



RESEARCH ON THE APPLICATION OF VR TECHNOLOGY BASED ON HUMAN POSTURE RECOGNITION IN ONLINE PHYSICAL EDUCATION

FUYANG HE*

Abstract. Offline and current online physical education classes are characterized by low learning efficiency and poor teaching quality. To solve this problem, the study combines a dynamic time regularization algorithm and local image feature action recognition with virtual technology, which is applied to online teaching to improve teaching quality. Experimental results indicate that the overall model accuracy reaches more than 94%, and in the public dataset MSR, the model accuracy reaches 91.4%. In the UTK dataset, the accuracy reaches 85.8%. When the accuracy of pose recognition is compared, the research model is superior to the two traditional models. The validation results against KNN and SVM are 62% and 71% as well as 79% and 84%, respectively, and the experimental results of the research model have improved by 9% and 5%. The research technique achieves an overall fit of more than 90% for different parts in the analysis of motion capture, which is superior to the traditional online instruction. The overall teaching satisfaction reaches more than 93.5% in the comparison of online physical education courses with offline courses. In summary, the optimized and improved posture recognition system is better for human motion capture, and the applied virtual technology can effectively improve the learning effect and teaching quality of online physical education courses.

Key words: VR technology; Online teaching; Gesture recognition; Motion capture

1. Introduction. With the progress of science and the change of educational needs, the combination of Virtual Reality (VR) technology and online education has become an important direction in teaching in recent years. Especially in physical education, VR is changing the way of learning, providing new possibilities for improving the interaction and learning effect of distance teaching. However, research usually focuses on the technology of teaching platforms or the learning theoretical models. But they pay less attention to the application and optimization of human posture recognition based on VR in physical education, especially in solving the low learning efficiency and low teaching quality in online teaching [1-3]. In the traditional offline physical education, there are many challenges that affect teaching efficiency and quality. For example, due to the large number of students, teachers are often unable to provide personalized guidance and feedback to each student. Especially when learning complex motor skills, such as basketball dribbling or soccer kicking, students often need immediate and specific motion corrections to avoid inveterate learning of wrong skills. In addition, students may be reluctant to actively participate because of shyness, strong self-awareness, or physical limitations in physical education classes. These conditions not only reduce the quality of teaching, but also affect the acquisition of students' motor skills and the cultivation of healthy habits. At present, human posture recognition technology is a crucial part of VR physical education. Although many studies have focused on this, the existing pose recognition methods still face challenges such as incomplete motion capture and insufficient accuracy. These technical limitations reduce the effectiveness of VR physical education and hinder its widespread application in education. To this end, this research focuses on the limitations of existing human posture recognition technology in online physical education and tries to develop effective solutions to improve its performance [4-6].

The current literature shows that while VR in education has evolved, there is still limited research on its technical details and enhancement of learning outcomes. Some early studies have provided valuable insights and foundations. But due to the rapid evolution of technology, it is necessary to introduce the advanced methods to ensure research relevance and innovation [7-8].

In this study, the specific needs of the combined application of VR technology and online physical education are first identified. Then, through in-depth analysis of the limitations of the existing pose recognition

*School of Physical Education, Xichang University, Xichang, 615000, China (Fuyang_He2023@outlook.com)

technologies, the research aims at proposing an improved human body pose recognition system to effectively improve the quality of online physical education and learners engagement. Subsequently, the experimental validation of the proposed method is presented, and the model performance is compared and analyzed to prove its effectiveness in different scenarios. Finally, the practical significance of this study and its potential contribution to the development of online physical education in the future are discussed.

2. Related Works. With the development of information technology, posture recognition technology has also received much concern, and many researchers at home and abroad have studied this technology in-depth and applied it to different fields. Hong Zhen et al. proposed a collaborative solution based on artificial intelligence and Internet of Things (IoT) to solve the high rate of misjudgment in motion capture in healthcare. And they proposed a multi-pose recognition offline algorithm implemented on wearable hardware for recognizing poses based on multidimensional data. The results showed excellent performance in terms of accuracy and reliability [9].

Zhang et al. explored gesture recognition and behavior tracking in swimming motion images under computer machine vision. They incorporated the Gaussian mixture model and filtering algorithm into the recognition tracking system. The experimental results indicated that this model could obtain the motion trajectory of the object relatively completely, and the filtering tracking algorithm could obtain the motion speed of the object better [10].

Xiaoyun Tong et al. proposed a second-order pooled convolutional neural network for facial expression recognition to explore the correlation information between facial features after deep network learning. The results indicated that the correct rate was above 88%, which was superior to other algorithms [11].

MGR Alam et al. proposed an improved loMT for an emotion recognition system to study the recognition of human emotion states, and the experimental results showed that the performance of the proposed method was analyzed using benchmark datasets with high classification accuracy to determine human emotion states [12].

Research related to online teaching has grown more rapidly since 2020. Dolighan et al. studied the relationship between specific variables, teaching experience, professional development, teaching support, and the self-efficacy of teachers who transitioned to online teaching during the pandemic. Results indicated that higher online teaching efficacy scores were associated with taking online additional driving qualification and professional development courses. The highest online teaching efficacy scores were associated with prior use of board-provided learning management systems and the virtual technology [13].

Daumiller et al. analyzed the attitudes and fatigue of students and teachers who switched from offline to online teaching in learning. The study showed that teachers' learning goals were positively associated with viewing the transition to online teaching as a positive challenge and useful for their own competency development. It was also noted that teachers' goals and attitudes were correlated with successful online teaching [14].

Siegel et al. taught the knowledge of skin cancer to nurses through online teaching and divided them into two groups for a controlled experiment. The results showed that participants in the intervention group scored significantly higher on the "behavior" and "role" indicators, suggesting that the intervention was successful in influencing these dimensions [15].

