Tennis Action Evaluation Model Based on Weighted Counter Clockwise Rotation Angle Similarity Measurement Method

Danni Jiang[1], Ge Liu[2]*
Department of Physical Education, Jinling Institute of Technology, Nanjing, China[1]
Department of Physical Education, Zhongnan University of Economics and Law, Wuhan, China[2]

***Abstract*—In order to intelligently analyze tennis movements and improve evaluation efficiency, a counter clockwise rotation angle of limbs is proposed to solve the direction problem of tennis movements. A dynamic time regularization algorithm is optimized by combining global time weighting and adjacent frame weighting. The results indicated that the proposed counter clockwise rotation angle feature of limbs could effectively represent changes in limb direction and clearly distinguish action postures. The average accuracy of this method in action classification on the Tennis Stroke Dataset was 97.60%. In the action evaluation mode, the average frame rate of the client was between 17.35FPS and 17.49FPS, and the overall average frame rate was about 17.40FPS. The server exhibits higher efficiency in action processing and evaluation, which can process video frames faster. It is more efficient in processing data capabilities and utilizing data resources. This indicates that the performance of the system is relatively consistent in different modes and has stability. The optimized method has a higher generalization ability in recognizing non-tennis movements on different datasets. When dealing with fine movements, the optimized method performs excellently and can better capture subtle differences in the movements. Meanwhile, this enhances the real-time performance of the system, making it suitable for evaluating tennis movements in practical application scenarios. This provides a new technical path for analyzing tennis movements and also serves as a reference for evaluating movements in other sports.**

***Keywords*—Action evaluation; counter clockwise rotation angle; weighting; dynamic time warping; tennis**

I. INTRODUCTION

With the rapid development of artificial intelligence, the combination of computer technology and tennis has become one of the hot topics in the field of sports. As a popular competitive sport, the analysis and evaluation of tennis technique movements are of great significance for athlete training and competition [1]. The Tennis Movement Evaluation System (TME) helps to obtain more accurate match and training data, which can improve athletes' performance and training efficiency, and more scientifically analyze the movement posture during tennis sports [2]. The data representation in tennis videos is in the form of time series. For time series processing, a common task is to compare the similarity between two sequences. Comparing the similarity of time series is more conducive to identifying patterns and trends, which is of great significance for discovering patterns in action data and predicting future behavior. As one of the most important similarity measurement methods in time series analysis, Dynamic Time Warping (DTW) is the process of elongating or shortening unknown variables until they match the length of the reference template. The time axis of unknown data is distorted or bent, so that its feature quantities correspond to the standard pattern. However, traditional action evaluation methods often rely on manual observation and analysis, which have problems such as low efficiency and strong subjectivity [3].

Regarding the evaluation and classification of tennis movements, many researchers have adopted different algorithms and techniques for in-depth optimization, and have achieved certain results. To analyze the performance of the tennis evaluation platform, Wu et al. collected data through wearable devices and selected the Z-score normalized Support Vector Machine (SVM) to classify hitting actions. The accuracy reached 98.4% [4]. Giles et al. proposed a hierarchical clustering method and tennis directional change technique to distinguish tennis movement styles, identify movement features, and analyze temporal movement characteristics such as change speed and directionality. The results showed that this method was feasible [5]. Perri et al. considered the hitting situation in tennis training and used wearable devices and prototype learning to detect different movements to determine exercise load. The results showed that this technology had high accuracy [6]. Wood et al. used Krippendorffs alpha analysis to evaluate the reliability of tennis serve features in 2D videos. This method had high measurement accuracy [7]. Liu et al. extracted different features through the acceleration of the action and the deep mode data. The spatial and temporal convolutional neural network were combined to realize the specific action recognition. The results showed that the accuracy of this method was more than 99% [8]. In order to analyze the trajectory of tennis serve, Hu et al. combined SVM with frame difference technology and median filtering algorithm to locate targets and recognize trajectories. The results showed that the classification accuracy was 97.5% [9]. Perri et al. used wearable GPS devices to record the training serving load for tennis serving information recording. The results showed that this method had good detection performance [10]. Setyawan et al. evaluated the serving movements of athletes of different genders to improve the success rate of serving and evaluate the

differences in tennis serving performance. The results showed that this method was effective [11].

Some scholars have attempted to improve the DTW method in identifying abnormal charging, time series analysis, and image change detection, and have achieved good results. Shuai et al. developed a DTW model with the longest similar substring to identify abnormal charging situations and avoid excessive regularization of DTW for the safety of electric bicycle charging. The average recognition accuracy reached 94% [12]. Deriso et al. proposed a method that combined DTW and iterative refinement techniques to address the trade-off between time regularization characteristics in traditional DTW signal alignment errors. The results showed that this method was feasible [13]. Zhang et al. designed a fast DTW and sequence decomposition method to analyze different components in time series. The time series was decomposed and the similarity between different components was measured. The results showed that this method had high classification accuracy [14]. Xing et al. detected changes in satellite image time series using remote sensing image time series values and DTW to calculate the change amplitude map. The change results were detected in advance. The accuracy was improved by up to 5.10% [15]. Froese et al. proposed two run length encoding time series to improve DTW calculation speed, which shortened the running time and reduced the factors affecting time. The results shows that this method was effective [16]. Kumawat et al. developed a new framework for DTW and adversarial training to enhance the robustness of deep neural networks. Different adversarial examples were created and random alignment paths were implemented. The framework had high efficiency [17]. He et al. proposed Anticor to improve DTW similarity calculation to identify differences between different sequences. The results showed that the algorithm has good practicality [18]. Vorpe et al. introduced non-parametric DTW to capture the leading and trailing relationships between time series for obesity rate prediction. The results showed that this method had good predictive performance [19].

In summary, both the evaluation of tennis movements and the improvement of DTW have shown high accuracy and effectiveness. However, due to the directional information involved in tennis movements and the large computational complexity of DTW, an innovative TME system is developed for this purpose. The feature of Counter Clockwise Rotation Angle (CRA) of limbs is used to quantify and analyze action direction, and global time weighting and adjacent frame weighting are selected to optimize DTW. The research aims to improve computational efficiency and enhance the real-time performance of the system, thereby providing technical support for the field of tennis movement analysis. This study consists of four main parts. Section II is methods and materials. Among them, Part A introduces the physical CRA of tennis actions. The content of Part B is WCRA between adjacent frameworks and global time. Section III is the results of the study. Among them, Part A analyzes the TME system based on WCRA similarity measurement method. Part B introduces the performance analysis of the TME system. Section IV is discussion and conclusion.

