VR-Motion Capture Method Design for the Teaching of Spinning Fitness Games in Universities

Hongzhou Ma
Division of Logistics and Infrastructure
Sichuan International Studies University
Chongqing, 400031, China
mhztgyx@163.com

Jiezhong Wu*
Department of Physical Education, Fuzhou University, Fuzhou, 350108, China, Graduate School, Beijing Sport University, Beijing,100091, China
Corresponding Author: jzwu99@126.com

Abstract: *As a part of physical education in colleges and universities, fitness game teaching is more and more applied to modern electronic equipment for auxiliary teaching. In order to improve the teaching quality and interestingness of college spinning teaching, a kind of teaching assistant technology of spinning fitness games combined with motion capture was proposed. In the process, kinect 3D depth camera is used for human motion capture, matrix operation is used for coordinate conversion, softmax layer is used for action classification and output classification results, integrated learning is used to supplement and reconstruct motion capture data, and the purpose of multi-person learning is achieved through photon server. The experimental results show that the loss value drops to 0.14 after 100 iterations in the test set. In the calculation accuracy test, the research method maintains 95.1% after 100s in Carnegie Mellon University (CMU) data set, which is higher than other methods. In the round-trip delay test, only 5 wave delay fluctuations occurred in the research method within 60s, and only 1 wave delay fluctuation exceeded 150ms. During the bone extraction test, the study method completed the restoration of 40 joints, and no bone loss occurred. The results show that the research method can more accurately capture the motion of spinning, and can effectively help improve the teaching quality of spinning games*.

Keywords: *Fitness games, Kinect, Motion capture, Virtual reality, Graph convolutional neural network, Ensemble learning*.

Received August 20, 2024; accepted October 12, 2025
https://doi.org/10.34028/iajit/23/3/3

1. Introduction

With the enhancement of people's health awareness and the increasing demand for fitness, the spinning bicycle, as an efficient fitness equipment, has been widely used [6]. In college education, fitness game teaching, as an innovative teaching method, is welcomed by more and more students [25]. Traditional spinning teaching often has the problem of lack of interest and interaction, which cannot stimulate students' active participation and lasting motivation. The assessment and feedback mean of spinning teaching are relatively limited, which cannot provide real-time movement guidance and personalized feedback [4, 16]. Traditional motion capture methods often require the use of complex sensor equipment, resulting in high deployment and use costs. The existing action classifiers have limited processing ability for multi-view data, and it is difficult to accurately identify and classify complex action sequences [9]. The traditional motion analysis and data transmission methods bring high delay in processing, and cannot accurately feedback the processing results in real time. Virtual Reality (VR) can build an environment that simulates the real world, allowing students to play fitness games in a more immersive environment. In order to reduce the complexity of motion capture, reduce the complexity of equipment used in fitness game teaching, and improve the interest and teaching quality of college fitness game teaching, the research innovatively designed a VR-motion capture method for fitness game teaching of spinning. The contributions of this paper are threefold. Firstly, the combination of virtual reality technology and motion capture technology provides a new interactive teaching method. Secondly, a low-cost motion capture system based on Kinect depth camera is designed, which improves the accuracy and system stability of capture through three-angle synchronous capture. Thirdly, research have developed a virtual interactive system to support multi-user interaction. Through Photon Server, different users can interact in the same virtual environment in real time, which enhances the interactive and interesting teaching.

The rest of this paper is organized as follows, the first section discusses and summarizes the current research achievements and applications of VR technology and motion capture technology. The second section is the design of VR-motion capture method for the teaching of spinning fitness games. The third section is the validity test and analysis of the research method. The last section is to discuss and summarize the full text.

2. Related Works

In recent years, the spinning exercise game has become a common method of fitness teaching in colleges and universities, and its related technologies have been paid

attention to by many professionals. Some scholars have discussed the technology of motion capture in fitness games. Aiming at the problem of motion capture in the study of running biomechanics, Carrier *et al*. [2] proposed a capture method using Garmin f nix 3HR. Three running conditions were set, flat, incline and descent, and data from the chest heart rate monitor was paired with the Garmin foot pod to record stride length, frequency and oscillation. The proposed method has a high motion capture accuracy and can effectively simulate the contact between motion and ground. Aiming at the motion capture problem of basketball indoor sports, Zhang *et al*. [26] proposed a positioning method based on position fingerprint. In the process, the Wifi fingerprint database is used to capture spatial features, and constraint points are added in the calculation process and multiple target points are selected in the region for simultaneous analysis, and the features of spatial objects are analyzed. The proposed method has high positioning accuracy Mihcin [14] proposed a motion capture system using a wearable inertial measurement unit to address the motion capture problem of leg wearables. In the process, the motor is used to simulate the motion range, and the parameters are set according to the motion state of hip joint, ankle joint and knee joint. Photoelectric and Inertial Measurement Unit (IMU) sensors are introduced to collect the data, and then the flexural extension value of the joint is calculated. The proposed method can effectively complete leg motion capture with low error [8]. Aiming at the problem of motion capture of human movement, scholars such as Qiu *et al*. [17] proposed a motion data analysis method based on gradient descent. During the process, 15 sensors are set at the key position of the human body to collect data, and the displacement is calculated using the zero-velocity update algorithm. Then the data in the sensor is reconstructed by fusion; the fusion data of displacement and human posture are obtained by unconstrained traversal. The proposed method has good capture accuracy Ligorio *et al*. [8] proposed a magnetometer-free capture technology aimed at the accuracy of human motion capture. In the process, Kalman filter is used to estimate the three-dimensional direction, the data fragment of the sensor is calibrated and the error is repaired by introducing compensation mechanism. The motion data is obtained by fusion of inertial sensor data and ground reality data. The proposed method has good motion capture accuracy.

There are also some scholars who have conducted relevant research on VR technology. Wu *et al*. [24] proposed a method of using VR equipment to get real interactive experience in view of the problem of tourists' real experience in tourism. In the process, a panoramic virtual reality video game was set according to the characteristics of the tourist area, and more than 400 tourists were arranged to carry out virtual reality identification and virtual reality familiarity experiment. The proposed method can effectively improve tourists' experience of scenic spots and facilities Javaid and Haleem [7] proposed a simulation method using VR technology for medical surgery and teaching. In the process, a simulated three-dimensional environment is established to simulate the actual situation, and the data of patients are imported to generate virtual images; the simulation observation and teaching of operation are realized by setting each parameter. The proposed method can effectively improve the teaching quality of medical surgery and provide more reference data for surgery. Getuli *et al*. [5] proposed an experience scheme using VR technology to solve the problem of safety training in engineering construction. In the process, Building Information Modeling (BIM) technology is used to establish the engineering construction environment, complete the learners' actions in the virtual environment through data fusion, and feedback the actions to the learners through the equipment. The proposed method can effectively improve the effectiveness of safety training and reduce the incidence of safety accidents Luo *et al*. [10] proposed a teaching method using VR technology to enhance interest in higher education environments. The VR technology in the past 20 years is analyzed and reviewed, and the teaching system is designed according to the characteristics of college teaching courses and students' needs. The proposed method is interesting and can improve students' concentration in the teaching process. Chaccour *et al*. [3] proposed a method using terahertz large bandwidth to solve the communication delay problem when users use VR devices. The experiment analyzed the end-to-end delay tail representation, quantified the VR communication risk, derived the probability distribution function of data delay, and inferred the reliability scenario. The proposed method has high data transmission rate and security reliability.