Truzoli et al. explored faculty experience with online teaching, risk factors (e.g., stress) and protective factors (e.g., sources of control), and their impact on satisfaction during social distance. The results showed that satisfaction with online teaching was affected by changes in the environment and teaching style and that increasing satisfaction required better maintenance of a teaching mindset and adaptation to teaching style changes [16].

Kwong et al. set up a control experiment to compare the effects of offline and virtual labs. The results of the control experiment showed that students became more sensitive in applying parameter adjustment and backtracking strategies. Questionnaires indicated that students found such virtual labs satisfactory [17].

A Lester et al. set up a controlled experiment with Australian students to maximize students' motivation in physical education classes, enhance youth understanding of physical education, and increase their interest in physical education. The results showed that AMPED achieved modest improvements in MVPA compared to previous face-to-face-only interventions better. Online teacher training can help facilitate the widespread dissemination of professional learning interventions [18].

To improve online teaching, Z Fen proposed a bioimmune algorithm framework using the GBDT algorithm encoding to improve the efficiency of teaching English online. And a bagging learning-based stream feature selection algorithm was proposed to solve the problem that redundant information between features could reduce the accuracy of the framework. The results showed that the model constructed by the study had high reliability [19].

The research on online education and human pose recognition by domestic and foreign scholars show that there are numerous studies on both, but there is still much room for optimization of how to accurately identify and apply online education. Therefore, this study incorporates the dynamic time regularization algorithm and Harris corner point detection algorithm into pose recognition to explore the optimization and development of online education.

3. Construction of a human posture recognition model based on skeleton features and local image recognition.

3.1. Construction of pose recognition system based on skeleton features. The human body's pose feature is rich in semantic information, which is an important part of biometric features. Human pose recognition can be solved as a time-series problem by transforming an action into a pose sequence in space and modeling the action in space. A pose sequence is then viewed as a combination of skeletal frame data for action. There are many methods to identify actions by skeleton features extracted from human joint point data. In online physical education, the study uses joint body data features for human gesture recognition. The Dynamic Time Warping (DTW) algorithm mainly addresses the similarity of two temporally unequal sequences, which is first applied in fields such as speech recognition. Although the movement speed of two people is not the same in action recognition, the movement trajectory and the relative movement amplitude are similar. DTW can be used to solve the action judgment of the same action sequence with the different time that exists between teachers and students. The training sample X is set as the horizontal axis and the test sample Y is set as the vertical axis. Then a time series matrix is constructed, in which an element value in the matrix represents the distance between two points and the similarity of these two joints, and the calculation formula is shown in Equation (3.1).

$$d(X_m, Y_n) = \sqrt{\sum_{k=1}^K d(X_{mk}, Y_{nk})} \quad (3.1)$$

In Equation (3.1), X_{mk} denotes the feature value corresponding to the m frame in the training sample, Y_{nk} denotes the feature value corresponding to the n frame in the test sample, and K denotes the feature vector dimension of one frame. Therefore, the nodal similarity problem can be transformed into a shortest distance problem between these two points, as shown in Equation (3.2).

$$f(X_m, Y_n) = d(X_m, Y_n) + \min \{f(X_m, Y_{n-1}), f(X_{m-1}, Y_n), f(X_{m-1}, Y_{n-1})\} \quad (3.2)$$

In Equation (3.2), $f(X_m, Y_n)$ refers to the cumulative distance $d(X_m, Y_n)$ is the Euclidean distance of the current point location. Therefore, the action recognition of the skeleton features is to match the joint point positions of the test data with the joint point positions of the training data to achieve action recognition. The flow of DTW-based pose recognition is shown in Figure 3.1.

The body somatosensory is transmitted into the test data, the joint point data are pre-processed for feature value extraction, and the training data are matched with the DTW template to derive the required recognition results. The depth image is evaluated using the skeletal tracking technology in the Kinect sensor to obtain joint point information of different parts of the human body. When teachers and students enter the Kinect detection position, the instrument automatically detects and obtains the coordinates of human joint point positions, as shown in Figure 3.2.

In Figure 3.2, the instrument detects the human body information entering the range, the orientation of the sensor is Z-axis, the upward direction of the sensor is Y-axis, and the vertical Z-axis extends in the left direction is X-axis. The collected data are stored in Xef format. The position of the human body joints in the image and the coordinates of the actual three-dimensional position can be obtained by saving the data

