International Journal of Innovative Computing, Information and Control
Volume **20**, Number **6**, December **2024**

ICIC International ©2024 ISSN 1349-4198
pp. 1837–1850

KINECT BASED RECOGNITION AND DETECTION OF FITNESS QIGONG MOVEMENTS

Zhehua Fan[1] and Keshuang Sun[2,*]

[1]College of Physical Education
Putian University
No. 1133, Xueyuan Road, Chengxiang District, Putian 351100, P. R. China
fanzhehua@ptu.edu.cn

[2]Department of Physical Education and Research
Fuzhou University
No. 2, Wulong Jiangbei Avenue, Fuzhou University Town, Fuzhou 350108, P. R. China
*Corresponding author: sks2009@fzu.edu.cn

Received November 2023; revised March 2024

Abstract. *The continuous advancement of computer technology and sensory games has provided players with a richer experience, but in the field of sports and fitness, especially in the teaching and training of fitness qigong, how to accurately identify and correct movements remains a major challenge. Therefore, this study focuses on the recognition of human movements and proposes a Kinect based recognition and detection of fitness qigong movements. The purpose is to more accurately grasp the data of body scans and body contours of athletes to correct their incorrect body movements, thereby reducing the difficulty of movement and overcoming the difficulties of traditional sports training. Firstly, the pose recognition method based on the static K-means algorithm is used to calculate and recognize human skeletal joints. Secondly, the dynamic time warping algorithm is combined to recognize and detect gymnastics posture features, and experimental analysis is conducted on the effectiveness and applicability of recognition and detection techniques. The experiment shows that the recognition accuracy for stretching and chest expansion movements is 247 and 253, respectively, while the recognition accuracy for these two movements in the preprocessed dataset is improved to 259 and 263, respectively. It can be seen that this study not only improves the accuracy of motion recognition, but also promotes the development of personalized fitness guidance. The method proposed by the research institute provides coaches with real-time feedback tools, optimizes athlete training plans, and achieves the improvement of athlete technical level and sports effectiveness. Not only did it help popularize the scientific fitness concept, but it also improved the overall teaching quality and efficiency of fitness qigong.*
Keywords: Kinect, Action recognition, DTW algorithm, Artificial neural network, Static K-means

1. **Introduction.** With the development of information technology and sensory games, the automatic recognition of gymnastics and sports movements has become the focus of widespread attention, but it is difficult to combine real-time and accurately [1]. As a 3D video capture device, Kinect can capture and record human movements, and its high-precision recording and low-cost efficiency in human dynamic behaviors have shown great potential in many fields [2]. The application of Kinect can help coaches and athletes find problems in the early stage of movement execution and correct them in time, as well as help athletes better understand and recognize gymnastics movements [3]. As a 3D sensing technology with broad application prospects, it can collect a large number

DOI: 10.24507/ijicic.20.06.1837

of human movement data in real time, so as to better serve the teaching and training of gymnastics [4]. In addition, this technology can not only capture movement trajectories, but also help to accurately evaluate the standard and effect of movement execution, thus promoting the modernization and scientific practice of qigong [5]. Based on this background, Kinect technology combines static and dynamic motion analysis to identify and detect fitness qigong movements, in order to achieve efficient calculation and recognition of human bones and joints. The research innovatively combines deep learning with Kinect sensor data to fill the gap between traditional fitness qigong and modern intelligent recognition technology. The proposed method can provide more accurate and real-time movement feedback for qigong enthusiasts, so as to promote the development of fitness Qigong movement recognition and detection research. The content is divided into four parts. Part 1 introduces the background of somatosensory technology and its application in human motion detection. The second part is a literature review, which introduces the research and achievements of many researchers on somatosensory technology in recognition and detection. The third part studies the recognition and detection of fitness Qigong movements built on Kinect. The first section is based on the Static K-means algorithm (SK-m) for clustering and identifying actions. The second section proposes the Dynamic Time Warping (DTW) algorithm to solve the problem of misaligned human pose feature sequences. The fourth part is an experimental analysis of the recognition accuracy and detection accuracy of the proposed algorithm. Trainers can choose different gymnastics training programs, learn the necessary movements at any time, and promptly correct their mistakes. The study provides a series of quantitative empirical indicators through feedback and data collection in practical applications. These indicators provide an empirical basis for future research on action recognition accuracy, user interaction design, and technological iteration, thereby enriching the theoretical basis for subsequent research work.

