VIDEO SKELETONIZATION AND AI: A NOVEL APPROACH FOR ANALYZING KINEMATICS IN TABLE TENNIS

Enrico Ferlinghetti[1], Nicola Abeni[1], Ryushi Kumamoto[2], Ping Yeap Loh[3], Matteo Lancini[4]

DIMI[1] & DSMC[4], Università degli Studi di Brescia, Brescia, Italy
Industrial Design Course[2] & Department of Human Life Design and Science[3], Kyushu University, Fukuoka, Japan

This study presents a marker-less vision system that measures joint angles of table tennis players and relates them to the moment in which the racket hits the ball. AI plays a key role in the present setup: human body skeletonization is executed by the Mediapipe framework and a custom trained YOLOv8 network is used to track the position of the ball. The system was tested on ten male participants using two *GoPro Hero 10* cameras, placed in front and on the right side the players. Intraclass correlation coefficient was used to compare the joint angles measured from the two cameras, suggesting that performing a weighted average of the measured angles is necessary to increase the reliability of the system.

KEYWORDS: table tennis, marker-less measurements, vision system, gopro, mediapipe

INTRODUCTION: Motion analysis in sport offers valuable insights into how to improve technique, prevent injuries and enhance performances for athletes of all levels. Nowadays, the most common solutions to collect motion analysis data are marker-based motion capture systems (Rigozzi et al., 2023; Xu, 2024) and inertial measurements units (IMU) (Keskinoğlu et al., 2023). The review presented by (Rigozzi et al., 2023) focuses on racket sports, where motion analysis is used to compare similarities and differences of players at different playing levels, to evaluate variability in racket, upper limb and joint movement patterns and to describe movement differences associated with different ball spin levels. In particular, in table tennis, motion analysis is relevant to understand how the upper body joint angles change when the racket hits the ball according to the different ball speed and rotations (i.e. during forehand and backhand topspin strokes) (Wong et al., 2020). Several studies explore this topic using data provided by IMUs or marker-based vision systems (Bańkosz & Winiarski, 2017, 2020; Xia et al., 2020). The setups of both of them are expensive and require the subject to wear markers or sensors that require field work and, however small, can limit movements and affect performance evaluation. Tracking the human body movement in a completely non-intrusive way and without any additional cost is possible through the use marker-less based vision system, such as Mediapipe Pose (Bazarevsky et al., 2020), an AI-based framework provided by Google, able to recognize 33 body landmarks on the RGB image and infer their coordinates in the real world reference, as shown in Figure 2. The reliability and validity of Mediapipe is proven by (Latreche et al., 2023), which uses intraclass correlation coefficient (ICC) (Fisher, 1925) to compare the joint angles measured by Mediapipe with respect to those obtained by a

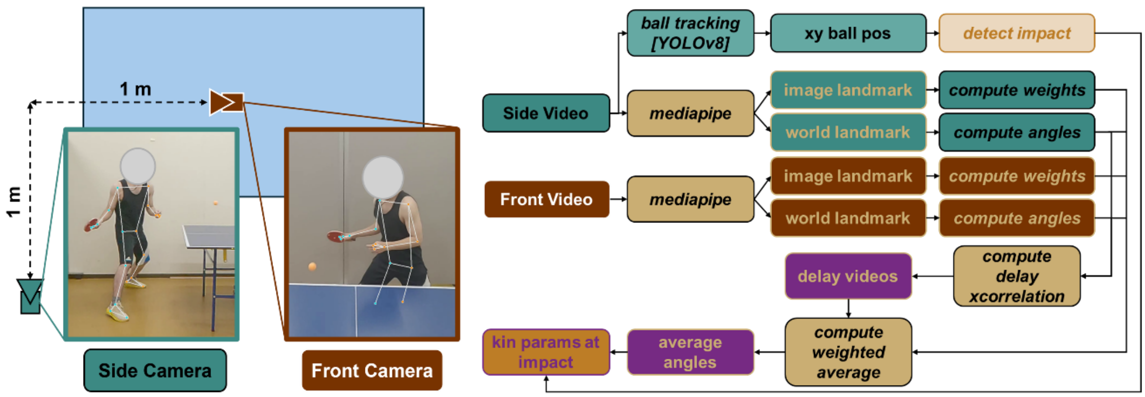

Figure 1: Flow Chart of the algorithm applied to identify the joint angles position at the moment of impact. In *italics* the operations, in roman the data

manual goniometer. The current paper describes the architecture of a marker-less based vision system for motion analysis in table tennis. The primary objective of the project is evaluating the feasibility of the setup, which focuses on the easy-to-use principle aiming to ease the field work for coaches and to remove burden to athletes. The setup consists of only two cameras collecting videos that are then elaborated by two AI based algorithms: the human body skeletonization (executed by Mediapipe) and the ball position tracking performed by a custom-made vision system based on YOLOv8 (Terven & Cordova-Esparza, 2023).