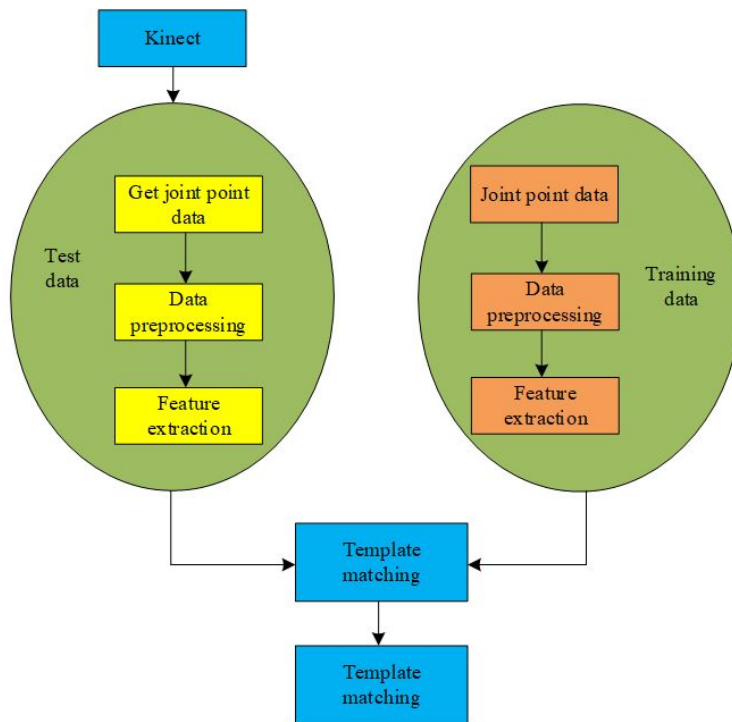


Fig. 3.1: Flow chart of attitude recognition based on DTW

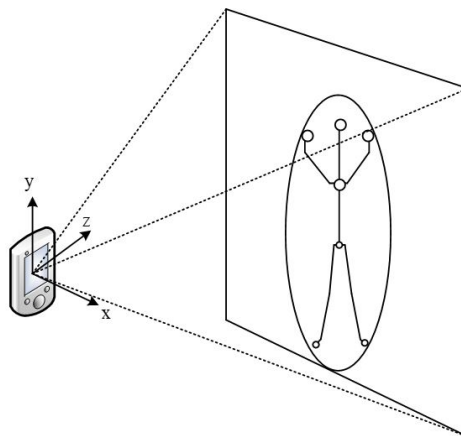


Fig. 3.2: Human joint detection

by frame through the file stream. After that, the data required for degree pre-processing are extracted by data segmentation and filtering with noise reduction. In the Kinect sensor range, a sequence of static and continuously changing images is used to represent the actions performed by a person. The motion recognition of the human body is performed by obtaining the coordinate data of the joint points to determine the activities of different parts. The raw skeleton data after pre-processing show incomplete inherent features of the action.

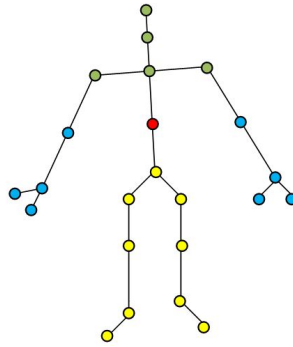


Fig. 3.3: Human skeleton model

To make the human pose model consistent with the actual action, the skeleton features are extracted from the sequence according to the displacement vector and relative position in the physical features. The definition formula of the displacement vector is shown in Equation (3.3).

$$v_i^s = \frac{p_i^{s+1} - p_i^{s-1}}{\Delta T} \quad 1 < s < n \quad (3.3)$$

In Equation (3.3), s refers to the skeleton sequence, the maximum of the sequence is n , p_i^s is the coordinate representation of the position of the joint in the sequence, ΔT indicates the interval between different times, and T refers to the number of skeleton frames in a sequence. The relative position describes the information of the human body and consists of the relative positions of different joints. The formula for calculating the relative positions of two joints in a sequence is shown in Equation (3.4).

$$j_{i,k}^s = p_i^s - p_k^s \quad i \neq k \quad (3.4)$$

In Equation (3.4), p_i^s and p_k^s represent the position coordinates of different joints in the same frame. The human skeleton consists of a fixed number of joints, and two skeletal models with 20 and 25 joints are used in this study.

In Figure 3.3, 25 skeleton models are counted into the neck as well as the fingers of the hand compared to 20 skeleton models, which can focus on different regions of the action metrics according to different needs. To get more representative skeleton features, the feature extraction method of global and local feature fusion is used to constitute the feature vector. Equation (3.4) can be transformed into Equation (3.5) required for the study.

$$f_{joint} = p_i(x_i, y_i, z_i) - p_1(x_1, y_1, z_1) \quad (3.5)$$

In Equation (3.5), $p_1(x_1, y_1, z_1)$ represents the location of the torso in model b of Figure (3.3), $p_i(x_i, y_i, z_i)$ is the 3D coordinates of the remaining skeleton points in model b of Figure (3.3), and f_{joint} is the new skeleton feature.

3.2. 3.2 Construction of recognition system based on local image features. In the pose recognition of skeleton features, the human body needs to be facing the sensor or completely back to the sensor to capture correct and complete joint point data. When the sensor is captured sideways or with different occlusion, incomplete joint acquisition and skeleton misalignment often occur, which do not accurately represent the human action features and reduce the feature differentiation. Therefore, local image features are added to the overall pose recognition to help describe the action features when the skeleton is occluded and to improve the action information under different scene interactions. In local feature extraction, it is most important to detect

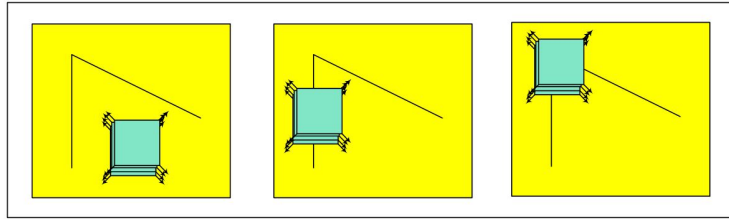


Fig. 3.4: HCD detection simulation diagram

the interest points in the temporal and spatial regions that can provide motion information and differentiation when the human motion mutates. So the study uses Harris Corner Point Detection Algorithm (HCD) for image interest point detection. The detection schematic diagram of HCD algorithm is shown in Figure 3.4.