II. Methods and Materials

A. Physical CRA of Tennis Actions

Tennis action videos are captured using cameras that can only capture images from one perspective, and these videos are processed and evaluated [20-21]. Considering that tennis courts include both indoor and outdoor environments, factors such as color and lighting affect video images. To cope with environmental impacts, video images need to be preprocessed. This includes adjusting the brightness of the image and performing noise reduction to improve image quality and ensure accuracy in subsequent processing. Monocular cameras can only capture channel information for the red, green, and blue colors of an image, but do not include distance information. Therefore, Two-Dimensional pose estimation is used to identify human skeletal joint points in images, that is, to determine the position information of each joint point. Each joint not only contains positional information, i.e. coordinates, but also semantic information, such as which joint is the shoulder and which is the knee. Simultaneously, the coordinates of joint points are extracted as the basis for subsequent action evaluation [22]. Because tennis movements are dynamic, there may be some joint points obscured in certain frames of the video, which can affect the accuracy of pose estimation. At this point, if the coordinates of the undetected joint point are set as the origin $O(0,0)$, the horizontal and vertical coordinate pair p_i for the i-th joint point is displayed in Eq. (1).

$$p_i(x,y)=\begin{cases} p_i(x,y), & \text{if } p_i \text{ detected} \\ O(0,0), & \text{if } p_i \text{ not detected} \end{cases} \quad (1)$$

Due to significant differences in body shape among individuals, even if two people perform the same action, the overlap rate of their contours may be low, resulting in inaccurate similarity calculation. However, in 2D skeleton nodes, only considering the angle between limbs has certain limitations. Because when the angle between two adjacent limbs formed by three adjacent joints is the same, i.e. $\theta_1=\theta_2$, the actual movement posture is not the same, which cannot accurately describe the movement characteristics [23-24]. The angle and CRA of the tennis movement limbs are shown in Fig. 1.

(a) Action 1 (b) Action 2

Fig. 1. The angle between tennis limbs and CRA.

In the TME system, to overcome the shortcomings of traditional contour overlap rate and simple angle analysis, this study aims to more accurately describe and evaluate human movements by combining directional information of limb angles. The CRA of the right wrist-right elbow-right shoulder in Fig. 1 (a) is denoted as α_1. The CRA of the right wrist-right

elbow-right shoulder in Fig. 1 (b) is denoted as α_2. The difference between angles α_1 and α_2 is obvious and will not be confused, so the CRA is used to more accurately describe the angle and human posture of the angle. The angle value range of CRA is between 0° and 360°. According to the cosine theorem, the angle between two vectors is calculated. Then, the vector product is used to determine their positional relationship [25]. The tennis movements exclude detailed changes in the head. The distribution of joint points and CRA in the limbs is shown in Fig. 2.

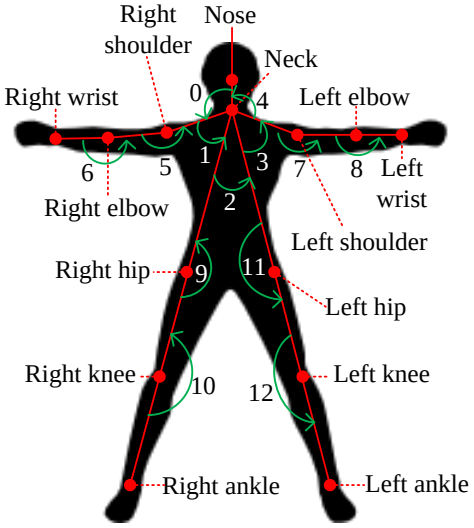


Fig. 2. Schematic diagram of limb joint points and CRA distribution.

In Fig. 2, there are 14 joint points in the limbs. In the template video, OpenPose is used to detect the human joint points of each frame image of two random tennis movements, obtaining 13 limb CRAs. The function of OpenPose is real-time multi person pose estimation and keypoint detection, which can detect the 2D poses of multiple people in images or videos in real time. It is suitable for single person and multi-person scenes and has excellent robustness. The input image or video uses a pre-trained model to identify key points in the human body, including the head, shoulders, elbows, wrists, hips, knees, and other key positions. By identifying and connecting these key points, OpenPose can generate a complete multi-person pose estimation result, thereby recognizing human actions [26-27]. The definition of partial limb CRA is displayed in Table I.

TABLE I. DEFINITION OF LIMB CRA

RA number	Joint number representation	CRA
0	0-1-2	Nose-neck-right shoulder
1	2-1-8	Left shoulder-neck-nose
2	8-1-11	Right Shoulder-neck-right Hip
3	11-1-5	Right Hip-neck-left Hip
4	5-1-0	Left Hip-neck-left Shoulder
5	3-2-1	Right elbow-right shoulder-neck

B. WCRA Between Adjacent Frames and Global Time

In tennis, athletes' movements and scenes change very quickly. Single frame real-time evaluation can quickly adapt to these changes, update evaluation results in a timely manner, and ensure accurate evaluation and feedback in dynamic scenes. Meanwhile, real-time processing of single frame data requires relatively less computation and can effectively utilize computing resources. Compared with processing the entire video stream, single frame evaluation can distribute the computational burden, avoid delays and lags, and ensure the real-time and smooth performance of the system [28]. The single frame evaluation is displayed in Fig. 3.

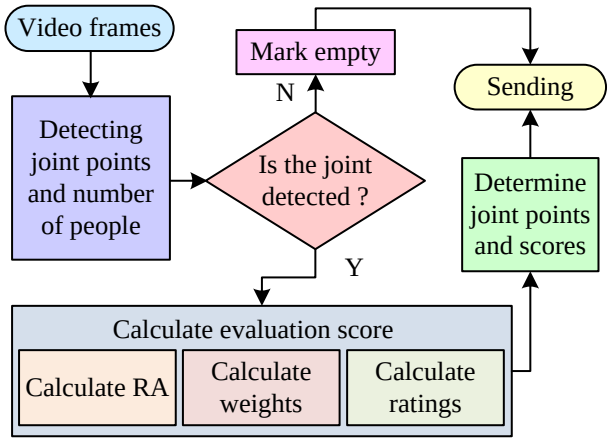


Fig. 3. Flowchart of single frame evaluation.

In Fig. 3, after inputting video frames with a resolution of 656×368, joint points and number of people are detected. If a joint is detected, the evaluation score is calculated from the Rotation Angle (RA), weight, and score to determine the joint points and score. If no joints are detected, it is marked as blank. To facilitate the evaluation of tennis movements in single frame videos, the frame rate of the template video and the input video are set to be the same [29]. The relationship between RA in various limbs of the human body is analyzed to obtain evaluation scores for each frame. The similarity score s_t of t-th frame has a value range of [0, 1], as shown in Eq. (2).

$$s_t = 1 - \frac{\sum_{i=1}^{n} W_i \left| \Delta \alpha_i^t \right|}{C} \tag{2}$$

In Eq. (2), n is the number of RAs. $\Delta \alpha_i^t$ is the difference between the t-th frame input video and the i-th RA of template video. C is the maximum value of RA. When s_t, it indicates that the actions of the template and the input video are completely consistent. When s_t is 0, it indicates that the actions of the template and the input video are completely different. The weight W_i for the i-th RA is expressed as Eq. (3).