METHODS: Ten male participants (23.6 ± 4.8 years, 172.2 ± 3.6 cm, 60.3 ± 10.3 kg) were recruited for the study. Prior to participation, all individuals provided a consent form approved by the Research Ethics Committee of the Faculty of Design at Kyushu University (approval number: 541). All participants were provided with the same table tennis racket, the Butterfly Stayer 2000 (Tamasu Co., Ltd., Japan), and used standard 40 mm Tigora table tennis training balls. Participants were tasked with successfully returning the ball to the opposite court. They were allowed to choose between forehand and backhand strokes for their returns. Each participant was required to complete 10 sets, with each set comprising 12 ball serves delivered by a table tennis training machine (CTR-18S, Sakurai Co., Ltd., Japan). The participants performed a synchronization movement waving their arm at the beginning and the end of every recording session to facilitate the synchronization of the two videos using cross-correlation. The system was conceived to work in a gym environment. The tests were conducted in normal training conditions without any specific light condition to facilitate ball recognition or body tracking. The only requirement for the camera is to frame the face of the participant to facilitate Mediapipe in the skeletonization phase. Two *GoPro Hero 10* cameras, collecting frames at 240 Hz, were placed in the front and on the right side of the participants at approximately 1 m.
A custom-made algorithm, briefly illustrated in Figure 1, was used to measure the joint angles, angular speeds and accelerations and to correlate them to the moment of impact between racket and ball. First of all, the position of the ball was tracked using the videos collected by the camera on the right side of the participant. The xy position of the ball were obtained by a custom-made vision system based on YOLOv8 (Terven & Cordova-Esparza, 2023) architecture. The x-coordinate was used to define a linear polygonal chain using Windowed Linear Least Squares Method. A vertex of the chain (where the slope inverts) indicated that the ball changed direction. This moment was considered as the moment of impact with the racket. Secondly, the human body skeletonization was obtained using the videos collected by both the cameras. The positions of the landmarks in real world coordinates obtained by Mediapipe were moved in a new reference frame, as shown in Figure 2, placed in the middle hip of the player, with v-axis pointing upward, w-axis pointing to the right and u-axis on the common line perpendicular to the v-axis and w-axis, as suggested in (Wu et al., 2005). It was consequently possible to compute the angles listed in Table I for the data obtained by the camera placed on the front and on the right side of the participant. When the distance between the landmarks defining the angle decreased, the angular value was different from the expected one. Consequently, the accuracy of the angular values was sensitive to the orientation and the distance of the player with respect to the camera, which changed during the execution of the movements. Therefore, the final angular value was obtained using a weighted average. The weights were computed as the minimum of the distances (expressed in pixels) between the two couples of landmarks defining the angle, as shown in the equation that follows.

$$w = \min\left(\sqrt{(x_a - x_b)^2 + (y_a - y_b)^2}, \sqrt{(x_b - x_c)^2 + (y_b - y_c)^2}\right)$$

Once computed the joint angular position, the signals were averaged over time with a sliding window of amplitude 0.2 s. Next, intraclass correlation coefficient (ICC) (Fisher, 1925) was used to assess the consistency between parameters computed

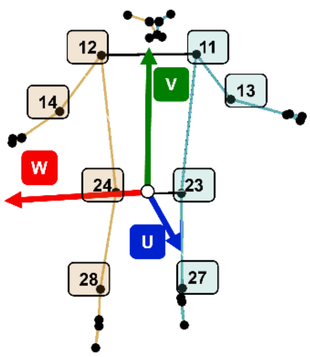

Figure 2: Reference frame placed in the middle hip. In brown the right side of the participant, in green the left side of the participant

Table I: Angles computed with projection plane and landmarks used. "Landmarks right" and "Landmarks left" refer to the landmarks placed on the right and left side of human body