The HCD algorithm uses a fixed local window moving in any direction to observe different grayscale situations. Detecting a significant sliding grayscale value change at any point in any direction is considered as the presence of a corner point. In the spatial domain, when the window is shifted, the gray value changes as shown in Equation (3.6).

$$E(u, v) = \sum_{u,v} w_{u,v} [f(x + u, y + v) - f(x, y)]^2 \tag{3.6}$$

In Equation (3.6), (u, v) represents the window offset, $f(x, y)$ and $f(x + u, y + v)$ represent the grayscale values of the pixel points before and after the shift, and $w_{u,v}$ represents the Gaussian window function. The gray value formula can be changed to Equation (3.7) by expanding the pixel point gray value Taylor.

$$E(u, v) \cong [u, v] M \begin{bmatrix} u \\ v \end{bmatrix} \tag{3.7}$$

In the human skeleton model, when the interest point data are sparse, iteration can be used for adaptive scale selection and cannot be accelerated, which can detect the required number of interest points in some scenes also to avoid excessive noise interference. When the simple background is too coefficient of interest points, it will lead to the lack of action features, which will have an impact on the recognition results. Therefore, Histogram of Orientation Gradient (HOG) and Histogram of Optical Flow Field Orientation (HOF) are chosen to describe the local information of interest points in detail. In the HOG, the RGB image is converted to grayscale map and then the X-axis and Y-axis gradient components are obtained by convolution operation using the set matrix template, and the calculation formula is shown in Equation (3.8).

$$\begin{cases} G_x = f(x - 1, y) - f(x + 1, y) \\ G_y = f(x, y - 1) - f(x, y + 1) \end{cases} \tag{3.8}$$

In Equation (3.8), $f(x, y)$ represents the grayscale value of the pixel at the corresponding position. G_x and G_y represent the horizontal and vertical gradients at the corresponding position. HOF is the amplitude weighted statistics of the optical flow and the horizontal axis angle and encodes the feature vector of the image, which reflects the change of the gray value of the pixel point as time passes. HOF is calculated as shown in Equation (3.9).

$$f(x + u\Delta t, y + v\Delta t, t + \Delta t) \tag{3.9}$$

In Equation (3.9), u is the component of optical flow motion in the horizontal direction at the corresponding moment of frame space position (x, y) . v is the component of optical flow motion from the vertical direction at the corresponding moment of frame space position. t is the corresponding moment. When the interest points

are characterized, the bag-of-words model is used for template matching. And the K-means clustering method is chosen as the bag-of-words model, and the cost function of K-means is Equation (3.10).

$$J = \sum_{i=1}^c J_i = \sum_{i=1}^c \left(\sum_{k, x_k \in G_i} \|x_k - c_i\|^2 \right) \quad (3.10)$$

In Equation (3.10), $J_i = \sum \|x_k - c_i\|^2$ denotes the intra-class objective function.

When using DTW for pose recognition and local image features for action recognition, it is necessary to temporally synchronize the two types of data to ensure that the pose recognition results are fused in the same time. Therefore, the decision fusion of Bayesian algorithm is investigated to fuse the recognition results of both classifiers to improve the speed and accuracy of the classification system. When two sets of skeleton data are obtained, the decision fusion using Bayesian algorithm is expressed by Equation (3.11) when the recognizers are independent of each other.

$$p(s|\omega_k) = p(s_1, s_2, \dots, s_L|\omega_k) \prod_i^L p(s_j|\omega_k) \quad (3.11)$$

In Equation (3.11), $p(s_j)$ is the probability that the classifier will classify the sample x in the class s_j .

In general, as a classical time series analysis method, DTW is used to measure the similarity between two time series. It allows the sequence to stretch or compress elastically across the timeline to achieve the best match. The key steps of DTW are as follows.

The construction of a Cumulative distance matrix, namely cost matrix. A matrix of $N \times M$ is constructed for two time series, where N and M are the lengths of the two time series, respectively. Each element (i, j) of the matrix represents the distance between point i of sequence A and point j of sequence B. Next is to find the optimal path. Starting from the upper-left corner of the cumulative distance matrix and ending in the lower-right corner, it needs to find a path that minimizes the cumulative distance on that path. This path shows how the two time series are aligned. Then the final distance is calculated. The overall similarity between two time series is obtained by summing the distance values on the optimal path. Harris corner detection algorithm is a classical algorithm used for corner detection in image processing. It is based on the change of gray level generated by local window movement to determine the corner points. The key steps of the algorithm are as follows.

First it is the calculation of the gradient values of each point of the image. This step is usually achieved using the Sobel operator or other image derivative filters. Next it is the calculation of the second moment matrix of the gradient. For each pixel, the second moment of the gradient is calculated, which reflects how much the image changes within the window of that point. Then the Harris response function is applied. The Harris response value is calculated for each pixel, with high response values usually corresponding to corner points.

4. Performance verification and empirical analysis of motion capture system based on DWT algorithm and local image feature action recognition.

4.1. DWT algorithm and local image feature action recognition algorithm performance verification. In the gesture recognition experiment, two public datasets, MSR and UTK, and a self-collection dataset SCD were used. The MSR dataset collected 8 common actions in teaching scenes as a complete Kinect collection process. UTK dataset was a complete Kinect acquisition process performing 5 common teaching gestures by the same person. The SCD dataset was a complete Kinect acquisition process with the same person performing five common actions such as measurement and recording. The MSR dataset was collected from multiple online teaching scenarios, covering eight common teaching actions. The dataset consisted of 100 participants from different backgrounds, each performing each action five times, for a total of 4,000 action samples. Each sample contained the time series data of the keyframe and the corresponding skeleton model coordinates. In the data preprocessing stage, the standardization process was carried out to eliminate the influence of attitude scale, and the noise filtering algorithm was applied to reduce the random error in the acquisition process. The UTK

