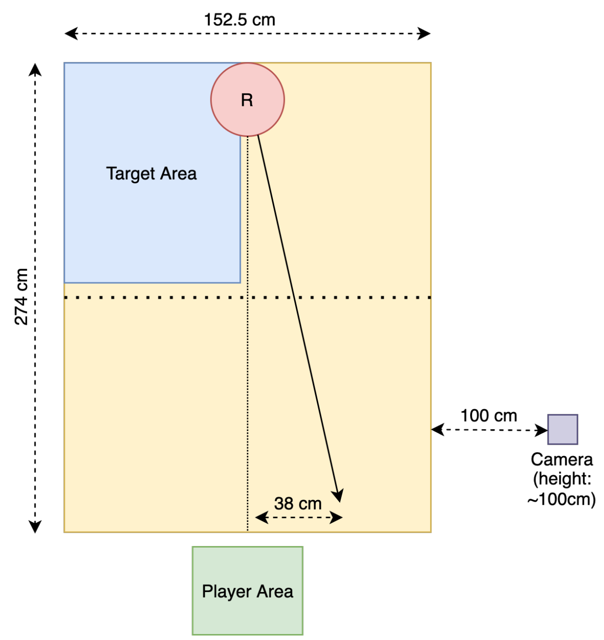

Fig. 1: Experiment table and equipment setup

or casual players who fall short of the requirement for the first group. The participants consist of **three standard**, and **seven beginner players**. To maintain consistency, the participants were restricted to right-handed players.

B. Experiment Setup

The main actions for table tennis are the forehand and backhand strokes used to return the ball to the opponent's side of the table. Thus, we focused on the stroke of receiving the ball from the opponent by using forehand and backhand. Similar to a prior study [14], we set up a ball-shooting robot at the center of one side of the table, and let the players stand on the opposite side. The shooting angle is adjusted based on the stroke being captured at the time (approx. 7.9 degrees to the right side of the table for forehand, and to the left for backhand). Participants were positioned in the center area and allowed to freely move laterally to either side. Figure 1 shows the setup dimensions for the forehand experiment. Participants were instructed to return the ball from the robot to the target area. If the ball is hit and returned to the target area, a **hitting score** is counted. In the case of the backhand, the setup was simply mirrored horizontally. In this study, participants consecutively performed 10 forehand strokes, and 10 backhand strokes.

We recorded the participants using an Apple Watch 5, and a GoPro8 action camera. We tracked the motion of the wrist using a 3-axis accelerometer data logger from the Apple Watch. For the camera, we tracked four positions: (1) the shoulder joint, (2) the elbow joint, (3) the wrist joint, and (4) the center of the racket. These positions were considered because they can be used to determine the stroke at the time the player hit the ball. The camera was set to perform video capture of 10 consecutive stroke events for both forehand and backhand over 35 seconds. We configured the device capture rate at 60 frames per second with a 640x360 video resolution. Furthermore, to ensure that player movements are properly captured, the focal length of the camera was set to 19 mm to avoid visual distortion and placed approximately 100 cm away from the table.

C. Accelerometer and Preprocessing

To measure the consistency of each stroke, we analysed the x, y, and z-axis of an accelerometer signal data gathered using an Apple Watch 5 with a sampling rate of 50Hz. Kalman filtering [15] was performed on the accelerometer data to filter out noise in the signal for each of the axis. The key idea of this filtering technique is to produce estimates of hidden variables based on inaccurate and uncertain measurements. This allows us to remove jitters in data without eliminating some signal frequencies when compared to a moving average noise reduction approach.