$$W_i = aW_i^{\sigma} + bW_i^{\tau} \tag{3}$$

In Eq. (3), a and b are both parameters, with a value of 0.5. W_i^{σ} is the inter frame RA weight of i-th RA input video

actions. W_i^{τ} is the global time RA weight of the template video action for the i -th RA. To analyze TME more comprehensively, the study considers the action situation of the input video. Adjacent Frame Weighting (AFW) focuses on the change details between adjacent frames in the action sequence, which can capture small changes in the action and enhance the matching precision [30-31]. During the DTW process, weight allocation is performed on adjacent frame pairs. AFW considers the changes and interrelationships between adjacent frames, which is suitable for capturing local dynamic changes in sequences. In TME, subtle movement changes can also affect technical evaluation, and AFW can improve the sensitivity of DTW in details. AFW needs to first calculate the cumulative change in limb RA, with a required range between adjacent frames of the current frame, in order to obtain the RA weights between adjacent frames. W_i^{σ} is shown in Eq. (4).

$$W_i^{\sigma} = \begin{cases} \dfrac{\theta_i^{'} + \rho_i^{\sigma}}{\sum_{i=1}^{n} \theta_i^{'} + n\rho_i^{\sigma}} & 0 < \Delta T < t_c \\ 0 & \Delta T \geq t_c \end{cases} \tag{4}$$

In Eq. (4), t_c is the current frame. ΔT is between adjacent frames, and $\Delta T \in (0, t_c)$. $\theta_i^{'}$ signifies the ΔT cumulative change of the i -th RA in the input video from the previous t_c. ρ_i^{σ} is the smoothing term, as shown in Eq. (5).

$$\rho_i^{\sigma} = \sum_{t=t_c-\Delta T}^{t_c} \alpha_i^t \big/ \Delta T + 1 \tag{5}$$

$\theta_i^{'}$ is shown in Eq. (6).

$$\theta_i^{'} = \sum_{t=t_c-\Delta T-1}^{t_c-1} \left| \alpha_i^{t+1} - \alpha_i^t \right| \tag{6}$$

Global Time Weighting (GTW) is used to allocate reasonable weights to the entire action sequence in the time dimension, ensuring that each stage of the entire action process receives appropriate attention. It considers the temporal characteristics of the entire sequence and assigns higher weights to certain key moments or key action points. GTW can balance the importance of different time periods in action sequences, making the DTW algorithm more accurate in matching actions and avoiding certain important time periods from being ignored or underestimated. W_i^{τ} is shown in Eq. (7).

$$W_i^{\tau} = \frac{\theta_i + \rho_i^{\tau}}{\sum_{i=1}^{n} \theta_i + n\rho_i^{\tau}} \tag{7}$$

In Eq. (7), ρ_i^{τ} is the smoothing term, which serves to avoid a weight of 0. If the weight is 0, the corresponding RA information will be ignored, affecting the result. ρ_i^{τ} is shown in Eq. (8).

$$\rho_i^{\tau} = \sum_{t=1}^{T} \alpha_i^t \big/ T \tag{8}$$

In Eq. (8), T signifies the frame number of the template video. θ_i is the degree of change of a specific RA in the time series in the template video. By calculating θ_i, the dynamic variation characteristics of the angle during the action process can be captured. If a certain RA changes dramatically between different frames, θ_i will be larger, indicating that the action point has high dynamism in the entire action sequence and may need to be given higher weight. θ_i is shown in Eq. (9).

$$\theta_i = \sum_{t=1}^{T-1} \left| \alpha_i^{t+1} - \alpha_i^t \right| \tag{9}$$

To better reflect the true s_t, AFW and GTW are applied to the data. The overall evaluation score S for a certain tennis action is shown in Eq. (10).

$$S = \sum_{t=1}^{T} (w_t \cdot s_t), \sum_{t=1}^{T} w_t = 1 \tag{10}$$

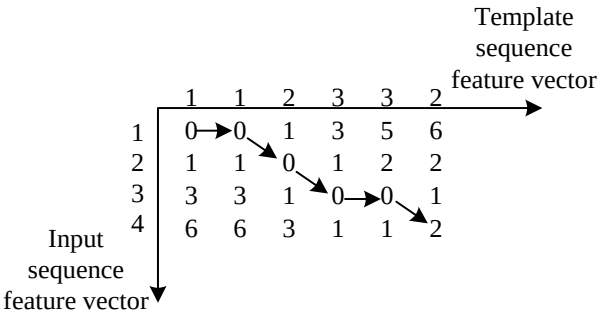
In Eq. (10), w_t is the weighted score, and $w_t \in (0,1)$. AFW focuses more on local changes, emphasizes the dynamic relationship between adjacent frames, and improves the precision and smoothness of matching. GTW focuses more on the temporal characteristics of the entire sequence, emphasizing the matching of critical moments to improve the accuracy and robustness of overall evaluation. Combining these two weighting methods can capture the importance of global key moments in action evaluation while not ignoring the details of local dynamic changes, achieving higher evaluation accuracy and robustness.

C. TME System Based on SDTW

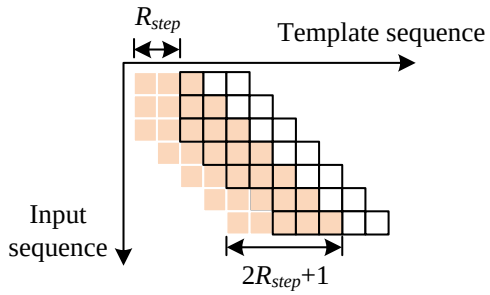
There are multiple similarity or distance functions in time series data in videos. DTW is used to calculate the similarity between two time series. DTW minimizes the cumulative distance between one time series and another by calculating the nonlinear mapping relationship between the two. DTW may not be able to output results in a timely manner, as it requires traversing a large amount of frame data to find the path with the minimum cumulative distance. To solve the real-time of the DTW algorithm, an improved algorithm called Segmented DTW (SDTW) is proposed. The core idea of SDTW is to segment the actions in the template video and perform path search within each segment. This method reduces time complexity and improves computational efficiency by limiting the search range. By segmenting and reducing computational complexity, SDTW can output calculation results in a timely manner at the end of frame processing in action videos, thus meeting real-time requirements. The schematic diagram of path search for DTW and the search area for SDTW with a step size of $R_{step} = 2$ are shown in Fig. 4.

In Fig. 4 (a), a 2D distance matrix D is constructed, with the total frame numbers of the input video and template video being n and m, respectively. The matrix element $d_{(i,j)}$ at the current position (i, j) is displayed in Eq. (11).

$$d_{(i,j)} = d_{\min} + \sqrt{\sum_{k=1}^{n} (\Delta \alpha_k)^2} \tag{11}$$



(a) DTW path search diagram



(b) SDTW's search area

Fig. 4. DTW path search and SDTW search area.