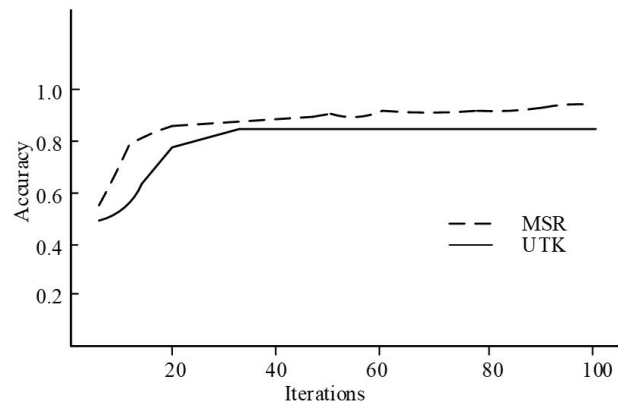


Fig. 4.1: Recognition accuracy of network model in MSR and UTK datasets

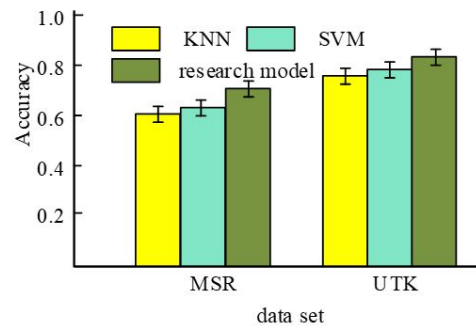


Fig. 4.2: Accuracy comparison of different attitude recognition methods

dataset consisted of five instructional gestures performed by a single participant, designed to capture gesture accuracy. The dataset contained 1500 samples, each recording the three-dimensional motion coordinates of gestures and the corresponding time stamp. The interference of irrelevant background action was eliminated by filtering algorithm, and the motion trajectory was smoothed to improve the accuracy of subsequent analysis. Self-collection dataset SCD was designed specifically for the needs of this study, recording 5 kinds of actions such as measurement and recording. 20 students participated in the collection, and each performed each action 10 times, which contained 1000 action samples in total. In addition to the use of standardization and filtering processing, data enhancement techniques such as rotation and scaling were introduced to improve the generalization ability of the model.

In Figure 4.1, during the iterations of the MSR dataset, the accuracy of the model increased substantially to more than 80% when the iteration started at 10. After about 60 iterations, the accuracy became stable gradually. After 100 iterations, the accuracy of the model reached 91.4%. Compared with the MSR dataset, the overall accuracy of the UTK dataset was slightly lower, with a significant increase in accuracy in the first 20 iterations and a gradual decrease in accuracy change after about 50 iterations, but still by an increase. Finally, it reached an accuracy of 85.8% after 100 iterations. Compared with the traditional recognition rate, the research model had effective convergence on the recognition process of different human postures, and the recognition accuracy was significantly improved. To verify the effectiveness of the improved model in pose recognition, a comparison experiment based on skeletal feature vectors for behavior recognition was designed on two public datasets with K-nearest Zero Classifier (KNN) and Support Vector Machine (SVM). The experimental results are shown in Figure 4.2.

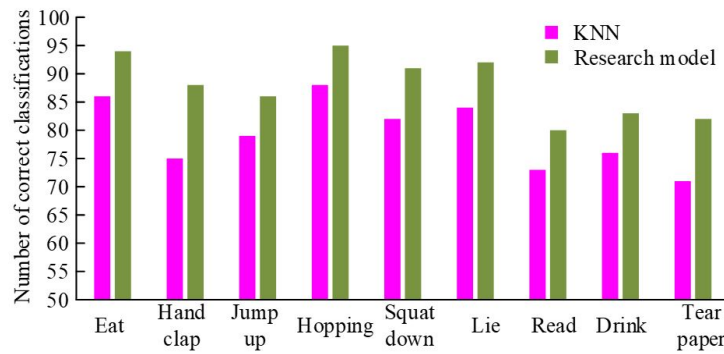


Fig. 4.3: Comparison of the number of different actions identified by the model in the MSR dataset

From the experimental results in Figure (3.6), the accuracy of the research model was superior compared to the two traditional models for pose recognition. The validation results compared with KNN and SVM were 62% and 71% as well as 79% and 84%, respectively. The experimental results of the research model were improved by 9% and 5%. The reason is that the research model can capture and recognize behavioral features at different time lengths using interest point detection, and the improvement is more obvious for more complex human gesture behaviors. The better ability to characterize action features with mask occlusion allows the model to learn with better robustness characteristics. Therefore, the research model has better performance and higher recognition accuracy compared to KNN over SVM. And to verify whether the research model can have good gesture recognition effect in different scenes, representative behaviors with different time complexity in life are selected as tests and compared with KNN model for analysis and verification, and the experimental results are shown in Figure 4.3.

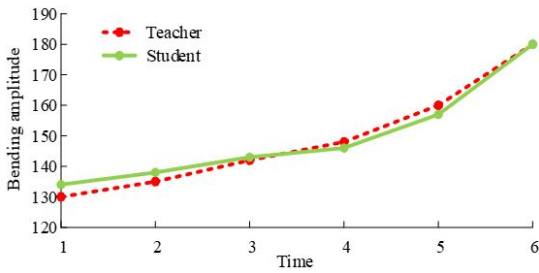
Figure 4.3 shows the comparison of the number of different categories recognized by KNN and the research model on the MSR dataset. The overall trend showed that the recognition of different actions by the research model was generally higher than that of the traditional human posture recognition. In the recognition of hand clapping, the traditional model was 13 lower than that of the research model. In the Hopping recognition, both models could recognize the action effectively, and the KNN model had a recognition rate of 88, and the research model had an advantage in the recognition of the fully frontal action, reaching a recognition rate of 95. In the recognition of objects with occlusion, such as the three actions of reading, drinking and tearing paper, the research model with HOG and HOF was more sensitive to grayscale changes. And its dense interest point detector was also more sensitive to complex motion backgrounds, which could reduce the amount of computation and data redundancy. So the recognition system can effectively improve the action recognition rate and better real-time performance. It can better capture semantic information in different poses, has stronger descriptive ability for human pose, and has more advantages with other human recognition systems in certain complex action recognition.