D. Video-based Semi-automatic Detection of Arm's Position

We used DeepLabCut, a supervised CNN-based tool for animal posture quantification, to analyze the players' stroke from the videos [12]. As shown in Figure 2, this tool allows us to automatically annotate the positions of objects in videos, once their positions are manually provided in a few frames. Since DeepLabCut is based on ResNet50, a supervised learning model, it requires a sample of our videos for training the model. To create the training dataset, we randomly selected 10 frames per video, a total of 200 frames for training, and manually annotated the positions of three parts of the arm and the table tennis racket. Once trained, the tool was then used to estimate the positions throughout the remaining frames. In the end, we obtained the sequences of x and y coordinate of four tracking positions for each participant. We treated the results of the arm's position as time-series data in the same fashion as the x, y, and z-axis of accelerometer signal.

E. Dynamic Time Warp for Stroke Consistency Evaluation

We used Dynamic Time Warping (DTW) and Root Mean Square Error (RMSE) on pairs of two strokes from each player, in order to evaluate the consistency between the stroke motion signals from video motion tracking and the Apple Watch accelerometer. Prior works applied this algorithm for analysis on various fields such as basketball [16], and surgery [17]. The use of DTW allows for matching the peaks of the strokes by reducing the effects of time and shifting distortion between the two signals and detects how similar their phase and shapes as described by Algorithm 1.

RMSE was used as the metric to compare the consistency of the stroke motions after using DTW on the signals by measuring the square root of the difference between two compared signals. In this context, a sizeable difference in RMSE implies greater inconsistency and erraticism in the stroke motion. For each player, we defined the reference signal as the signal that produces the lowest averaged RMSE when paired with other strokes. It should also be noted that we discarded the first and last strokes because they contained

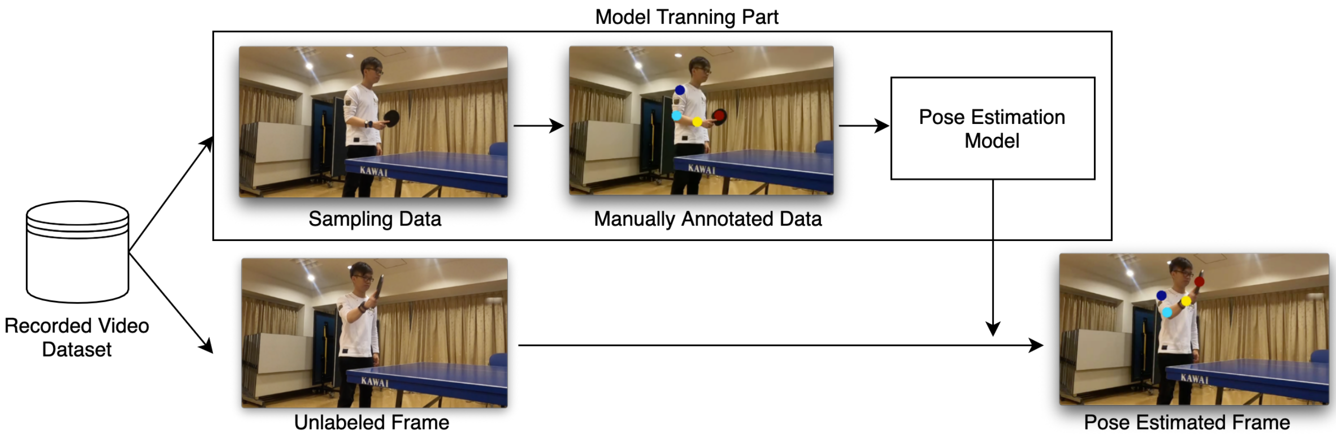

Fig. 2: Procedure of Semi-automatic detection via DeepLabcut

signals from initial and ending postures during the recording. For each player, we calculated RMSE from the set of signals, and used it as the feature for analysis. The video tracking approach contains eight features, x and y coordinate of four tracking points from the GoPro8 camera. For the accelerometer tracking approach, there are three features, x-axis, y-axis, and z-axis of the accelerometer from Apple Watch.