Projection Plane	Landmarks right	Landmarks left	Refers to
UV	24-12-14	23-11-13	angle between arm and torso on sagittal plane
VW	24-12-14	23-11-13	angle between arm and torso on frontal plane
UW	11-12-14	12-11-13	angle between arm and torso on transverse plane
UVW	12-14-16	11-13-15	elbow flexion extension
UV	12-24-26	11-23-25	angle between leg and torso on sagittal plane
VW	12-24-26	11-23-25	angle between leg and torso on frontal plane
UVW	24-26-28	23-25-27	knee flexion extension

by front camera and right side camera. This evaluation was crucial for determining the accuracy of angle representation and assessing the necessity of implementing a weighted average for improved precision.

RESULTS and DISCUSSION: Since the current stage of the project only aimed to assess the feasibility of the system and not its validation, no ground truth data was collected. The ICC values, computed using all the data points available in the 100 tests (10 tests for 10 participants) for a total of 1366894 data points, are shown in Table II. Comparing these values for the right and left sides of the body, the ICC is higher for the right side of the body. This can be justified considering that the cameras were placed in front and on the right side of the participant and not on the left side, therefore the left side angles were more difficult to be detected. Values of ICC lower the 0.81 suggest that the use of the weighted average is necessary. Figure 3 shows the right shoulder angle when executing five times a flexion extension in the sagittal plane for the full range of motion. In the first three times (from 55 s to 60 s), the participant was looking at the front camera, showing his side to the camera placed on his right (as shown in the two pictures on the left). The low peaks of the side weights (green line in the bottom graph, highlighted in green) happened when the arm-torso angle is close to 90°, namely the moment in which the landmarks of the elbow and the shoulder are seen superimposed by the camera placed on the side. In the last two times (from 61 s to 64 s) the participant was in an intermediate position, as shown in the two pictures on the right. Consequently, the peaks are present for both the weights signal. Lastly, in the yellow zoom, it is possible to appreciate the effects of the weighted average: since the weight for the side camera (in green in the bottom graph) is higher, the average angular value is closer to the one of the side camera (in green inside the yellow zoom).

CONCLUSION: The proposed solution, not requiring any equipment worn by the participant, allows the implementation of an easy-to-use setup, reducing the field work for coaches and the burden to athletes. In practical applications, the current system could be used for studies such as those presented in (Rigozzi et al., 2023), where the focus is towards the joint movement patterns and the movement differences associated with different ball spin levels. The use of

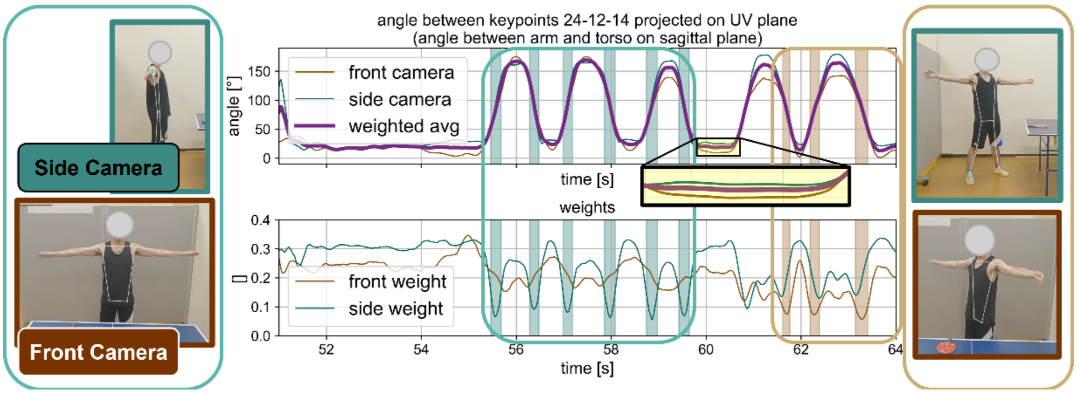

Figure 3: Comparison and weights of flexion extension angle of the right shoulder in the sagittal plane when executing five times the movement for the full range of motion

Table II: Intraclass correlation coefficients according to right and left side of the participant

Refers to	ICC [Right]	ICC [Left]
angle between arm and torso on sagittal plane	0.65	0.63
angle between arm and torso on frontal plane	0.68	0.54
angle between arm and torso on transverse plane	0.19	-0.04
elbow flexion extension	0.59	0.54
angle between leg and torso on sagittal plane	0.46	0.21
angle between leg and torso on frontal plane	0.23	0.13
knee flexion extension	-0.33	-0.49

two cameras allows to compute a weighted average of the angles. The next step of the study is the metrological validation of the measured angles. Investigating the addition of cameras placed either on top of the participant or on its left to enhance accuracy might be interesting. Lastly, Mediapipe Pose provides a “visibility” parameter associated with each joint, that might be used in the computation of the weights.