To evaluate model performance more comprehensively, accuracy, precision, recall, and F1 scores were calculated to reflect the model's recognition performance on different datasets. Precision refers to the proportion of samples correctly identified as positive examples to all samples identified as positive examples. Recall refers to the proportion of samples correctly identified as positive examples to all samples that are actually positive examples. The F1 score is a harmonic average of precision and recall, which provides a single metric to evaluate the balance of these two aspects. Table 1 lists the performance specifications.

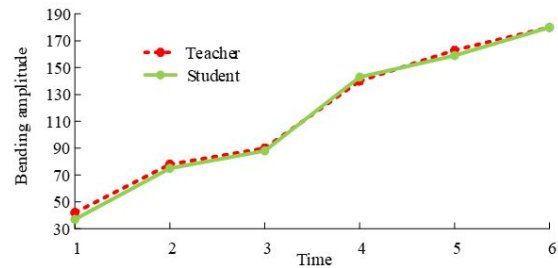
In Table 4.1, three different datasets were used to evaluate the performance of the model, including the four key metrics of accuracy, precision, recall, and F1 score. The accuracy of the model on the MSR dataset was 91.4%, with a slightly higher precision of 92.5%, meaning that 92.5% of all samples identified as positive cases were actually positive cases. The recall was 89.9%, indicating that the model could capture most of the actual positive cases. The F1 score, namely the harmonic average of accuracy and recall, was 91.2%, showing

Table 4.1: Performance evaluation

Datasets	Accuracy	Precision	Recall	F1 Score
MSR	0.914	0.925	0.899	0.912
UTK	0.858	0.87	0.835	0.852
Self-collection dataset	0.94	0.953	0.918	0.935



(a) Leg Range Comparison when shooting



(b) Arm Amplitude Comparison when shooting

Fig. 4.4: Analysis of the action capture of the shooting action by teachers and students

that the model maintained a good balance between them. Compared to the MSR dataset, the performance on the UTK dataset was slightly lower, with 85.8% accuracy, 87.0% precision, 83.5% recall, and 85.2% F1 score. Finally, on the self-collection dataset, the model showed the highest performance with 94.0% accuracy, 95.3% precision, 91.8% recall and 93.5% F1 score. These results show that the model has high recognition accuracy on different datasets, especially on self-collection datasets.

In general, through comprehensive analysis of different models, although KNN is simple and easy to implement, it is limited by computational efficiency when dealing with large datasets, and its recognition performance is very sensitive to the size of selected neighbors. The SVM model performs well in high-dimensional spaces, especially for datasets with significant gaps, but when the datasets contain more overlap, the SVM's performance may decline, and its parameter adjustment process may be complicated.

4.2. Empirical analysis of the research model motion capture system. In online physical education, the human posture recognition system allows that student to understand the teacher's power points of certain movements from the data, three-dimensional, and other levels, instead of the traditional offline. This can be achieved through observation and thinking about the normality of the action by themselves. So the disassembly teaching process of shooting action in basketball training is chosen as the empirical evidence of motion capture. The experimental results are shown in Figure 4.4.

In Figure 4.4, the teacher's disassembly process of the shooting action and the time of student imitation are shown, disassembling the shooting posture for students to experience the change of body amplitude during shooting. By the two pose folds, the study utilized the limb tracking accuracy, namely the joint angle, as the algorithm performance index. In (a), for the change of leg bending amplitude between the teacher and the student during the shooting, the teacher's joint angle in the shooting demonstration was easier for the students to imitate and understand. And the students bent more than the teacher in the initial action, changed their subsequent posture in the middle of the shooting, and finally reached the same action with the teacher in the end, with an overall fit of more than 90%. In (b), for the change in arm bending amplitude during the shooting, the overall curve fit reached more than 96% because the arm bending was more easily perceived by themselves. Therefore, the above results show that the research improved pose recognition can separate the magnitude of the action, improve the anti-interference ability of the sensor, and accurately capture the correct motion pose.

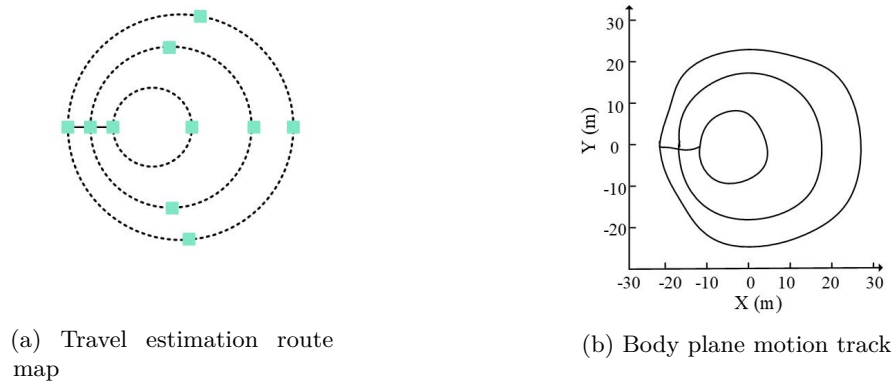


Fig. 4.5: Route walking experiment results

It not only has a higher accuracy rate but also has a lower error angle than the traditional pose recognition method, thus achieving accurate motion capture.