To sum up, the application of motion capture technology in fitness game teaching has been proven to improve student participation and teaching effect. Although VR technology and motion capture technology have been widely studied and applied in many fields, how to effectively combine these technologies to improve the teaching experience and effect is still a problem worthy of in-depth discussion. In view of this, the research proposes VR and motion capture methods for the teaching of sports games in colleges and universities, in order to provide more technical references for the teaching of sports games in colleges and universities.

3. Design of VR-Motion Capture Method for the Teaching of Spinning Fitness Games

The integration of effective motion capture methods into VR technology can improve the quality of spinning game teaching in universities. This section will focus on

the VR-motion capture method of research design and the technical means used in the method.

3.1. Design of Motion Capture Method for Spinning Fitness Based on Depth Camera

Spinning has gradually become a popular content in the teaching of fitness games in colleges and universities. However, in the teaching process, it is difficult for teachers to accurately analyze and guide students' movements only by means of video recordings, and students also lack appropriate feedback systems when learning by themselves [23]. Virtual Reality integrates vision and hearing to create a virtual experience environment. By interacting with the virtual environment, users can simulate actions and the real environment to the maximum extent [15, 22]. When using VR, the user's movements need to be captured. The traditional connected device is large and complex, and is not suitable for the fitness motion capture of the spinning bike. In recent years, the technology of motion capture using computer vision has gradually matured, and the capture of complex actions has good performance. The research uses Kinect 3D depth camera for human motion capture [12, 13]. The data acquisition method of Kinect camera is shown in Figure 1.

As can be seen from Figure 1, the Kinect camera first uses infrared light to obtain the position of the human body. The Kinect camera is equipped with an infrared light transmitter, which is capable of emitting specific wavelengths of infrared light. When infrared light interacts with an object, a portion of it is reflected back. After extracting the outline of the human body, the human body parts are classified. In the classification process, irrelevant parts of the human body such as skin, organs and hair are omitted, and the skeletal framework is analyzed. With multiple points of depth information, Kinect is able to build a mannequin in three-dimensional space. The bones are divided into rotatable parts, and the parts consisting of multiple bones such as the spine and hands are simplified. Finally, the basic human skeleton diagram formed by 25 nodes is obtained, and nodes are added and refined in the subsequent calculation. In order to avoid the blind field of visual field caused by a single camera, three depth cameras were selected to be connected by a Universal Serial Bus (USB) to ensure the collection of complete image data. The designed three-angle motion capture system is shown in Figure 2.

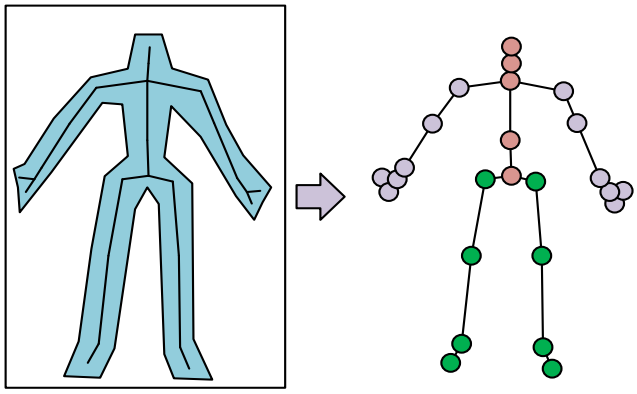


Figure 1. Kinect camera data collection.

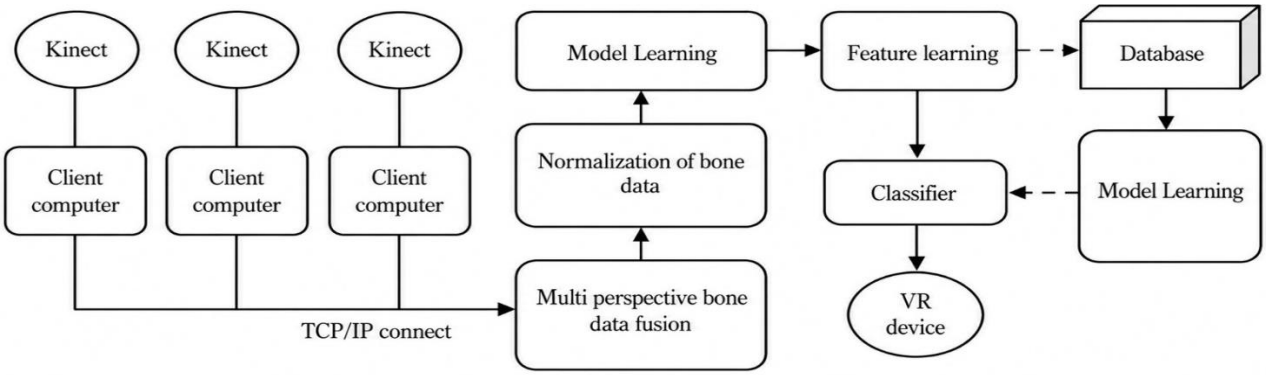


Figure 2. Three view motion capture system.

In Figure 2, the three depth cameras are connected to their respective client computers via USB, and the client computers stream data over Transmission Control Protocol/Internet Protocol (TCP/IP). The human bone data captured from the three perspectives were fused and normalized to correct the abnormal bone positions. After modifying the bone position, the bone features are extracted, and the processed data is introduced into the database for model learning. The learned classifier is used to analyze the feature data, and then the analyzed data is output to the VR device. The position and Angle of the three depth cameras are different, so the coordinate system needs to be unified before data fusion. When calibrating bone data from multiple perspectives, each depth camera's field of view contains a skeleton view. The initial Angle calculation of the skeleton is shown in Equation (1).

$$\theta = \arctan\left(\frac{Z_r - Z_l}{X_r - X_l}\right) \tag{1}$$

In Equation (1), Z_r represents the z coordinate of the right shoulder joint point; Z_l represents the z coordinates of the left shoulder joint point; X_r represents the x coordinate of the right shoulder joint point; X_l represents the x coordinate of the left shoulder joint point; θ represents the initial angle. The initial center position of

the skeleton is obtained by calculating the average value of joint coordinates. The calculation of coordinates in the world coordinate system is shown in Equation (2).

$$\begin{cases} K_j'\left(K_{k',j},Y_{k',j},Z_{k',j}\right)=\left(X_{k,j}-X_c,Y_{k,j}-Y_c,Z_{k,j}-Z_c\right) \\ W_j\left(X_{w,j},Y_{w,j},Z_{w,j}\right)=\left(X_{k',j}\cos\theta+Z_{k',j}\sin\theta,Y_{k',j},Z_{k',j}\cos\theta-X_{k',j}\sin\theta\right) \end{cases} \tag{2}$$

In Equation (2), X_c, Y_c, Z_c represents the coordinates of the initial center position; K_j represents the coordinates to be converted, where the coordinate values on different axes are $X_{k,j}$, $Y_{k,j}$, $Z_{k,j}$ respectively; K_j' represents the converted coordinates; W_j represents the coordinates in the world coordinate system, where the values on the different axes are $X_{w,j}$, $Y_{w,j}$, $Z_{w,j}$. To facilitate the transfer of data from world coordinates to a depth camera perspective, coordinate conversion is performed using matrix operations. The Software Development Kit (SDK) reports the status of the bone point, which is *Tracked*, *Inferred*, and *Not Tracked*, representing three different ways of tracking the bone point. The report is calculated as shown in Equation (3).