In Eq. (11), $d_{\min}$ signifies the minimum distance value obtained in the previous step. $\Delta\alpha_k$ is the difference between the k-th CRA of the input video and the template video. At the initial position (1,1), $d_{\min}$ is 0. At different positions, $d_{\min}$ is shown in Eq. (12).

$$d_{\min} = \min\left(d_{(i-1,j)}, d_{(i,j-1)}, d_{(i-1,j-1)}\right) \tag{12}$$

In Eq. (12), $(i-1,j)$ and $(i,j-1)$ are the adjacent positions on the upper and left sides of (i,j), respectively. $(i-1,j-1)$ is the upper left corner position of (i,j). DTW first starts from the initial position $(1,1)$ of the time series matrix, gradually accumulates and calculates the distance of each step by moving to the right, down, and right adjacent positions in the matrix, and finds the path in this process to minimize the total accumulated distance. Finally, the DTW algorithm uses this path to find the minimum cumulative distance at the endpoint position (n,m) of the matrix, which represents the minimum similarity distance between two time series. A small distance indicates that the two sequences are more similar in the time dimension, demonstrating that the shapes or patterns of the two sequences are closer. In Fig. 4 (b), when the template video time series is segmented, the step size is set to R_{step}. To better limit the range of path search, the width on both sides of the diagonal is specified to ensure that the search path is concentrated near the diagonal rather than spreading throughout the entire matrix R_{step}. In the case of segmented processing, the total width range of path search is determined by the step size R_{step}, with a maximum width of $2R_{step}+1$ and a minimum width of $R_{step}+1$. This means that the path search range will be limited to the diagonal and its left and right range R_{step}, thereby reducing computation and complexity. The endpoint position of the first segment of SDTW is (n, j_e), and the vertical axis satisfies $j_e \in \left(m+1-R_{step}, m+1+R_{step}\right)$. At this time, the search range of the path is on the diagonal, with a width of $2R_{step}+1$. The computational workload is reduced to avoid global search. Subsequently, by moving R_{step} steps each time, different path searches can be achieved. The average distance $\bar{d}$ for each step is shown in Eq. (13).

$$\bar{d} = \frac{c}{\kappa} \tag{13}$$

In Eq. (13), c and κ are the distance and length of the path. The conversion score s_d is shown in Eq. (14).

$$s_d = \frac{1}{1+h\bar{d}} \tag{14}$$

In Eq. (14), h is the coefficient that adjusts the magnitude of the score decrease, and $h = 0.25\kappa$. As $\bar{d}$ increases, the similarity between two time series decreases. At this time, the result of s_d is closer to 0. Two time series are completely identical, and $s_d = 1$. To evaluate human movements in a single frame video, it is necessary to ensure that the input video and the template video have equal frame rates. Analyzing the relationship between RA in various limbs of the human body can obtain a single frame calculation evaluation score. In practical operation, extracting human joint information is a time-consuming process and may cause delays during network transmission. To alleviate these issues, an offline processing module is added to the TME system. The function of this module is to process the template video in advance, extract and save the joint coordinates of each frame to the local file. Therefore, in the actual real-time evaluation process, the system only needs to read the stored joint information from the local file without recalculating and extracting, greatly reducing the time required for evaluation and improving efficiency. Therefore, the constructed architecture diagram of the TME system is shown in Fig. 5.

In Fig. 5, the client and server of the TME system use Socket to achieve data transmission. The client is connected to the monocular camera through USB, and can display a video interface and collect video data. Meanwhile, customers can preprocess video frame data and then transmit it to the server. The focus of the server is on extracting joint points and evaluating actions. The monocular camera is transmitted to the server via USB connection, and the server receives the data information and transmits the results to the client.

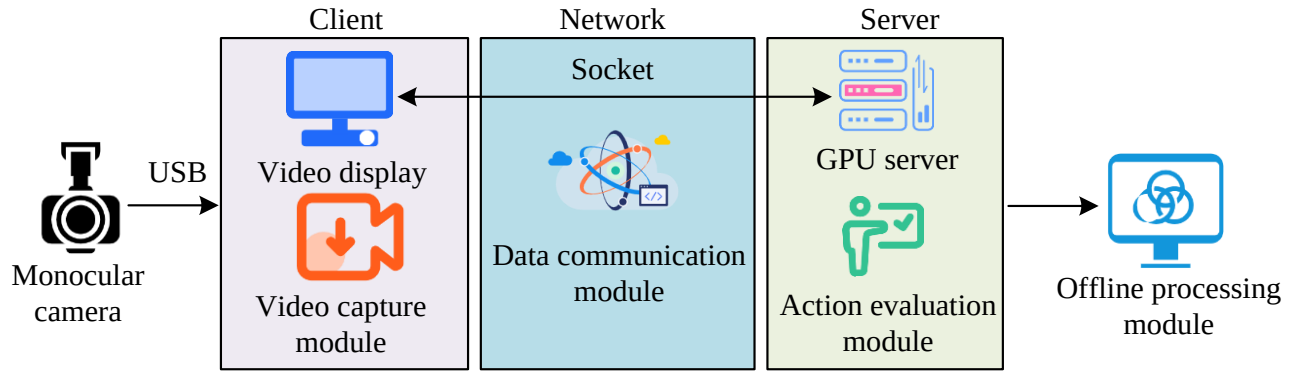


Fig. 5. TME system architecture diagram.

III. Results

A. Analysis of TME System Based on WCRA Similarity Measurement Method

The experimental datasets are the Tennis Videos Dataset and the Tennis Stroke Dataset, with the former containing a large number of tennis matches and training videos, typically used for video analysis, action recognition, and strategy analysis. The latter includes data on various tennis hitting movements, such as smashes, Forehand Strokes (FS), Backhand Strokes (BS), Forehand Cut (FC), and Backhand Cut (BC). The experimental setup is displayed in Table II.

TABLE II. Hardware Equipment and Software Environment for Experimental Setup

Hardware device	Device type	Software environment	Configuration information
GPU server	NVIDIA Tesla P40	Operating system	Ubuntu 14.04.5 LTS
GPU architecture	NVIDIA Pascal	Deep learning framework	Caffe
High definition digital camera	SONY HDR-CX405	GPU parallel computing architecture	CUDA 8.0
Monocular camera	RER-USBFHD01M	Computer vision library	OpenCV 2.4.13

The research indicators include similarity evaluation scores and accuracy. The similarity evaluation score is used to assess the similarity between actions of the same type (such as multiple backhands) and actions of different types (such as backhands and smashes). By calculating and comparing these scores, the similarity between different action types before and after improvement can be analyzed to evaluate the accuracy and effectiveness of the algorithm. Accuracy is used to measure the performance of algorithms in identifying and classifying action types, such as kill actions. By comparing the accuracy and average accuracy of different methods, the effectiveness and reliability of each method can be evaluated under different levels of action performance. To evaluate the impact of different algorithms on real-time performance when processing monocular camera videos using OpenPose. A monocular camera is conFig.d, with a video frame rate of 60FPS and a resolution of 1280×720. The required algorithm environment is set on the server, including OpenPose. Different algorithms are run sequentially and combined with OpenPose to process video streams, recording the actual frame rate per second during the processing. The experiment is conducted 20 times to ensure reliability and statistical significance, with each experiment lasting 10 minutes. In each experiment, the processed video frame rate and changes in frame rate are analyzed, as shown in Fig. 6.