Figure 4.5 shows the verification of whether the studied human posture recognition system can correctly locate the human position and correctly capture the position according to the joint nodes. The study conducts experiments on the lasso route for the improved model, and from the experimental results. In Figure 4.5 (a), the proposed walking trajectory is shown, with the repeated action of walking a large circle from the starting point and then walking a small circle inward. Figure 4.5 (b) shows the simulation of the torso motion trajectory of the human posture recognition system for the human model. The range of movements captured by the research improved model was similar to the actual walking trajectory of the set of circles, showing that the research model could correctly identify the human body movements while effectively capturing the general path, and its accuracy of capture was high. The positional changes depicted were able to meet the teaching needs for teaching real-world simulations. In the comparative analysis, three key performance indicators, including learning satisfaction, learning effect, and teaching efficiency, were adopted to comprehensively evaluate the application effect of human posture recognition system in physical education teaching. Quantitative questionnaires were used to evaluate the results on a scale of 1 to 100, with higher scores indicating better results. Learning satisfaction reflects students' satisfaction with course content and teaching methods. Learning effect refers to the degree to which students absorb and master knowledge. Teaching efficiency focuses on the speed and effect from teachers on students' understanding and application of knowledge. Through this evaluation system, a single subject scored each indicator independently, and the scoring results were comprehensively compared according to the average values of the two groups of samples. This scoring mechanism is used in three different teaching scenarios: online physical education based on DTW and local image feature action recognition, traditional offline physical education, and online physical education supported by traditional pose recognition technology. The results of the analysis visually reveal the potential impact of various teaching methods on improving learning experience and outcomes. The comparison results are shown in Table 2.

From the scores of the three teaching methods through the two groups of experimental subjects in Table 4.1, the DWT algorithm and local image feature action recognition-based system scored higher in learning effect, learning satisfaction, and teaching effect than traditional offline education and traditional online teaching. The effect of absorbing knowledge in physical education class was better than the other two teaching methods. In the DWT algorithm and local image feature action recognition, the average learning satisfaction score of the two experimental groups was 95, the average learning effect score was 92, and the average teaching effect score was 93.5. The results show that the improved human posture recognition system is more effective for online physical education teaching, more popular among students than traditional offline physical education classes, and has better teaching effects and quality for teachers.

Table 4.2: Comparison of different teaching satisfaction

/	/	Research model	Offline teaching	Traditional online teaching
First group	Learning satisfaction	94	70	86
	Learning effect	91	78	84
	Teaching effectiveness	96	69	88
Second group	Learning satisfaction	90	72	87
	Learning effect	93	75	86
	Teaching effectiveness	91	71	89

Table 4.3: Algorithm comparison

Peculiarity	RM	KNN	SVM	DT	Instructions
Accuracy	0.914	0.88	0.9	0.85	The accuracy of this method on MSR dataset exceeds that of other systems, showing better recognition ability.
Real-time	Intermediate	High	Low	Intermediate	Although KNN has better real-time performance, this research method provides a balance of accuracy and response time.
Computational efficiency	High	Intermediate	High	Low	The method in this study uses efficient algorithm optimization, which makes the model run fast even on complex datasets.
Generalization	Strong	Intermediate	Strong	Weak	Through testing on multiple datasets, this method shows strong generalization ability.
Ability to handle complex movements	Outstanding	Good	Normal	Poor	This research improves the recognition ability of complex actions by introducing local features and deep learning techniques.
User adaptability	Strong	Weak	Intermediate	Intermediate	This research method performs well in user adaptability, adapting to learners with different body types and behavioral characteristics.
Ability to cope with occlusion situations	Strong	Weak	Strong	Normal	By combining various sensor inputs and advanced image processing technology, this method can better deal with the sight occlusion problem.
Complexity of experimental setup	Low	High	Strong	Low	Compared with KNN and SVM, this research method does not require complex experimental Settings and reduces the threshold of users in real application scenarios.
Expandability	High	Intermediate	Intermediate	Low	Due to the flexibility of the model architecture, this research method can be easily extended to different application scenarios and tasks.
Cost-effectiveness	High	Low	Intermediate	Intermediate	Considering the implementation cost and computing resources required, this research method provides a considerable cost-benefit ratio.

In the comparative analysis in Table 4.3, the pose recognition method achieved 91.4% accuracy on MSR dataset, which was better than 88.0% of SVM, 90.0% of KNN and 85.0% of DT. Although slightly inferior to SVM in terms of real-time performance, this method provided a balance of accuracy and response speed, reflecting a moderate compromise of real-time performance and accuracy in the recognition. The method also demonstrated strong generalization performance and robustness to motion occlusion, and surpassed DT in both user adaptability and ability to handle complex actions, proving its effectiveness in challenging environments. In addition, thanks to its low complexity experimental setup and high scalability, the cost effectiveness of the method in practical applications also showed competitiveness.

In summary, the pose recognition system in this study significantly improves student engagement by providing an immersive learning environment. Through virtual reality technology, the system allows students to

observe and imitate actions from a first-person perspective, creating a more interactive and interesting learning experience. In this way, students can more intuitively understand the correct posture and essentials of sports actions and get immediate simulation feedback, which greatly enhances the efficiency of content delivery. Students' satisfaction is improved not only because they are able to participate more actively in their learning, but also because they perceive a significant improvement in their personal skills.

5. Conclusion. VR technology has continuously come into the public's view. In education, teaching is a new way to effectively improve students' learning effects. To improve the effectiveness of online physical education teaching, the study proposed to integrate DWT and local image feature recognition into the human pose recognition system and applied the improved recognition system to online physical education courses. The experimental results show that the accuracy of the model can reach 91.4% in the public dataset MSR. In the UTK dataset, the model achieves 85.8% accuracy after 100 iterations. Compared with traditional recognition rates, the research model has effective convergence in the recognition of different human postures, and the recognition accuracy is significantly improved. Compared to the other two traditional pose recognition models, the research model has higher accuracy. The validation results compared with KNN and SVM are 62% and 71% as well as 79% and 84%, respectively, and the experimental results of the research model have improved by 9% and 5%. In the analysis of the disassembly of the shooting action, the degree of leg bending fit is more than 90%, the arm bending fit is more than 96%. The research with improved pose recognition can correctly separate the magnitude of the action and accurately capture the correct motion pose. If the improved system is applied to online physical education classes, all aspects of learning are better than traditional offline physical education classes, effectively improving teaching quality.