$$w(s_{j,f})\begin{cases} 1.0, s_{j,f}\ is\ Tracked \\ 0.5, s_{j,f}\ is\ Inferred \\ 0.0, s_{j,f}\ is\ Not\ tracked \end{cases} \tag{3}$$

In Equation (3), $w(\cdot)$ represents tracking performance; $s_{j,f}$ represents the tracking state of the f bone point coordinates of the j depth camera. The tracking position accuracy is calculated, as shown in Equation (4).

$$\mu\left(d_{j,f}\right)=1-\left(\frac{0.4946e^{0.7d_{j,f}}-1.1457}{5.7316-1.1457}\right) \tag{4}$$

In Equation (4), $0.4946e^{0.7d_{j,f}}$ is the evaluation error fitting function; $\mu(\cdot)$ represents the range tracking position accuracy of the target and the depth camera. Then, the bone joint data extracted by different depth cameras were normalized to keep the bone sizes captured by all depth cameras consistent, and the bone position data were corrected. The feature extraction is accomplished by analyzing the joint position, joint Angle, linear velocity and angular velocity of the bone. In this study, the position of the central joint of the two hips of the bone was set as the coordinate origin, and the joint speed was calculated using two consecutive frames, as shown in Equation (5).

$$V_j[n]=\frac{N_j[n]-N_j[n-1]}{t}, j=1,2,3,...,24 \tag{5}$$

In Equation (5), $V_j[n]$ represents the speed of joint j of frame n; t represents the interval between two consecutive frame times; $N_j[n]$ represents the joint position of j of frame n; $N_j[n\text{-}1]$ represents the j joint position of the first frame before n frame. The angular velocity of the joint is calculated by two consecutive frames, as shown in Equation (6).

$$V_{angle,j}[n]=\frac{A_j[n]-A_j[n-1]}{t} \tag{6}$$

In Equation (6), $V_{angle,j}[n]$ represents the angular velocity of joint j of frame n; $A_j[n]$ represents the j joint angle value of the n frame; $A_j[n\text{-}1]$ represents the j joint angle value of the first frame before n frame. In order to recognize the motion better, a multi-view motion classifier is designed to classify the motion, and the accuracy of the motion classification is improved by comprehensively considering the depth camera data from different angles. The multi-view action classifier is shown in Figure 3.

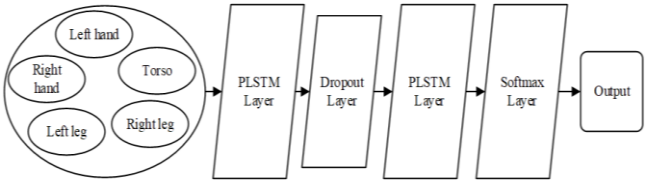


Figure3. Multi perspective human motion classifier.

In Figure 3, the multi-view human action classifier divides the human body into five parts: trunk, left arm, right arm, left leg and right leg. The five actions are stored independently in the partial perceptual long- and short-time memory network, and the combination containing contextual information is output. The Dropout layer is inserted into part of the perceptual short-duration memory network to prevent overfitting, and only the Softmax layer is used to classify the actions and output the classification results. Softmax converts the raw output to a probability distribution through a nonlinear transformation. This conversion is achieved by indexing the output and normalizing it, ensuring that the probabilities for all classes sum to 1. For each action category, Softmax outputs a value between 0 and 1 that represents the probability that the sample belongs to each category. During training, the Softmax classifier is used in conjunction with the cross-entropy loss function to help optimize the network parameters so that the probability distribution of the model output is as close as possible to that of the real label.

3.2. Optimization of Motion Capture Method for Spinning Bikes Based on Machine Learning and Design of User Interaction System

The motion capture and recognition accuracy of a spinning bike using only the depth camera and the motion classifier is insufficient, and it is prone to misjudgment in the face of complex fitness movements and high-speed moving joints [19, 21]. In motion capture and recognition, some data may be lost due to hardware limitations or recording conditions of the camera. In this study, ensemble learning is used to supplement and reconstruct motion capture data. The data reconstruction process is shown in Figure 4.

In Figure 4, in the process of data reconstruction, bone data needs to be preprocessed first, which includes

two steps of bilateral filtering and down-sampling. Then the incomplete motion capture sequence is generated according to the action sequence, and the trajectory parameters of the marked points are extracted by reference to the action inertia and other associated actions, and an independent recovery model is introduced. Constraints are applied using time and trajectory continuity, and then weighted averages are applied using distance probabilities. The recovered motion capture sequence is generated by reference to the marker distance likelihood. Two-sided filtering of motion capture data is calculated as shown in Equation (7).

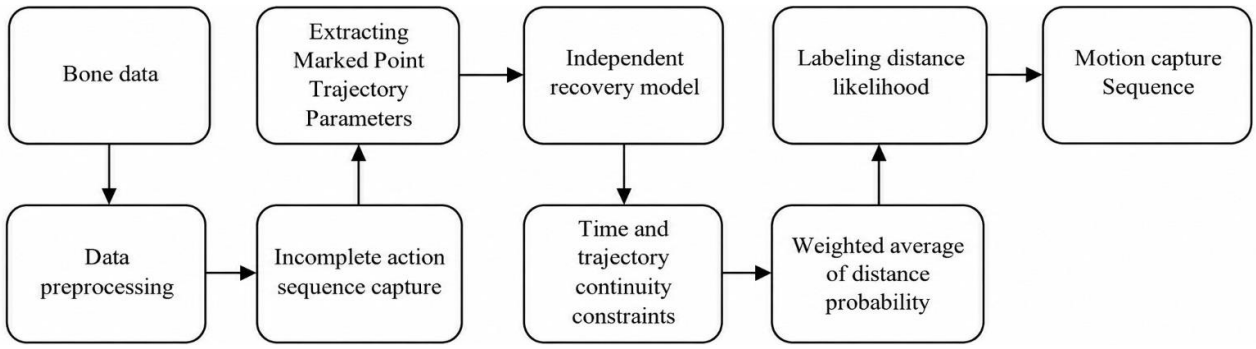


Figure 4. Data reconstruction process.

$$\hat{d}(x)=\frac{\sum_{i\in\Omega} w_s(i)w_r(i)d(x)}{\sum_{i\in\Omega} w_s(i)w_r(i)} \tag{7}$$

In Equation (7), $\hat{d}(x)$ represents the value of the motion capture data after filtering; Ω represents the scope of an area of subparagraph x; $w_s(i)$ and $w_r(i)$ are weights. The weights are calculated as shown in Equation (8).

$$w_s(i)=e^{-\frac{\|i-x\|^2}{2\sigma_s^2}}, w_r(i)=e^{-\frac{|d(i)-d(x)|^2}{2\sigma_r^2}} \tag{8}$$

In Equation (8), σ_s and σ_r represent smoothing parameters. The standard deviation of Euclidean distance between the marked trajectories to be recovered and other marked trajectories is calculated, as shown in Equation (9).

$$\sigma_j=std\left(\left\|\left(m-p_j\right)\right\|\right) \tag{9}$$

In Equation (9), σ_j represents standard deviation; m represents the marked track to be recovered; p_j represents other marked tracks. More distance changes are sorted as a potential reference for the trajectory to be reconstructed. In order to recover the motion trajectory accurately, the global linear regression model is introduced to predict and generate the complete motion sequence based on the existing motion data. The missing part of the marked trajectory is calculated, as shown in Equation (10).

$$\begin{cases}\beta_i=argmin_{b_i}\left(\sum_n (X(n)\cdot b_i-m_i(n))^2\right)\\ i=\{1,2,3\}, n\in\{1,\ldots,N\}\\ \tilde{m}\,[X\cdot\beta_i]_{1\le i\le 3}\end{cases} \tag{10}$$

In Equation (10), β_i represents the regression coefficient vector of m_i after the least squares error minimization extraction; m_i represents column i of the recovery matrix; $m_i(n)$ represents the n element in column i of the recovery matrix; $\tilde{m}$ represents the recovery trajectory calculated by the global linear regression model.