Algorithm 1: Dynamic Time Warping Algorithm

Data: s, t: the sequences
Result: d: the minimum distance
n = s.length() ;
m = t.length() ;
DTW = array[0...n][0...m] ;
for $i \leftarrow 0$ **to** n **do**
  for $j \leftarrow 0$ **to** m **do**
    DTW[i][j] = infinity
  end
end
DTW[0][0] = 0 **for** $i \leftarrow 0$ **to** n **do**
  for $j \leftarrow 0$ **to** m **do**
    cost = RMSE($s[i]$, $t[i]$) ;
    DTW[i][j] = cost + minimum(DTW[$i-1$, j], DTW[i, $j-1$], DTW[$i-1$, $j-1$])
  end
end
return DTW[n][m];

To investigate which factor(s) were correlated with the consistency of the stroke motions, we performed statistical comparisons between the following pairs:

- successful v.s. failure results when the beginners tried forehand strokes.
- successful v.s. failure results when the beginners tried backhand strokes.
- standard v.s. beginner players when the players tried forehand strokes.
- standard v.s. beginner players when the players tried backhand strokes.
- successful v.s. failure results in case of forehand strokes.
- successful v.s. failure results in case of backhand strokes.

The comparison was implemented by two-sample t-test (significance level = 1.0e-4 w/o multiple comparison) for each coordinate.

For the relationship between the player's stroke and the hitting score, we defined the **stability error** for representing the player's stroke consistency. We applied Principle Component Analysis (PCA), which is a technique for reducing the dimensionality of the features [18]. We then selected the first rank component as the stability error. For example, we reduced the dimensional of the eight features of the video tracking approach into one feature called stability error. This procedure was also used with the features for the accelerometer tracking approach.

III. Results

Table I shows the averaged and standard deviation of RMSE for video tracking and accelerometer tracking, and the hitting score of each player. We have eight features for video tracking, and three features for accelerometer tracking. For simple analysis, we average the feature depending on the feature group and the dominant hand such as video tracking of forehand, video tracking of backhand, accelerometer-based tracking of forehand, accelerometer-based tracking of backhand. Noted that the results of the video tracking approach are based on the pixel-to-pixel difference of footage from the GoPro8 camera, while the results of accelerometer tracking are based on the inertial measurement unit (IMU) in the Apple Watch 5.

For pair-wise comparisons, the coordinates showing significant difference were as follows:

As results of pair-wise comparisons,

- **standard v.s. beginner players when the players tried forehand strokes**: x-axis, y-axis, and z-axis of accelerometer, x and y coordinate of elbow joint, wrist joint, and the center of the racket.
- **standard v.s. beginner players when the players tried backhand strokes**: y-axis, and z-axis of accelerometer.

Figure 3 shows the relationship between stability and the hitting score for video tracking and accelerometer tracking approach, separately. It is obvious that there are the difference

TABLE I: Average and standard deviation of RMSE within players forehand and backhand strokes

Participant	Video Tracking (RMSE)		Accelerometer Tracking (RMSE)		Hitting score	
	Forehand	Backhand	Forehand	Backhand	Forehand	Backhand
Standard#1	215.15 ± 73.54	237.34 ± 113.03	6.92 ± 0.71	4.27 ± 0.92	10	10
Standard#2	162.39 ± 100.22	538.77 ± 476.27	7.29 ± 0.74	6.76 ± 0.61	10	9
Standard#3	246.07 ± 178.81	205.62 ± 105.84	3.10 ± 0.60	4.34 ± 0.85	10	7
Beginner#1	664.18 ± 387.56	87.91 ± 36.28	10.25 ± 3.76	10.80 ± 3.21	10	10
Beginner#2	1064.78 ± 869.57	148.96 ± 66.21	8.59 ± 0.90	9.87 ± 2.31	7	6
Beginner#3	402.72 ± 230.67	336.84 ± 208.53	11.66 ± 4.06	7.45 ± 1.40	7	8
Beginner#4	266.43 ± 124.81	392.09 ± 275.54	10.43 ± 1.11	13.30 ± 1.80	10	8
Beginner#5	700.82 ± 319.90	559.41 ± 164.06	17.52 ± 1.63	11.35 ± 1.85	6	7
Beginner#6	603.71 ± 457.87	394.84 ± 428.63	7.86 ± 2.20	6.42 ± 1.01	8	9
Beginner#7	727.38 ± 633.50	291.62 ± 141.68	21.38 ± 5.90	9.10 ± 4.16	7	6