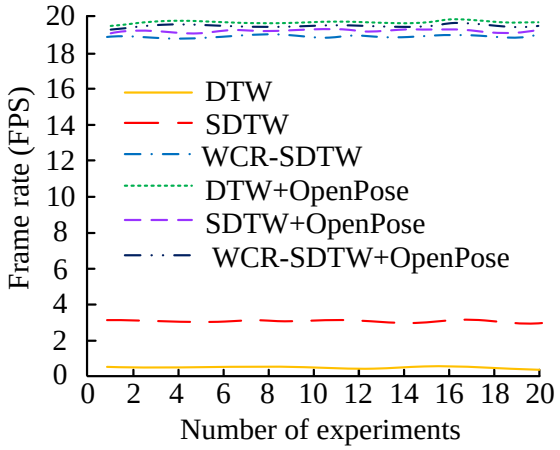


Fig. 6. The frame rate variation of different methods.

In Fig. 6, the average frame rate of DTW without OpenPose processing was the lowest, only 0.5FPS. The average frame rate under the action of WCR and SDTW was19.1FPS, while the average frame rate processed by OpenPose was 19.5FPS, which was time-consuming, but achieved high accuracy. The proposed algorithm, combined with OpenPose, can better maintain the real-time frame rate of the video.

To verify the effectiveness of the designed limb CRA, an analysis is conducted on the changes in CRA values of BS action and smash action, and a comparison between the two was obtained, as shown in Fig. 7.

In Fig. 7 (a), the angle value of CRA 8 fluctuated the most within 30 frames, indicating a more drastic change in the left arm. In Fig. 7 (b), there was a significant fluctuation in the angle value of the right arm CRA 6 during the smash action, ranging from 36° to 260°. Therefore, the proposed limb CRA can effectively represent changes in limb direction, clearly distinguishing between movement and posture situations.

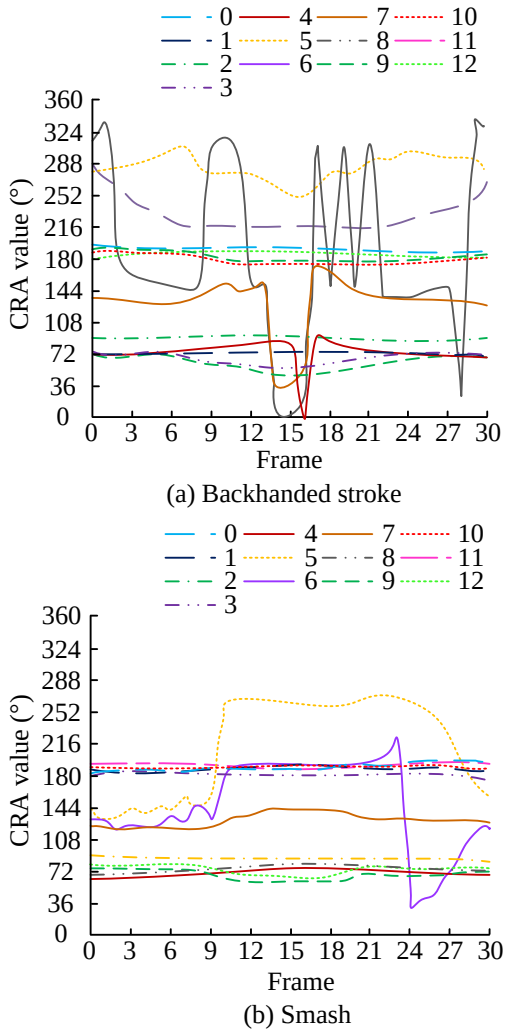


Fig. 7. Comparison of CRA value changes in backhand and smash movements.

To verify the proposed improvement method, the similarity scores between different action types are evaluated to analyze the performance of the improved method. The study randomly selects 250 tennis training videos from the dataset and selects five types of movements: FS, BS, FC, BC, and smash. The similarity scores are analyzed by action type. The similarity scores of same action types and different action types are calculated. The similarity of different action types before and after improvement is compared, as shown in Fig. 8.

In Fig. 8 (a), the similarity scores between actions of the same type and between actions of different types both exceeded 0.75 points. However, according to the similarity score results, the method before improvement was not able to distinguish action similarity scores well. The difference between the maximum and minimum scores for actions of the same type ranged from 0.02 to 0.16 points. In Fig. 8 (b), there was a clear distinction between the five types of actions, with a score difference ranging from 0.27 to 0.49 points. Therefore, the improved method has better discrimination than before.

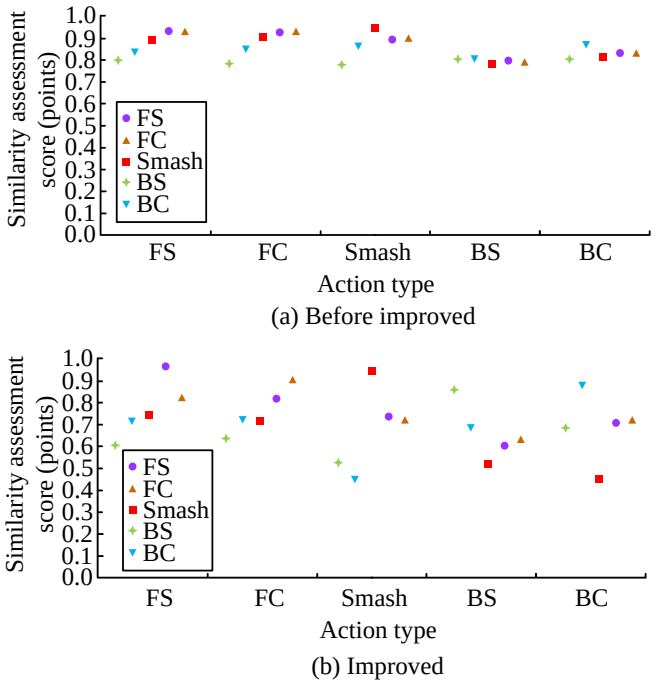


Fig. 8. The similarity score between different action types before and after improvement.

To analyze the accuracy of the improved method, different methods are selected for comparison, including 3D motion capture technology [32], Gaussian Distance-Improved DTW (GD-IDTW) [33], and Joint Angles and Movement Similarity (JA-MS) [34]. The action performance of the dataset is divided into three levels: good, average, and poor. The score for poor action performance ranges from 0.0 to 0.6, the score for average action performance ranges from 0.6 to 0.8, and the score for good action performance ranges from 0.8 to 1.0. The accuracy and average accuracy of different methods of smash actions among the three levels of performance are shown in Fig. 9.

In Fig. 9 (a), the accuracy of the improved method for the smash action was 98.12%, 100.00%, and 100.00% for good, average, and poor performance, respectively. The accuracy of GD-IDTW was relatively close to the improved method, with accuracy rate of 96.97%, 85.77%, and 100.00% for good, average, and poor action performance, respectively. The overall accuracy of the improved method was higher than other methods, because the improved method was significantly better than the GD-IDTW in handling fine actions, which could better capture subtle differences in these actions. In Fig. 9 (b), in the similarity evaluation with good action performance, the improved method had an average accuracy close to GD-IDTW, with 88.15% and 88.01% respectively, showing superior performance. In terms of average action performance, the improved method had an average accuracy of 90.12% in action performance, which was higher than other methods. In the similarity evaluation of poor action performance, the average accuracy of the improved method was 92.55%. The improved method not only enhances computational efficiency, but also improves the accuracy of similarity evaluation.