Graph convolutional neural network is a machine learning method that can extract and analyze structural feature changes of bone over continuous time. Traditional convolutional neural network has good performance in the analysis of video and image data, and the two can be applied to the analysis of exercise movements in spinning [1, 11]. Traditional convolution and graph convolution are shown in Figure 5.

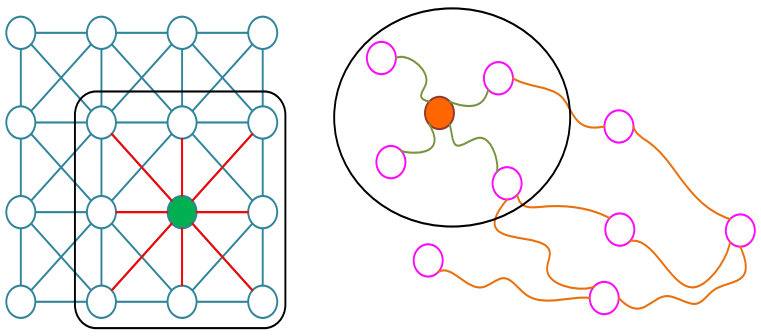


Figure 5. Traditional convolutions and graph convolutions.

In Figure 5, traditional convolution is processed with ordered neighbor nodes and a limit on node size. If the filter size is 3×3, the middle node neighbors are the eight nearby nodes. Different from traditional convolutional neural networks, the core of graph convolutional neural network is that it can capture the complex relationship between nodes. In motion capture, bone data can be viewed as a graph where nodes represent bone joints and edges represent the connections between joints. Through the graph convolution operation, the graph convolutional neural network can effectively learn the spatial relationship between nodes, so as to classify and recognize the actions accurately [18, 20]. In this paper, the image convolutional neural network is optimized by combining viewpoint transformation to extract fitness action information better. The graph convolutional neural network model based on viewpoint transformation is shown in Figure 6

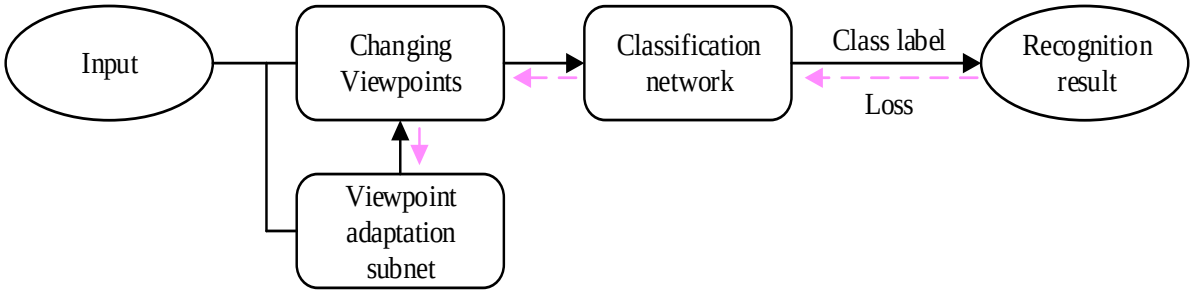


Figure 6. Graph convolution neural network action recognition.

In Figure 6, in the process of action recognition, the motion characteristics of the spinning cycle fitness will be analyzed first, and the appropriate angle will be selected by entering the viewpoint adaptation subnet to complete the viewpoint transformation. Then, the action data is entered into the classification network for learning and the appropriate calculation parameters are generated. Then, the flexibility of the model is increased through the graph topology to obtain more general features for data capture of different action samples, and the first and second order information of the skeleton is simulated simultaneously using the two-flow structure. The graph convolution operation on the vertices of spatial dimension pairs is shown in Equation (11).

$$f_{out}\left(v_i\right)=\sum_{v_j\in \mathrm{B}_i}\frac{1}{Z_{ij}}f_{in}\left(v_j\right)\cdot u\left(l_i\left(v_j\right)\right) \tag{11}$$

In Equation (11), f represents feature mapping; v represents a vertex on a space-time graph; B_i denotes the convolution sampling area representing the vertices; v_j denotes the domain 1 node of the vertex; u represents traditional convolution operations; l_i represents the mapping function; Z_{ij} denotes the contributions of each subset are averaged. The skeletal sequence can be regarded as a three-dimensional third-order tensor with the shape of channel number, frame number and joint number. After the tensor is expanded and rearranged into a matrix, the input feature map and output feature map can be redefined. The normalized embedded Gaussian function is used to estimate the feature similarity between graph nodes, as shown in Equation (12).

$$f\left(v_i,v_j\right)=\frac{e^{\vartheta\left(v_i\right)^T\phi\left(v_j\right)}}{\sum_{j=1}^{b}e^{\vartheta\left(v_i\right)^T\phi\left(v_j\right)}} \tag{12}$$

In Equation (12), b represents the number of nodes; T stands for transpose operation; ϑ and ϕ are embedded functions. The similar matrix is calculated, as shown in Equation (13).

$$\begin{cases} M_{\vartheta k}=\left|W_{\vartheta k}X_{(1)}\right|_{C_e\times T\times N} \\ M_{\phi k}=\left|W_{\phi k}X_{(1)}\right|_{C_e\times T\times N} \\ C_k=SoftMax\left(M_{\vartheta k(3)}M_{\phi k(3)}^T\right) \end{cases} \tag{13}$$

In Equation (13), $M_{\vartheta k}$ and $M_{\phi k}$ are two embedded feature graphs; W_ϑ and W_ϕ are parameters of the embedded function; C_k stands for similarity matrix. The output value of the feature vector after processing by Softmax is shown in Equation (14).

$$SoftMax\left(d_i\right)=s_i=\frac{e^{d_i}}{\sum_{j=1}^{d}e^{d_j}},\forall i\in1,\ldots,q \tag{14}$$

In Equation (14), d represents the component of the eigenvector; s_i represents the probability that the input action is calculated and predicted to be the i action. Multiple probabilities form a classification probability vector, and the fusion probability is calculated, as shown in Equation (15).

$$t_i=s_i^g+s_i^j,\forall i\in1,\ldots,n \tag{15}$$

In Equation (15), s_i^g represents the predicted action probability of bone flow; s_i^j represents the probability of joint flow predicting action; t_i represents the predicted action label. The predicted action label with the largest value is the predicted action result. In order to improve the interactive and immersive feeling of the spinning fitness game, the research uses the virtual interactive

system to run. Translate user movements into feedback in a VR environment in real time, providing a more intuitive and personalized fitness experience. In actual use, learners need to use the VR function through the user interaction system. Based on the designed motion capture method, the virtual interaction system is designed, as shown in Figure 7.