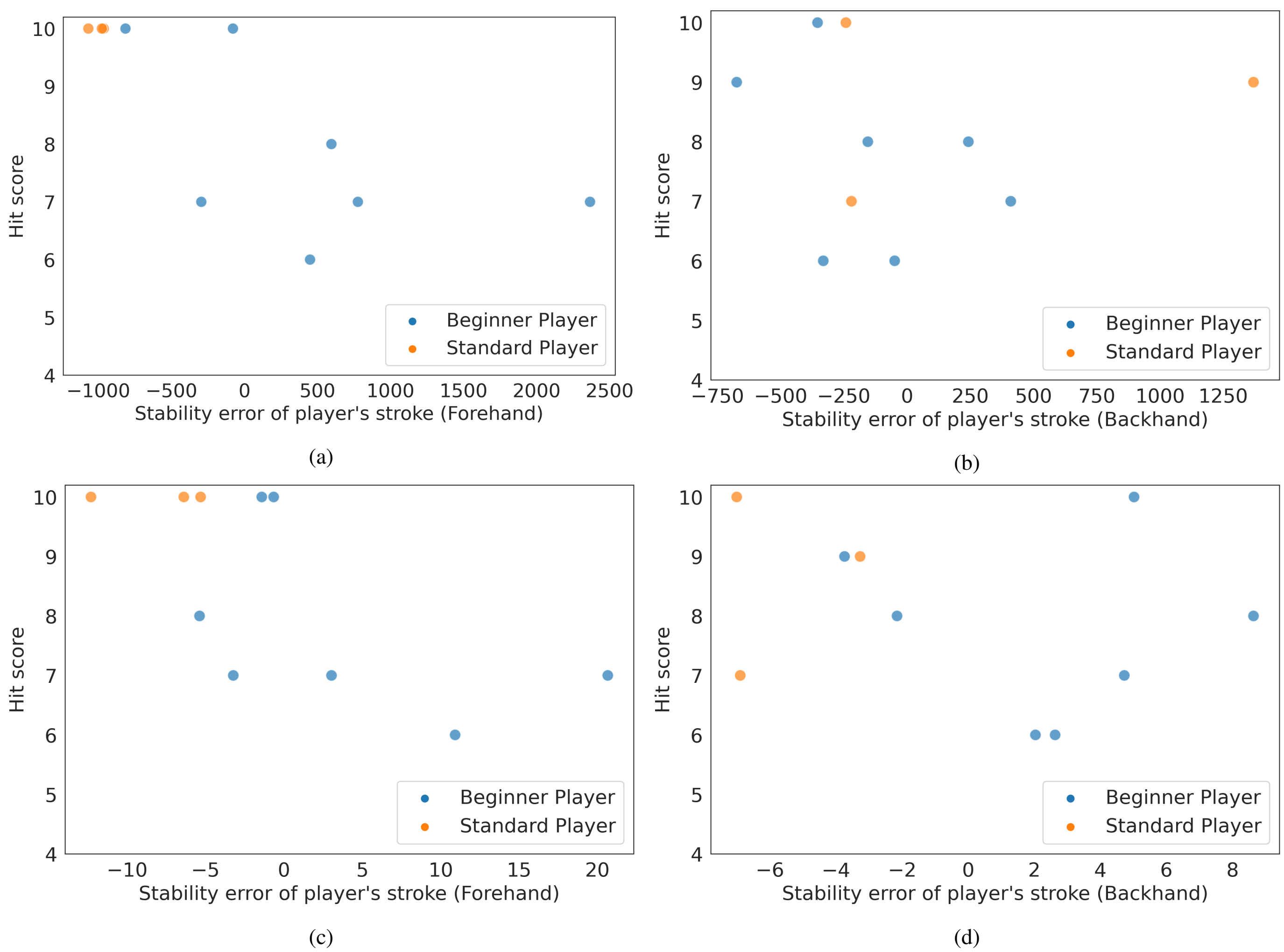

Fig. 3: Scatter plot between stability error of player's stroke and their hitting score. (a) and (b) are the results of video tracking approach. (c) and (d) are the results of accelerometer tracking approach

of stability error between standard and beginner players for forehand results for both approaches. On the other hand, there are unclear differences between two players groups for backhand. These imply that stability error can be an efficient candidate for evaluating player's skill.

IV. Discussion

Our results suggest that there is a difference in stroke consistency between standard and beginner players based on the features from video tracking and the accelerometer tracking approach. Additionally, analysis shows significant difference between the successful and fail stroke groups. For the relationship between the stability error and the hitting score, we see a clear difference between the two groups during the forehand stroke for both video and accelerometer-based approaches. However, the results of backhand seem to be unclear, possibly indicating an effect due to the player's dominant hand.

There are three main limitations that affect the accuracy and reliability of our result. The first is that the result of using DeepLabCut is affected by the camera position, especially for the backhand. Regarding the full stroke motion for the forehand of the right-handed player, we set the camera position to see the entire stroke. On the other hand, when the player performs the backhand, their left hand can be intercepted by their body that leads to cause of inaccurate analysis for the backhand. To resolve this issue, we aim to place another camera to capture the entire player's stroke for the backhand. The second is that DTW discards the timing information. While DTW is able to capture how consistent is the stroke motion, the duration of the stroke between two players might be important for analysis. Hence, measuring additional insight such as speed of the stroke can be considered given that stroke consistency, hit timing, and hand-eye coordination movements are affected by speed not only in table tennis but in