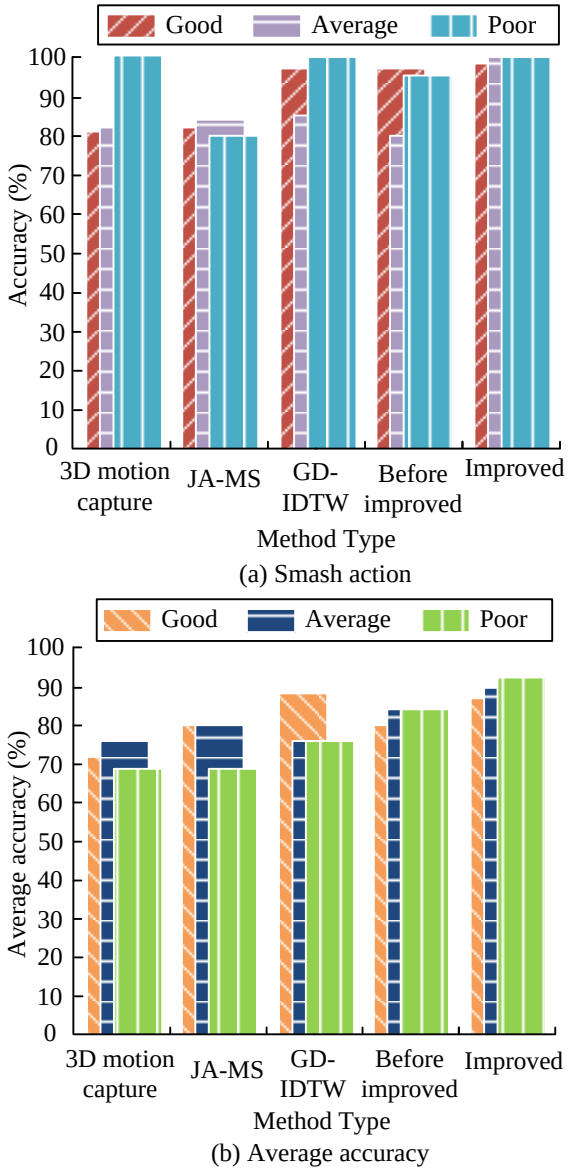


Fig. 9. Comparison of accuracy of different methods.

To verify the accuracy of the improved method for classifying tennis movements, the experiment selects five different types of movements from two datasets, the Tennis Videos Dataset and the Tennis Stroke Dataset, for analysis. The matrix element is the ratio of the number of recognized actions to the number of tested actions. The confusion matrix obtained for the five types of tennis movements is shown in Fig. 10.

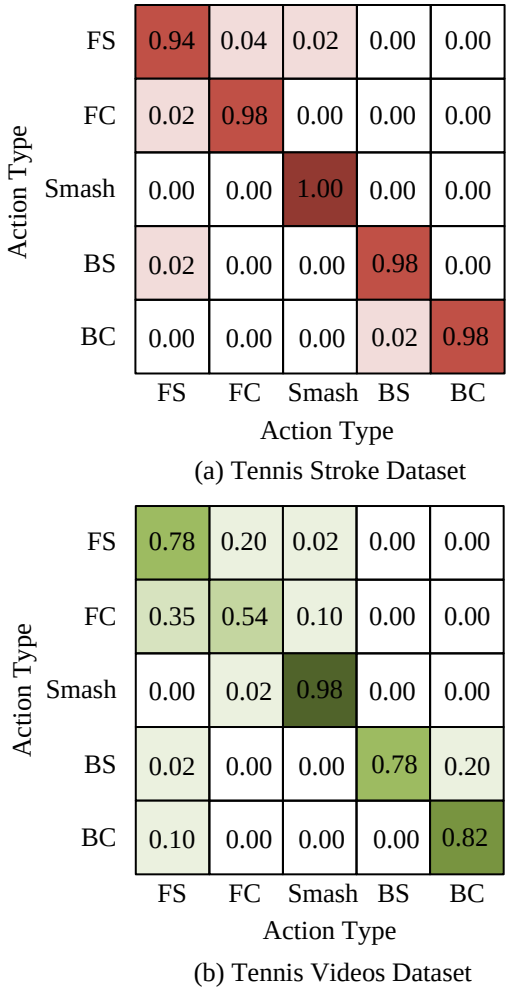


Fig. 10. Comparison of confusion matrices for different datasets.

In Fig. 10 (a), the five columns represent the five recognized action types, and the 5 rows represent the 5 tested action types. The average accuracy of action classification in the Tennis Stroke Dataset was 97.60%. In Fig. 10 (b), there were many classification errors between FS and FC actions in the Tennis Videos Dataset, with an average accuracy of 78.00% for action classification.

To analyze the generalization of the improved method, the study identifies non-tennis movements in two datasets. The number of tests for each type of action is 9, and duplicate videos are filtered out to obtain the recognition accuracy, as displayed in Table III.

TABLE III. RECOGNITION ACCURACY OF NON TENNIS MOVEMENTS

Dataset	Action Type	Bending	Capriole	Running	Walking	High five
Tennis Videos Dataset	Correct number	7	8	7	9	7
	Accuracy	0.78	0.89	0.78	1.00	0.78
Tennis Stroke Dataset	Correct number	6	7	6	8	9
	Accuracy	0.67	0.78	0.67	0.89	1.00

In Table III, for the five movements of bending, Capriole, running, walking, and high five, the average accuracy obtained using the improved method reached 82%. The recognition accuracy of the Tennis Videos Dataset was generally high, with walking movements achieving 100%, and the high five movements of the Tennis Stroke Dataset also achieving 100%. The proposed method has a certain degree of generalization ability in recognizing non-tennis movements on different datasets.

B. Performance Analysis of TME System

Client performance evaluates the real-time performance of video capture and transmission, ensuring that video frames captured by monocular cameras can be transmitted to the server in a timely manner. This includes the frame rate of the camera, data transmission delay, and stability of network transmission. Server performance refers to evaluating the real-time performance of the server in receiving, processing, and analyzing video frames. The server needs to quickly perform algorithm processing and return results after receiving video frames to ensure the overall system response speed. To comprehensively evaluate and optimize the performance of the system, and ensure that the entire TME system can run efficiently and stably in practical applications, real-time testing is conducted on the client and server of the TME system. Five tennis movements are selected for the experiment, and tested five times on both the client and server sides of the system. The average frame rate obtained is presented in Table IV.

In Table IV, the average frame rates of the server in action evaluation and action recognition were 18.52FPS and 18.51FPS, respectively, both higher than those of the client. In the action evaluation mode, the average frame rate of the client was between 17.35FPS and 17.49FPS, and the overall average frame rate was about 17.40FPS. The server exhibited higher efficiency in action processing and evaluation, which could process video frames faster. It is more efficient in processing data capabilities and utilizing data resources. In both modes, there is not much difference in frame rate between the client and server, but the overall trend is consistent. This indicates that the performance of the system is relatively consistent in different modes and has a certain degree of stability.