In Figure 7, when learners use the system, they establish a connection with the running environment based on multi-perspective motion capture; and convert the user's actions into VR content, reflecting the real user's spinning fitness actions. By analyzing the user's actions, the computing engine calculates the possible scene changes caused by each action. After that, the system's calculation content is transformed into an image by the display device and displayed in the learner's field of vision. In order to improve the applicability of the system in the teaching of spinning fitness games in colleges and universities, the multi-user interaction function was added to the system. Through photon server, multiple users can enter the same virtual environment through different ports, and the data of each port is continuously uploaded and downloaded, so as to achieve the purpose of learning and teaching together, so that the teaching quality is improved.

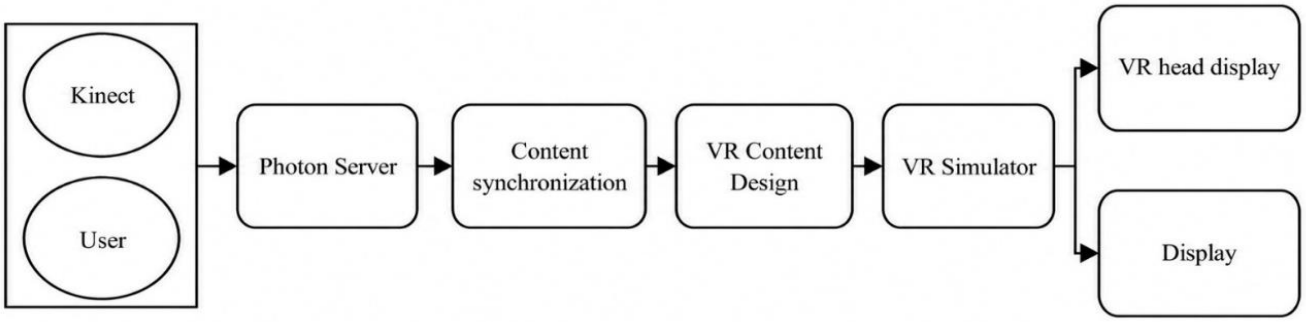


Figure 7. Virtual interactive system.

4. Performance Test and Application Analysis of VR-Motion Capture Method for Teaching Spinning Fitness Games

Virtual Reality-motion capture technology can make spinning fitness games more interesting. This section will test the performance of the research method and perform an applied analysis in a real-world environment to determine the effectiveness of the research method.

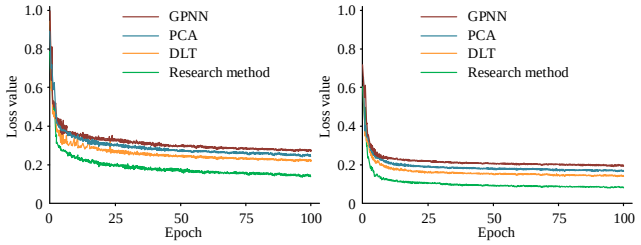
4.1. Performance Test of VR-Motion Capture Method for Spinning Fitness Game Teaching

To analyze the effectiveness of the VR-motion capture method in the teaching of spinning fitness games in colleges and universities, the performance of the research method was tested first. The basic hardware environment settings of the experiment are shown in Table 1.

Table 1. The experimental basic environmental parameters.

Parameter variables	Parameter selection
Operating system	Windows10
Operating environment	Matlab
System PC side memory	16G
CPU dominant frequency	3.20GHz
GPU	RTX-2060
Central Processing Unit (CPU)	Intel⑧CoreTM i7-10400

The research uses the High Tech Computer Corporation (HTC) Vive headset jointly developed by HTC and VALVE for screen display, and the theoretical delay of the device can be maintained at about 20ms. The movement data sets of spinning bikes are small, so the Carnegie Mellon University (CMU) dance data set and University of Mons (UMONS) data set, which also have complex and high-speed movements, are used for performance testing. The CMU Dance dataset, collected by the CMU Graphics Lab, contains captured data on a variety of dance movements. Each movement in the dataset is performed by professional dancers in an optically captured environment with high temporal resolution and spatial precision. This dataset was chosen because it contains a variety and complexity of movements similar to those you might encounter in a spin class, making it suitable for evaluating the performance of motion capture methods. The UMONS dataset, collected by the University of Namur in Belgium, focuses on human motion capture in everyday life. It includes a variety of dynamic actions such as walking, running, and jumping, and also has high-precision time and space information. The UMONS dataset was chosen because it provides a dynamic range similar to that of a spinning bike action, helping to verify the accuracy and robustness of the proposed method when dealing with high-speed movements. In the experiment, the three Kinect depth cameras were distributed at 120 degrees, and the Graph Convolutional Network (GCN) model used 5 convolutional layers of the graph, each layer had 32 filters. With the Adam optimizer, the learning rate is initially set to 0.001 and decays after 100 epochs. To ensure the reliability of the results, all experiments were repeated five times and the average was calculated. In the test, General Regression Neural Network (GRNN), Principal Component Analysis (PCA), Direct Linear transformation Direct Linear Transform (DLT) is designed to compare the methods. Firstly, the loss curve of the research method is tested. As shown in Figure 8.

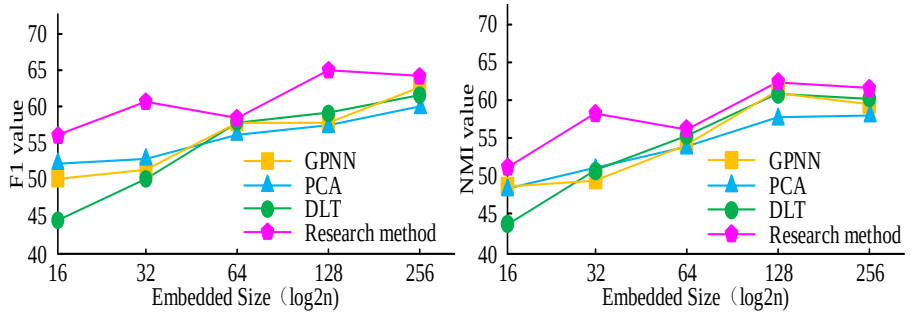


a) Testing losses. b) Training losses.

Figure 8. Loss curve testing.

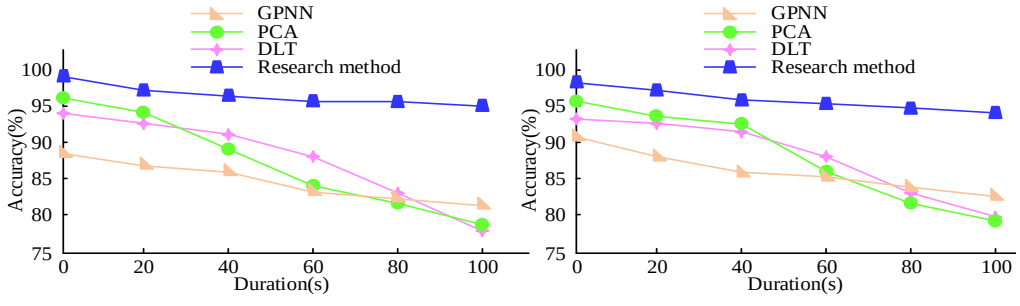
In Figure 8, when the loss value test was carried out, the four methods showed a rapid decline in the early stage and a steady decline in the later stage in the test set and training set. In the test set, four methods declined rapidly in the first 12 iterations and slowly thereafter. The loss of GPNN decreases to 0.36 at the 12th iteration and 0.27 at the 100th iteration. The loss value of DLT decreases to 0.30 at the 12th iteration and 0.22 at the 100th iteration. The loss value decreased to 0.23 at the 12th iteration and 0.14 at the 100th iteration. In the training set, the four methods declined rapidly in the first 10 iterations and slowly thereafter. The loss value of DLT decreases to 0.19 at the 10th iteration and 0.13 at the 100th iteration. The loss value decreased to 0.12 at the 10th iteration and 0.08 at the 100th iteration. It shows that the method has better iteration efficiency and better generalization ability in calculation. The results can be attributed to the effective integration of the Softmax classifier and GCN adopted. Softmax classifiers perform well in multi-classification tasks, while GCN is particularly suited for working with graph structure data, such as human bone data. This combination makes use of the advantages of the two models and improves the learning efficiency. The Normalized Mutual Information (NMI) and F1 values of the research method were tested, as shown in Figure 9.