Despite the positive results of the study, there are certain limitations. For example, the real-time performance of current models may be limited when dealing with a large range of complex scenarios, especially in multi-user environments where performance has not been fully validated. In addition, the generality of the model across student groups of different ages and physical conditions still needs further study. In view of these limitations, in future work, the study plans to introduce more efficient algorithm optimization techniques, tests the model in a broader application field, and studies cross-population adaptation strategies. These measures are expected to provide strong support for improving the usability and popularization of the model.

Future research can be explored in the following directions. First, deep learning techniques have the potential to continue to improve the accuracy and real-time performance of gesture recognition. Especially, the latest neural network architectures are utilized to handle the recognition and prediction of complex movements. Second, the study of larger and more diverse data will help verify the generality and scalability of the model. In addition, empirical research on the psychological and cognitive impact of VR technology in education is also a necessary follow-up work, which will provide a deeper understanding and optimization direction for online physical education.

REFERENCES

- [1] Dong, Z. & Wang, X. An improved deep neural network method for an athlete's human motion posture recognition. *International Journal Of Information And Communication Technology*. **1**, 45-59 (2023)
- [2] Hong, Z., Hong, M., Wang, N., Ma, Y., Zhou, X. & Wang, W. A wearable-based posture recognition system with AI-assisted approach for healthcare IoT. *Future Generations Computer Systems: FGCS*. **2022**, 286-296
- [3] Chen, L., Hu, D. & Han, X. Study on forearm swing recognition algorithms to drive the underwater power-assisted device of frogman. *Journal Of Field Robotics*. **39**, 14-27 (2022)
- [4] Liang, L. Applied Research of VR Technology in Civil Engineering Teaching[C]// 2021 International Conference on Internet, Education and Information Technology 2021,4(IEIT):16-18.
- [5] Abrudan, C. & Horea, I. The Perks and Downsides of Teaching English Online. *Annals Of Faculty Of Economics*. **30** pp. 465-472 (2021)
- [6] Farra, S. & Hodgson, E. Evaluation of two simulation methods for teaching a disaster skill. *BMJ Simulation And Technology Enhanced Learning*. **7**, 92-96 (2021)
- [7] Yin, J. The method of table tennis players' posture recognition based on a genetic algorithm. *International Journal Of Biometrics*. **13**, 243-257 (2021)
- [8] Yang, H. Research on leg posture recognition of sprinters based on SVM classifier. *International Journal Of Biometrics*. **14**, 367-382 (2022)
- [9] Hong, Z., Hong, M., Wang, N. & Others A wearable-based posture recognition system with AI-assisted approach for healthcare IoT. *Future Generations Computer Systems: FGCS*. **2022**, 286-296

- [10] Zhang, Z., Huang, C., Zhong, F. & Others Posture Recognition and Behavior Tracking in Swimming Motion Images under Computer Machine Vision. *Hindawi Limited*. **202** pp. 1-9
- [11] Tong, X., Sun, S. & Fu, M. Data Augmentation and Second-Order Pooling for Facial Expression Recognition. *IEEE Access*. **(7)** pp. 86821-86828 (2019)
- [12] Alam, M., Abedin, S., Moon, S. & Others Healthcare IoT-Based Affective State Mining Using a Deep Convolutional Neural Network. *IEEE Access*. pp. 75189-75202 (2019)
- [13] Dolighan, T. & Owen, M. Teacher Efficacy for Online Teaching During the COVID-19 Pandemic. *Brock Education Journal*. **30**, 95-116 (2021)
- [14] Daumiller, M., Rinas, R., Hein, J. & Others Shifting from face-to-face to online teaching during COVID-19: The role of university faculty achievement goals for attitudes towards this sudden change, and their relevance for burnout/engagement and student evaluations of teaching quality - ScienceDirect. *Computers In Human Behavior*. **7**, 106677 (2021)
- [15] Siegel, V., Moore, G. & Siegel, L. Improving Nursing Students' Knowledge and Assessment Skills Regarding Skin Cancer Using Online Teaching Resources. *Journal Of The Dermatology Nurses' Association*. **2021**, 305-308
- [16] Truzoli, R., Pirola, V. & Conte, S. The impact of risk and protective factors on online teaching experience in high school Italian teachers during the COVID-19 pandemic. *Journal Of Computer Assisted Learning*. **37**, 940-952 (2021)
- [17] WongWing-Kwong, C. & Others Online Scaffolding for Data Modeling in Low-Cost Physical Labs. *International Journal Of Distance Education Technologies (IJDET)*. **17**, 1-20 (2019)
- [18] Lester, A., Owen, K., White, R. & Others An internet-supported school physical activity intervention in low socioeconomic status communities: results from the Activity and Motivation in Physical Education (AMPED) cluster randomised controlled trial. *British Journal Of Sports Medicine*. **53**, 341-347 (2019)
- [19] Fen, Z. Efficiency improvement of English online teaching system based on bagging learning flow feature selection. *Journal Of Intelligent & Fuzzy Systems: Applications In Engineering And Technology*. **4**, 6695-6705 (2021)

Edited by: Mudasir Mohd

Special issue on: Scalable Computing in Online and Blended Learning Environments: Challenges and Solutions

Received: Nov 23, 2023

Accepted: May 7, 2024