In the TME system, three individuals are selected for testing. Among the five types of tennis movements, three sets of imitative movements are tested for each movement. The performance levels are good, average, and poor. The similarity evaluation scores for different action types are shown in Fig. 11.

TABLE IV. COMPARISON OF AVERAGE FRAME RATES BETWEEN SERVER AND CLIENT

Mode type	Action Type	FS	FC	Smash	BS	BC	Average frame rate (FPS)
Average frame rate of action evaluation mode (FPS)	Client	17.38	17.35	17.37	17.39	17.49	17.40
	Server	18.50	18.57	18.56	18.43	18.54	18.52
Average frame rate of action recognition mode (FPS)	Client	17.24	17.28	17.38	17.28	17.27	17.29
	Server	18.53	18.51	18.48	18.51	18.49	18.51

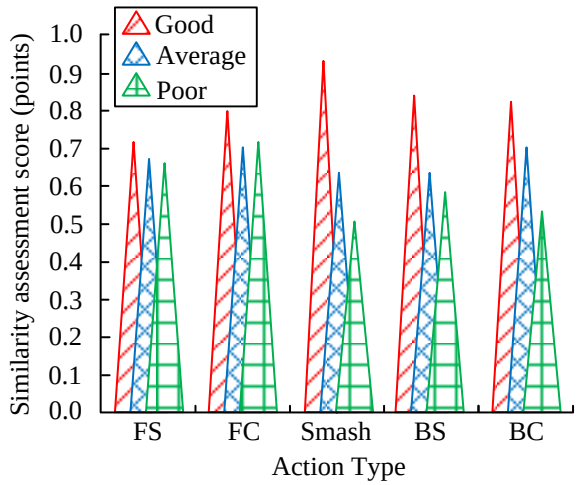


Fig. 11. Similarity evaluation scores for different action types.

In Fig. 11, the similarity evaluation score for the smash action with good performance was the highest, at 0.94, while the similarity evaluation score for this action with poor performance was the lowest, at 0.5. The evaluation scores for smash, BS, and BC actions were relatively high, while the similarity evaluation scores for FS and FC actions were relatively close, resulting in a lower degree of differentiation between the two.

IV. DISCUSSION AND CONCLUSION

There are significant challenges in efficiently and accurately evaluating tennis movements at present. In order to improve computational efficiency and real-time performance of the system, a TME system was developed. The limb CRA was used to quantify and analyze action direction. Then, AFW and GTW weighting were used to optimize DTW. The results showed that the angle value of CRA 8 fluctuated the most within 30 frames, indicating that the changes in the left arm were more severe. The angle value of CRA 6 in the right arm of the smash action fluctuated greatly, ranging from 36° to 260°. The similarity scores between actions of the same type and between actions of different types both exceeded 0.75 points. However, according to the similarity score results, the method before improvement was not able to distinguish action similarity scores well. To address this issue, Zhang et al. focused on evaluating the quality of actions in videos by using distributed autoencoders, likelihood loss sampling scores, and learning uncertainty parameters [35]. The difference between the maximum and minimum scores for actions of the same type ranged from 0.02 to 0.16 points. There was a clear distinction between the five types of actions, with a score difference of 0.27-0.49 points. Therefore, the improved method had better discrimination than before. For the smash action, the accuracy rates of the improved method for good, average, and poor performance were 98.12%, 100.00%, and 100.00%, respectively. The accuracy of GD-IDTW was relatively close

to the improved method, with accuracy rates of 96.97%, 85.77%, and 100.00% for good, average, and poor action performance, respectively. The overall accuracy of the improved method outperformed other methods, because the improved method was significantly better than the GD-IDTW in handling fine actions, which could better capture subtle differences in these actions. In the similarity evaluation of good action performance, the improved method had an average accuracy close to GD-IDTW, with 88.15% and 88.01%, respectively, showing superior performance. As for average action performance, the improved method had an average accuracy of 90.12%, which was higher than other methods. In the similarity evaluation of poor action performance, the average accuracy of the improved method was 92.55%. The improved method not only enhances computational efficiency, but also improves the accuracy of similarity evaluation. This method has a certain positive effect on the Estevam team in obtaining semantic information from videos for recognizing actions [36]. The improved method effectively achieves accurate and real-time action similarity evaluation, which is of great significance for various training and evaluation systems that require real-time feedback. Meanwhile, this innovative method will make significant contributions to the development of sports science research and intelligent sports evaluation systems. However, this study does not consider the impact of lighting environmental factors on the evaluation results. Excessive or insufficient lighting can lead to the loss of details in the image, which in turn affects the accuracy of action recognition and the reliability of similarity evaluation. In the future, experiments can be extended to different lighting environments, including natural light, artificial light sources, and changes in lighting intensity. By testing the performance of the algorithm under various lighting conditions, the robustness of the algorithm in different lighting environments can be identified and optimized.

ACKNOWLEDGMENT

The research is supported by: A Study on Teaching Reform and Learning Effectiveness Improvement of College Physical Education Curriculum Based on Mobile Application Embedding (No.2024GZJX072).

REFERENCES

[1] C. Schneider, J. Rothschild, and A. Uthoff, "Change-of-direction speed assessments and testing procedures in tennis: A systematic review," J. Strength Cond. Res., vol. 37, no. 9, pp. 1888-1895, September 2023.

[2] M. Aslanyan, "On mobile pose estimation and action recognition design and implementation," Pattern Recognit. Image Anal., vol. 34, no. 1, pp. 126-136, January 2024.

[3] J. Purohit and R. Dave, "Leveraging deep learning techniques to obtain efficacious segmentation results," Arch. Adv. Eng. Sci., vol. 1, no. 1, pp. 11-26, July 2023.

[4] M. Wu, R. Wang, Y. Hu, M. Fan, Y. Wang, Y. Li, and S. Wu, "Invisible experience to real-time assessment in elite tennis athlete training: Sport-specific movement classification based on wearable MEMS sensor data," Proc. Inst. Mech. Eng. P J. Sports Eng. Technol., vol. 237, no. 4, pp. 271-282, October 2021.

[5] B. Giles, P. Peeling, S. Kovalchik, and M. Reid, "Differentiating movement styles in professional tennis: A machine learning and hierarchical clustering approach," Eur. J. Sport Sci., vol. 23, no. 1, pp. 44-53, December 2021.

[6] T. Perri, M. Reid, A. Murphy, K. Howle, and R. Duffield, "Differentiating stroke and movement accelerometer profiles to improve prescription of tennis training drills," J. Strength Cond. Res., vol. 37, no. 3, pp. 646-651, March 2023.

[7] D. Wood, M. Reid, B. Elliot, J. Alderson, and A. Mian, "The expert eye? An inter-rater comparison of elite tennis serve kinematics and performance," J. Sports Sci., vol. 41, no. 19, pp. 1779-1786, December 2023.

[8] S. Liu, "Tennis players' hitting action recognition method based on multimodal data," Int. J. Biometrics, vol. 16, no. 3-4, pp. 317-336, April 2024.

[9] R. Hu, "IoT-based analysis of tennis player's serving behavior using image processing," Soft Comput., vol. 27, no. 19, pp. 14413-14429, July 2023.