In Figure 9, F1-values and Normalized Mutual Information (NMI) values of the four methods as a whole increase with the increase of embedding size. In the F1-value test, GPNN has F1-value of 50.2 when the embedding size is 16log2n, and increases to 62.7 when the embedding size is increased to 256log2n. The F1-value of the research method is 55.9 when the embedding size is 16log2n, and increases to 64.3 when the embedding size is increased to 256log2n. In the NMI value test, GPNN has an NMI value of 47.1 when the embedding size is 16log2n, and increases to 58.7 when the embedding size is increased to 256log2n. The NMI value of Principal Component Analysis (PCA) is 47.0 when the embedding size is 16log2n and increases to 57.7 when the embedding size is 256log2n. The NMI value of the research method is 51.8 when the embedding size is 16log2n and 62.1 when the embedding size is increased to 256log2n. Although the research method has declined in the process, it always maintains higher F1-value and NMI value, which indicates that the research method has better partitioning performance and better model performance during detection. The calculation accuracy of the research method was tested, as shown in Figure 10.



a) F1 value. b) NMI value.

Figure 9. F1 value.



a) CMU. b) UMONS.

Figure 10. Calculation accuracy.

In Figure 10, when the research method was tested for computational accuracy, the accuracy of both data sets decreased as the duration increased. In the CMU dataset, the initial accuracy of GPNN was 88.7%, which

dropped to 81.8% after 100s. The initial accuracy of the method was 98.8%, which decreased to 95.1% after 100s. In the UMONS dataset, the initial accuracy of GPNN was 91.3%, which dropped to 82.6% after 100s. The initial accuracy of PCA was 95.9%, which decreased to 78.9% after 100s. The initial accuracy of DLT was 93.1%, which dropped to 79.3% after 100s. The initial accuracy of the research method was 98.1%, which decreased to 93.7% after 100s. The high accuracy and good persistence of the model on the CMU data set prove the robustness of the motion capture method in processing complex motion sequences. The results relate to GCN's ability to capture the complex relationships between bone points that are crucial to understanding the nature of movement. The action confusion matrix of the research method was generated, as shown in Figure 11.

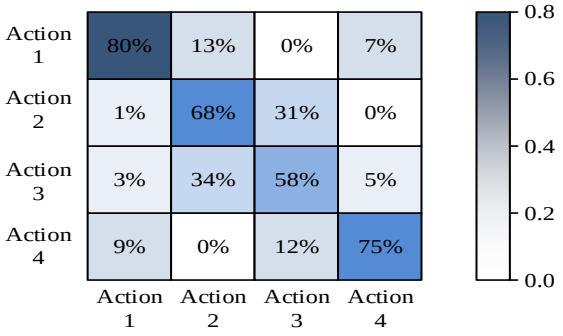
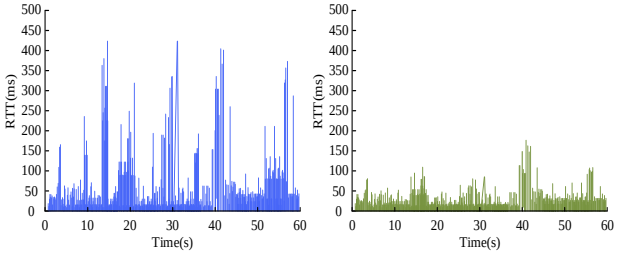


Figure 11. Action confusion matrix.

In Figure11, the probability of misjudgment between action 2 and action 3 with similar motion shapes is relatively high, reaching more than 30% but less than 35%. Considering that there is not a large number of similar movements in the spinning fitness teaching game, the error rate of less than 35% is acceptable. The probability of misjudgment is small between movements with large differences in movement patterns, and the probability of misjudgment between movements 4 and 2 can be as low as 0%. It shows that the research method has high accuracy in the judgment of movement with obvious characteristic difference.

4.2. Application Analysis of VR-Motion Capture Method for Spinning Fitness Game Teaching

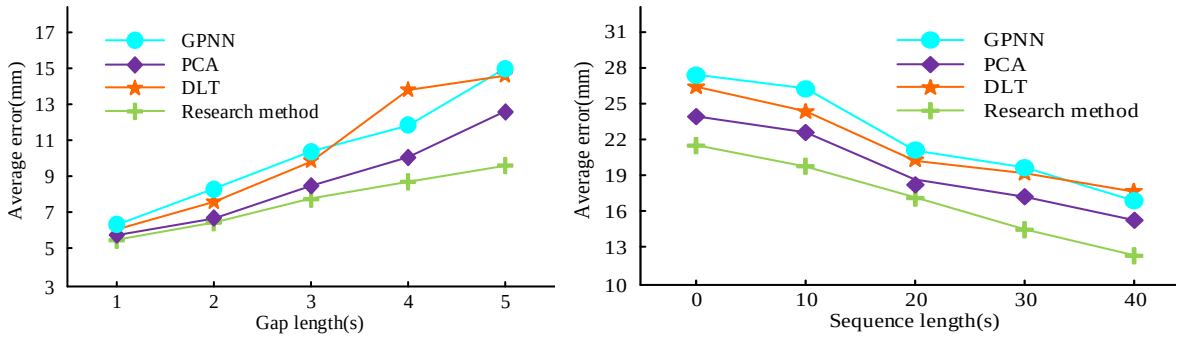
In the application analysis of the research method, 2 testers were set up to conduct the system operation in accordance with the normal teaching process of spinning fitness games. Round-trip delay of the research method is generated, as shown in Figure 12.



a) Wireless signal transmission. b) Research method.

Figure 12. Round-trip delay curve.

In Figure 12, during the 1-minute data transmission process, the round-trip delay of both the wireless signal transmission method and the research method showed partial fluctuations. The minimum delay of wireless signal transmission method reaches about 20ms; In most time domain, the delay is kept within 20ms-60ms. Within 60s, 8 waves of large delay fluctuation occurred. Four of the waves reached more than 300ms. The minimum time delay of the research method reaches about 20ms. In most time domain, the delay is kept within 20ms-30ms. 5 wave delay fluctuations occurred within 60s. Among them, 4 waves remain below 150ms; Wave 1 exceeds 150ms but stays below 200ms. It shows that the data transmission delay of the research method is lower, and it can provide more accurate and timely control feedback when applied to VR equipment. The results are closely related to the data synchronization strategies adopted by the universal serial bus and TCP/IP protocol, which ensure the fast data transmission and processing. The average error of the research method at different gap lengths and sequence lengths was tested, as shown in Figure 13.



a) Gap length change. b) Sequence length change.