REFERENCES

Bańkosz, Z., & Winiarski, S. (2017). The kinematics of table tennis racquet: Differences between topspin strokes. *Journal of Sports Medicine and Physical Fitness*, *57*(3), 202–213. https://doi.org/10.23736/S0022-4707.16.06104-1

Bańkosz, Z., & Winiarski, S. (2020). Using Wearable Inertial Sensors to Estimate Kinematic Parameters and Variability in the Table Tennis Topspin Forehand Stroke. *Applied Bionics and Biomechanics*, *2020*. https://doi.org/10.1155/2020/8413948

Bazarevsky, V., Grishchenko, I., Raveendran, K., Zhu, T., Zhang, F., & Grundmann, M. (2020). *BlazePose: On-device Real-time Body Pose tracking*. http://arxiv.org/abs/2006.10204

Chénier, F. (2021). Kinetics Toolkit: An Open-Source Python Package to Facilitate Research in Biomechanics. *Journal of Open Source Software*, *6*(66), 3714. https://doi.org/10.21105/joss.03714

Fisher, R. A. (1925). *STATISTICAL METHODS FOR RESEARCH WORKERS*.

Keskinoğlu, C., Özgünen, K. T., & Aydın, A. (2023). Designing and implementation of IMU-based wearable real-time jump meter for vertical jump height measurement. *Transactions of the Institute of Measurement and Control*, *45*(14), 2648–2657. https://doi.org/10.1177/01423312231158857

Latreche, A., Kelaiaia, R., Chemori, A., & Kerboua, A. (2023). Reliability and validity analysis of MediaPipe-based measurement system for some human rehabilitation motions. *Measurement: Journal of the International Measurement Confederation*, *214*. https://doi.org/10.1016/j.measurement.2023.112826

Rigozzi, C. J., Vio, G. A., & Poronnik, P. (2023). Application of wearable technologies for player motion analysis in racket sports: A systematic review. *International Journal of Sports Science and Coaching*, *18*(6), 2321–2346. https://doi.org/10.1177/17479541221138015

Terven, J., & Cordova-Esparza, D. (2023). *A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS*. https://doi.org/10.3390/make5040083

Wong, D. W. C., Lee, W. C. C., & Lam, W. K. (2020). Biomechanics of table tennis: A systematic scoping review of playing levels and maneuvers. In *Applied Sciences (Switzerland)* (Vol. 10, Issue 15). MDPI AG. https://doi.org/10.3390/app10155203

Wu, G., Van Der Helm, F. C. T., Veeger, H. E. J., Makhsous, M., Van Roy, P., Anglin, C., Nagels, J., Karduna, A. R., McQuade, K., Wang, X., Werner, F. W., & Buchholz, B. (2005). ISB recommendation on definitions of joint coordinate systems of various joints for the reporting of human joint motion - Part II: Shoulder, elbow, wrist and hand. *Journal of Biomechanics*, *38*(5), 981–992. https://doi.org/10.1016/j.jbiomech.2004.05.042

Xia, R., Dai, B., Fu, W., Gu, N., & Wu, Y. (2020). Kinematic Comparisons of the Shakehand and Penhold Grips in Table Tennis Forehand and Backhand Strokes when Returning Topspin and Backspin Balls. *Journal of Sports Science and Medicine*, *19*, 637–644. https://pubmed.ncbi.nlm.nih.gov/33239936/

Xu, J. (2024). Application of optical motion capture device based on android intelligent platform in sports field auxiliary recognition system. *Optical and Quantum Electronics*, *56*(3). https://doi.org/10.1007/s11082-023-05878-3

ACKNOWLEDGEMENTS: The authors want to thank the creator of the Python package “Kinetics Toolkit” (Chénier, 2021), which was used to create Figure 2.