[10] T. Perri, M. Reid, A. Murphy, and R. Duffield, "Tennis serve volume, distribution and accelerometer load during training and tournaments from wearable microtechnology," Int. J. Perform. Anal. Sport, vol. 24, no. 4, pp. 285-297, December 2024.

[11] H. Setyawan, S. Suharjana, S. Purwanto, J. V. García-Jiménez, and A. Adriansyah, "The differences result in serve skill of junior tennis players assessed based on gender and age," Retos: Nuevas Perspectivas de Educación Física, Deporte y Recreación, vol. 54, pp. 272-278, 2024.

[12] C. Shuai, Y. Sun, X. Zhang, F. Yang, X. Ouyang, and Z. Chen, "Intelligent diagnosis of abnormal charging for electric bicycles based on improved dynamic time warping," IEEE Trans. Ind. Electron., vol. 70, no. 7, pp. 7280-7289, July 2023.

[13] D. Deriso and S. Boyd, "A general optimization framework for dynamic time warping," Optim. Eng., vol. 24, no. 2, pp. 1411-1432, August 2022.

[14] Q. Zhang, C. Zhang, L. Cui, X. Han, Y. Jin, G. Xiang, and Y. Shi, "A method for measuring similarity of time series based on series decomposition and dynamic time warping," Appl. Intell., vol. 53, no. 6, pp. 6448-6463, July 2022.

[15] H. Xing, L. Zhu, B. Chen, L. Zhang, D. Hou, and W. Fang, "A novel change detection method using remotely sensed image time series value and shape based dynamic time warping," Geocarto Int., vol. 37, no. 25, pp. 9607-9624, December 2021.

[16] V. Froese, B. Jain, M. Rymar, and M. Weller, "Fast exact dynamic time warping on run-length encoded time series," Algorithmica, vol. 85, no. 2, pp. 492-508, September 2022.

[17] T. Belkhouja, Y. Yan, and J. R. Doppa, "Dynamic time warping based adversarial framework for time-series domain," IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 6, pp. 7353-7366, June 2022.

[18] H. He and H. Li, "A new boosting algorithm for online portfolio selection based on dynamic time warping and anti-correlation," Comput. Econ., vol. 63, no. 5, pp. 1777-1803, May 2024.

[19] K. Vorpe, S. Hessinger, R. Poth, and T. Miljkovic, "Clustering regions with dynamic time warping to model obesity prevalence disparities in the United States," J. Appl. Stat., vol. 51, no. 4, pp. 793-807, March 2023.

[20] M. Hasanvand, M. Nooshyar, E. Moharamkhani, and A. Selyari, "Machine learning methodology for identifying vehicles using image processing," Artif. Intell. Adv., vol. 1, no. 3, pp. 170-178, April 2023.

[21] M. Poignard, G. Guilhem, M. Jubeau, E. Martin, T. Giol, B. Montalvan, and F. Bieuzen, "Cold-water immersion and whole-body cryotherapy attenuate muscle soreness during 3 days of match-like tennis protocol," Eur. J. Appl. Physiol., vol. 123, no. 9, pp. 1895-1909, September 2023.

[22] L. Wang, L. Lin, Y. Sun, S. Hou, and J. Ren, "The effect of movement speed on audiovisual temporal integration in streaming-bouncing illusion," Exp. Brain Res., vol. 240, no. 4, pp. 1139-1149, April 2022.

[23] A. Visser, D. Büchel, T. Lehmann, and J. Baumeister, "Continuous table tennis is associated with processing in frontal brain areas: an EEG approach," Exp. Brain Res., vol. 240, no. 6, pp. 1899-1909, June 2022.

[24] B. M. Pluim, M. G. T. Jansen, S. Williamson, C. Berry, S. Camporesi, K. Fagher, and C. L. Ardern, "Physical demands of tennis across the different court surfaces, performance levels and sexes: a systematic review with meta-analysis," Sports Med., vol. 53, no. 4, pp. 807-836, April 2023.

[25] H. He, H. Sun, and Y. Chen, "Analysis and assessment of ground motion characteristics and similarity using dynamic time warping distance," J. Seismol., vol. 27, no. 6, pp. 1013-1033, December 2023.

[26] M. Kumawat and A. Khaparde, "Development of adaptive time-weighted dynamic time warping for time series vegetation classification using satellite images in Solapur district," Comput. J., vol. 66, no. 8, pp. 1982-1999, August 2023.

[27] J. Xia and Y. Yamashita, "Online batch process monitoring with a combination of normal operating history data and physical knowledge," J. Chem. Eng. Jpn., vol. 55, no. 1, pp. 38-50, January 2022.

[28] A. Sabarinath, A. N. Rajesh, and S. S. G. V. L. Kumar, "Application of deep learning algorithms to correct bias in CMIP6 simulations of surface air temperature over the Indian monsoon core region," Int. J. Climatol., vol. 43, no. 16, pp. 7496-7515, December 2023.

[29] M. V. L. Pazini, H. de Abreu Corrêa, H. Haan, G. Zanon, and T. G. R. Clarke, "A dynamic time warping approach to access fatigue damage in composite pipes," Exp. Mech., vol. 64, no. 6, pp. 839-849, June 2024.

[30] Z. Yu, Z. Wang, and J. Wang, "Continuous wavelet transform and dynamic time warping-based fine division and correlation of glutenite sedimentary cycles," Math. Geosci., vol. 55, no. 4, pp. 521-539, April 2022.

[31] J. P. Choe, I. W. Hwang, J. H. Park, C. Amo, and J. M. Lee, "How valid is the commercially available tennis match analysis mobile application? Is it good enough?," Int. J. Perform. Anal. Sport, vol. 24, no. 1, pp. 58-73, October 2023.

[32] J. S. Israel, S. R. Loushin, S. U. Tetzloff, T. Ellenbecker, K. R. Kaufman, and S. Kakar, "Wrist motion assessment in tennis players using three-dimensional motion capture and dynamic electromyography," J. Wrist Surg., vol. 13, no. 3, pp. 264-271, July 2024.

[33] A. X. Chen, B. W. Yang, B. Lu, D. G. Nie, E. M. Jin, and F. X. Wang, "Gesture scoring based on Gaussian distance-improved DTW," Elektron. Elektrotech., vol. 30, no. 2, pp. 18-27, April 2024.

[34] W. Yan, X. Cao, and P. Ye, "Application of human posture recognition algorithms based on joint angles and movement similarity in sports assessment for physical education," Scalable Comput. Pract. Exp., vol. 25, no. 4, pp. 2385-2397, June 2024.

[35] B. Zhang, J. Chen, Y. Xu, H. Zhang, X. Yang, and X. Geng, "Auto-encoding score distribution regression for action quality assessment," Neural Comput. Appl., vol. 36, no. 2, pp. 929-942, October 2023.

[36] V. Estevam, R. Laroca, H. Pedrini, and D. Menotti, "Tell me what you see: A zero-shot action recognition method based on natural language descriptions," Multimed. Tools Appl., vol. 83, no. 9, pp. 28147-28173, September 2023.