Figure 13. Average error.

In Figure 13, the average error of the four methods increases with the increase of gap length and decreases with the increase of sequence length. When the gap length changes, the average error of GPNN is 6.3mm

when the gap length is 1s, and 15.1s when the gap length increases to 5s. The average error of DLT when the gap length is 1s is 6.1mm, and the average error increases to 14.7s when the gap length increases to 5s. The average error of the research method is 5.4mm when the gap length is 1s, and 9.5mm when the gap length is increased to 5s. When the sequence length changes, the average error of PCA is 23.8mm when the sequence length is 0s, and decreases to 15.1mm when the sequence length is 40s. The average error of DLT is 26.5mm when the sequence length is 0s, and decreases to 16.9mm when the sequence length increases to 40s. The average error of the method is 21.5mm when the sequence length is 0s, and decreases to 12.4mm when the sequence length increases to 40s. The research method has a lower average error and can better restore the real motion in motion capture. The modified Z-coordinate curve of the research method was tested, as shown in Figure 14.

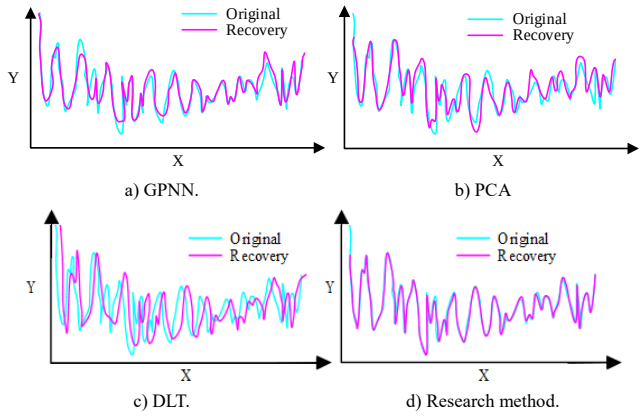


Figure 14. Correction of coordinate curve testing.

In Figure 14, when coordinate curve correction was carried out, the corresponding coordinate curve was generated by the four methods. The trend of the curve generated by GPNN is basically consistent with the original curve, but there are more than 10 deviations from the original movement amplitude in the movement segment. The overall amplitude deviation of the curve generated by PCA is small, but there are three misjudgments in the movement section. The amplitude and direction of the curve generated by DLT are basically the same as the original curve, but there is a significant delay from the original action on the X-timeline. The amplitude and trend of the curve generated by the research method are basically consistent with the original curve, and there is no significant delay on the X-time axis. It shows that the research method can generate the coordinate change curve of the motion node more accurately. The low level of average error and high consistency of coordinate curve indicate that the data preprocessing and reconstruction process adopted by the acting database effectively improves the quality of motion capture data. The application of bilateral filtering, down-sampling and global linear regression model can improve the prediction accuracy of the model. The bone extraction results of the research method were tested, as shown in Figure 15.

In Figure 15, when bone extraction was carried out, the bone extraction results of GPNN showed the absence of 4 bones and 2 nodes in the leg part. The results of PCA bone extraction showed that there were 3 missing bones in the leg and arm. The extraction results of DLT showed that 5 bones and 2 joints were missing in the leg and arm parts. The results of bone extraction showed that the bones and 40 joints were recovered well, and there was no missing condition. It shows that the method can extract human bones well and accurately.

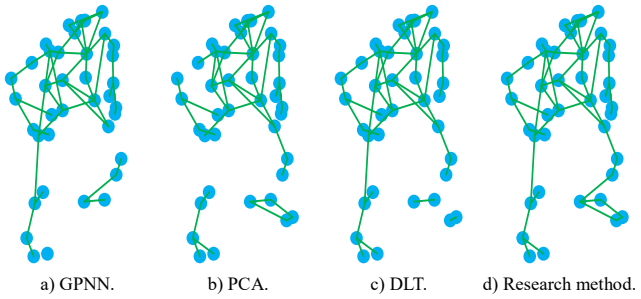


Figure 15. Bone extraction results.

5. Conclusions

The technology selection of college spinning fitness game teaching affects the student experience and teaching effect of fitness game teaching. Based on VR technology, 3D depth camera is used to capture human motion and generate bone map. Then, three depth cameras are connected by universal serial bus to construct a three-view motion capture system. Then ensemble learning is introduced to reconstruct motion capture data and trajectory recovery is calculated by global linear regression model. Finally, the virtual interactive system is designed, and the effectiveness of the research method is analyzed. In the test of loss value, the loss value of the research method decreases to 0.12 after 10 iterations in the training set, and to 0.08 after 100 iterations, which is lower than other methods. In the F1 value and NMI value test, when the embedding size of the research method is 256log2n, the F1 value is 64.3 and the NMI value is 62.1, which is higher than other methods. In the calculation accuracy test, the accuracy of the research method decreases from 98.1% to 93.7% after 100s in the UMONS data set, which is higher than other methods. When the round-trip delay test is carried out, the delay of the research method is maintained within 20ms-30ms under stable condition, and the delay fluctuation of less than 200ms only occurs 5 times within 1 minute. In the average error test, the error of the research method is 9.5mm when the gap length is 5s, and 12.4mm when the sequence length is 40s. The results of curve modification are basically consistent with the original curve. The integrity of the bones and 40 nodes was maintained by the research method during

bone extraction. The above results show that the method can accurately complete the motion capture of the spinning cycle and has better system stability. However, this experiment was conducted in the Local Area Network (LAN) environment, and the subsequent large-scale test will be conducted in the internet environment to analyze the impact of network delay, bandwidth fluctuation and data packet loss on the real-time and stability of the system, and optimize the data transmission protocol and fault tolerance mechanism. And explore the system performance in high concurrency scenarios, study the distributed server architecture and load balancing strategy, and improve the fluency of multi-user synchronous interaction. Adapt to mobile terminals, diversified VR devices and operating systems, develop lightweight clients, and enhance the universality and flexibility of the system.

Data Availability Statement

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Reference

[1] Atlasov B. and Selskiy A., "The State and Prospects of Using Virtual reality Technologies in Sports: A Brief Review," *Russian Journal of Information Technology in Sports Technology*, vol. 2, no. 1, pp. 13-21, 2025, DOI: 10.62105/2949-6349-2025-2-1-13-21

[2] Carrier B., Creer A., Williams L., Holmes T., and et al., "Validation of Garmin Fenix 3 HR Fitness Tracker Biomechanics and Metabolics (VO2max)," *Journal for the Measurement of Physical Behaviour*, vol. 3, no. 4, pp. 331-337, 2020. DOI: 10.1123/jmpb.2019-0066

[3] Chaccour C., Soorki M., Saad W., Bennis M., and Popovski P., "Can Terahertz Provide High-Rate Reliable Low-Latency Communications for Wireless Vr?," *IEEE Internet of Things Journal*, vol. 9 no. 12, pp. 9712-9729, 2022. DOI: 10.1109/JIOT.2022.3142674

[4] Dhaya R., "Improved Image Processing Techniques for User Immersion Problem Alleviation in Virtual Reality Environments," *Journal of Innovative Image Processing*, vol. 2, no. 2, pp. 77-84, 2020. DOI: 10.36548/jiip.2020.2.002