other sports as well [19]. The last limitation is the small size of participants and stroke types in this study. Our data has a small amount of fail strokes, and are lacking in such strokes from standard players. These analyses should be validated by increasing the number of participants and strokes for all player groups in further studies.

V. Conclusion and Future Works

In this work, we present a multimodal approach for table tennis stroke evaluation. The results suggest that there is a difference in stroke consistency between standard and beginner players based on the features from video tracking and accelerometer tracking approach. Furthermore, we introduce the stability error which is represented as the quality of the player's stroke consistency.

With our evaluation procedure, we can monitor the beginner player's skill by using the stroke consistency through the prevalent equipment. Regarding the training of beginner players, one possible way is to use the stroke analysis of the standard player as a template of proper stroke for beginner players to follow. Since our data consists of the arm's tracked position and accelerometer tracked signal, we can build the stroke template from this data. Afterward, the beginner players can use the template to improve their strokes to be closer to the standard player. From an engineering standpoint, we aim to implement these applications on both smartwatch and mobile phone as a monitoring and training device for convenience and accessibility purposes in future directions of this study.

References

[1] A. Lees, "Science and the major racket sports: a review," *Journal of Sports Sciences*, vol. 21, no. 9, pp. 707–732, 2003.

[2] I. Faber, M. Elferink-Gemser, F. Oosterveld, and M. Nijhuis-van der Sanden, "Revision of two test items of the dutch motor skills assessment measuring ball control in young table tennis players: a reproducibility and validity study," *Annals of Research in Sport and Physical Activity*, vol. 5, pp. 53–69, 01 2015.

[3] M. Kondric, A. Zagatto, and D. Sekulic, "The physiological demands of table tennis: A review," *Journal of sports science & medicine*, vol. 12, pp. 362–70, 10 2013.