[5] Getuli V., Capone P., and Bruttini A., "Planning, Management and Administration of HS Contents with BIM and VR in Construction: An Implementation Protocol," *Engineering, Construction and Architectural Management*, vol. 28, no. 2, pp. 603-623, 2021. DOI: 10.1108/ECAM-11-2019-0647

[6] Haghighat P., Prince A., and Jeong H., "Graph Convolutional Networks for Exercise Motion Classification," *in Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Baltimore, pp. 685-689, 2021. DOI: 10.1177/1071181321651255

[7] Javaid M. and Haleem A., "Virtual Reality Applications Toward Medical Field," *Clinical Epidemiology and Global Health*, vol. 8, no. 2, pp. 600-605, 2020. DOI: 10.1016/j.cegh.2019.12.010

[8] Ligorio G., Bergamini E., Truppa L., Guaitolini M., and et al., "A Wearable Magnetometer-Free Motion Capture System: Innovative Solutions for Real-World Applications," *IEEE Sensors Journal*, vol. 20, no. 15, pp. 8844-8857, 2020. DOI: 10.1109/JSEN.2020.2983695

[9] Long S., He X., and Yao C., "Scene Text Detection and Recognition: The Deep Learning Era," *International Journal of Computer Vision*, vol. 129, no. 1, pp. 161-184, 2021. DOI: 10.1007/s11263-020-01369-0

[10] Luo H., Li G., Feng Q., Yang Y., and Zuo M., "Virtual Reality in K-12 and Higher Education: A Systematic Review of the Literature from 2000 to 2019," *Journal of Computer Assisted Learning*, vol. 37 no. 3, pp. 887-901, 2021. DOI: 10.1111/jcal.12538

[11] MacDowell P., Jaunzems-Fernuk J., Clifford J., Ghani A., and Hoy B., "Virtual Reality in History Education: Instructional Design Considerations for Designing Authentic, Deep, and Meaningful Learning," *Journal of Applied Instructional Design*., vol. 14, no. 1, pp. 6-48, 2025, DOI: 10.59668/2033.19032

[12] Maihulla A., Yusuf I., and Bala S., "Reliability and Performance Analysis of a Series-Parallel System Using Gumbel-Hougaard Family Copula," *Journal of Computational and Cognitive Engineering*, vol. 1, no. 2, pp. 74-82, 2022. DOI: 10.47852/bonviewJCCE2022010101

[13] Meng Y., Shen J., Zhang C., and Han J., "Weakly-Supervised Hierarchical Text Classification," *in Proceedings of the AAAI Conference on Artificial Intelligence*, Honolulu, pp. 6826-6833, 2019. DOI: 10.1609/aaai. v33i01.33016826

[14] Mihcin S., "Simultaneous Validation of Wearable Motion Capture System for Lower Body Applications: Over Single Plane Range of Motion (ROM) and Gait Activities," *Biomedical Engineering/Biomedizinische Technik*, vol. 67, no. 3, pp. 185-199, 2022. DOI: 10.1515/bmt-2021-0429

[15] Minaee S., Kalchbrenner N., Cambria E., Nikzad N., and et al., "Deep Learning-Based Text Classification: A Comprehensive Review," *ACM Computing Surveys (CSUR)*, vol. 54, no. 3, pp. 1-40, 2021. DOI: 10.1145/3439726

[16] Norberg C. and Nordlund M., "A Corpus-Based Study of Lexis in L2 English Textbooks," *Journal*

of Language Teaching and Research, vol. 9, no. 3, pp. 463-473, 2018. DOI: 10.17507/jltr.0903.03

[17] Qiu S., Zhao H., Jiang N., Wu D., and et al., "Sensor Network Oriented Human Motion Capture Via Wearable Intelligent System," *International Journal of Intelligent Systems*, vol. 37, no. 2, pp. 1646-1673, 2022. DOI: 10.1002/int.22689

[18] Rafi K., Gani M., Hashim N., Rahman M., and Masukujjaman M., "The Influence of 360-Degree VR Videos on Tourism Web Usage Behaviour: The Role of Web Navigability and Visual Interface Design Quality," *Tourism Review*, vol. 80, no. 3, pp. 725-741, 2025. DOI: 10.1108/TR-06-2023-0383.

[19] Sachan D., Zaheer M., and Salakhutdinov R., "Revisiting Lstm Networks for Semi-Supervised Text Classification Via Mixed Objective Function," *in Proceedings of the Association for the Advancement of Artificial Intelligence AAAI Conference on Artificial Intelligence*, Hilton Hawaiian Village, Honolulu, pp. 6940-6948, 2019. DOI:10.1609/aaai. v33i01.33016940

[20] Stepanyan I. and Hameed S., "A Neuro Phenotypic Evolution Algorithm for Recognizing Human Motion Type," *The International Arab Journal of Information Technology*, vol. 21, no. 6, pp. 1015-1028, 2024. DOI: 10.34028/iajit/21/6/6

[21] Trost Z., France C., Anam M., and Shum C., "Virtual Reality Approaches to Pain: Toward a State of the Science," *Pain*, vol. 162 no. 2, pp. 325-331, 2021. DOI: 10.1097/j.pain.0000000000002060

[22] Wang X., Cheng M., Eaton J., Hsieh C., and Wu S., "Fake Node Attacks on Graph Convolutional Networks," *Journal of Computational and Cognitive Engineering*, vol. 1, no. 4, pp. 165-173, 2019. DOI: 10.47852/bonviewJCCE2202321

[23] Wright M., Twose D., and Gorter J., "Scootering for Children and Youth is more than Fun: Exploration of a Feasible Approach to Improve function and Fitness," *Pediatric Physical Therapy*, vol. 33, no. 4, pp. 218-225, 2021. DOI: 10.1097/PEP.0000000000000829

[24] Wu H., Ai C., and Cheng C., "Virtual Reality Experiences, Attachment and Experiential Outcomes in Tourism," *Tourism Review*, vol. 75, no. 3, pp. 481-495, 2020. DOI: 10.1108/TR-06-2019-0205

[25] Yildirim B., Topalcengiz E., Arikan G., and Timur S., "Using Virtual Reality in the Classroom: Reflections of STEM Teachers on the Use of Teaching and Learning Tools," *Journal of Education in Science Environment and Health*, vol. 6, no. 3, pp. 231-245, 2020. DOI: 10.21891/jeseh.711779

[26] Zhang J. and Mao H., "WKNN Indoor Positioning Method Based on Spatial Feature Partition and Basketball Motion Capture," *Alexandria engineering journal*, vol. 61, no. 1, pp. 125-134, 2022. DOI: 10.1016/j.aej.2021.04.078

Hongzhou Ma obtained his Bachelor's degree from the College of Physical Education, Henan University in 2010. He earned his Master's degree from the College of Physical Education, Henan University in 2013. From 2016 to 2018, he worked at Chongqing Vocational College of Culture and Arts. Since 2018, he has been working at Sichuan International Studies University as a lecturer. He has published three academic papers and participated in two scientific research projects.

Jiezhong Wu obtained his Bachelor's degree from the College of Physical Education, Jimei University in 2003. He obtained his Master's degree from the College of Physical Education and Sports Science, Fujian Normal University in 2011. He earned his Doctor of Education degree from Cavite State University in 2023. Since 2025, he has been studying at the School of Humanities, Beijing Sport University. He has been working at Fuzhou University as an associate professor since 2003. He has published 20 academic papers and 1 monograph, presided over 5 provincial-level research projects, and obtained 1 patent.