[4] J. McCann, D. Bryson, M. Malmivaara, R. Hurford, F. Saifee, L. RCA, D. Kane, A. Martin, P. Lam, G. Min, S. Morsky, X. Dong, I. Agnusdei, A. Taylor, D. Treadaway, M. Timmins, S. Underwood, J. Birringer, M. Danjoux, and W. Stahl, *Smart Clothes and Wearable Technology*, 02 2009.

[5] D. Connaghan, P. Kelly, N. E. O'Connor, M. Gaffney, M. Walsh, and C. O'Mathuna, "Multi-sensor classification of tennis strokes," in *SENSORS, 2011 IEEE*, 2011, pp. 1437–1440.

[6] M. Lapinski, E. Berkson, T. Gill, M. Reinold, and J. Paradiso, "A distributed wearable, wireless sensor system for evaluating professional baseball pitchers and batters," 09 2009, pp. 131–138.

[7] C. Egret, B. Nicolle, F. Dujardin, J. Weber, and C. Didier, "Kinematic analysis of the golf swing in men and women experienced golfers," *International journal of sports medicine*, vol. 27, pp. 463–7, 07 2006.

[8] W. Chen and Y. Zhang, "Tracking ball and players with applications to highlight ranking of broadcasting table tennis video," in *The Proceedings of the Multiconference on "Computational Engineering in Systems Applications"*, vol. 2, 2006, pp. 1896–1903.

[9] P. Blank, J. Hoßbach, D. Schuldhaus, and B. M. Eskofier, "Sensor-based stroke detection and stroke type classification in table tennis," in *Proceedings of the 2015 ACM International Symposium on Wearable Computers*, New York, NY, USA, 2015, p. 93–100.

[10] H. Hegazy, M. Abdelsalam, M. Hussien, S. Elmosalamy, Y. M. Hassan, A. M. Nabil, and A. Atia, "Online detection and classification of in-corrected played strokes in table tennis using ir depth camera," *Procedia Computer Science*, vol. 170, pp. 555 – 562, 2020.

[11] D. Strömbäck, S. Huang, and V. Radu, "MM-fit," *ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 4, pp. 1–22, 12 2020.

[12] A. Mathis, P. Mamidanna, K. Cury, T. Abe, V. Murthy, M. Mathis, and M. Bethge, "Deeplabcut: markerless pose estimation of user-defined body parts with deep learning," *Nature Neuroscience*, vol. 21, 09 2018.

[13] A. Zagatto and C. Gobatto, "Relationship between anaerobic parameters provided from maod and critical power model in specific table tennis test," *International journal of sports medicine*, vol. 33, pp. 613–20, 05 2012.

[14] T. Belli, M. S. Misuta, P. P. R. de Moura, T. d. S. Tavares, R. A. Ribeiro, Y. Y. S. d. Santos, K. J. Sarro, and L. R. Galatti, "Reproducibility and validity of a stroke effectiveness test in table tennis based on the temporal game structure," *Frontiers in Psychology*, vol. 10, p. 427, 2019.

[15] J. Woods and C. Radewan, "Kalman filtering in two dimensions," *IEEE Transactions on Information Theory*, vol. 23, no. 4, pp. 473–482, 1977.

[16] X. hu, S. Mo, and X. Qu, "Basketball activity classification based on upper body kinematics and dynamic time warping," *International Journal of Sports Medicine*, vol. 41, 01 2020.

[17] J. Jiang, Y. Xing, S. Wang, and K. Liang, "Evaluation of robotic surgery skills using dynamic time warping," *Computer methods and programs in biomedicine*, vol. 152, pp. 71–83, 2017.

[18] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, 2016.

[19] G. Delgado-García, J. Vanrenterghem, A. Muñoz-García, A. Molina-Molina, and V. M. Soto-Hermoso, "Does stroke performance in amateur tennis players depend on functional power generating capacity?" *The Journal of Sports Medicine and Physical Fitness*, vol. 59, 2019.