2. **Related Work.** Kinect is a data scanning device based on body sensing technology from Microsoft, allowing people to easily communicate with machines with just physical movements or sounds. Kinect has increased the fun of Human-Computer Interaction (HCI) prompting many scholars to actively study it. Antico et al. studied a new Kinect designed for developers, the Azure Kinect DK, to achieve application specialization. Compared to previous versions, the hardware of the new Kinect has undergone significant improvements. Compared with professional motion tracking systems, this system can achieve non-invasive and low-cost tracking, making it suitable for implementing home rehabilitation systems [6]. Song et al. proposed a dual attention directed network for automatic detection of action units, which uses spatial and channel attention models to selectively extract and integrate global and local depth features from the semantic level. In this network, features are aggregated through a fusion module designed to build an end-to-end action unit detection system. Experimental results show that the network achieves the most advanced performance in the field of image-based motion unit detection with F1 scores of 64.0% and 62.6% [7]. Cerfoglio et al. used Microsoft Kinect cameras to evaluate individual gait after stroke, and used them to evaluate several pathological gait parameters. However, other studies have focused on validating kinect based measurements and gold standard references. The heterogeneity of participants, measurement methods, and research objectives makes it difficult to fully compare results, leading to uncertainty about the advantages and disadvantages of this technology in this pathological state [8]. Kumar and Kumar proposed that human motion recognition has become a hot topic in the field of artificial intelligence, such as video surveillance. This paper discusses the advantages and disadvantages of the public data set of human motion recognition and various methods of

dimensionality reduction, motion representation and analysis, and shows the application performance of deep learning technology in human motion recognition. Research results show that deep learning methods perform well in dealing with the complexity of human behavior recognition [9]. Zhong et al. proposed a multimodal network recognition scheme that combines a dual subnet structure with a self-concerned mechanism to optimize human behavior recognition of skeleton and depth data. The framework self-attention subnet based on transformer and the depth self-attention subnet based on CNN are integrated into the system, and the spatial characteristics of motion coordination and the quantitative criteria of joint motion contribution are introduced. The research results show that in the application of NTU RGB+D and UTD-MHAD datasets, the proposed method achieves recognition rates of 90.5% and 94.7%, respectively [10].

Based on device based Human Motion Recognition (HMR), dedicated camera equipment is used to calculate and process real-time images or image data of human motion data, thereby achieving the goal of recognizing human posture and motion. This study has become the focus of attention for many scholars regarding this system. There are many studies on action recognition among them. Pareek and Thakkar discussed distinctive ML and DL technologies for HAR, the characteristics of public datasets, various motion recognition technologies, and a survey of HAR applications. This study ultimately presented the pros and cons of action representation, dimensionality reduction, and action analysis methods [11]. Muhammad et al. proposed an attention mechanism based on bidirectional long-term and short-term memory. This mechanism uses an extended Convolutional Neural Network (CNN) to selectively focus on effective features in input frames to identify different human behaviors in the video. Use improved loss functions to achieve higher performance in action classification. The recognition rates on the UCF11, UCF sports, and J-HMDB datasets reached 98.3%, 99.1%, and 80.2%, respectively [12]. Mazzia et al. introduced action converters. It is superior to more complex networks that mix convolutional layers, loop layers, and attention layers. Utilizing 2D pose representation with small time windows provides a low latency solution for real-time performance. Compared with several state-of-the-art architectures, the results demonstrated the effectiveness of the ACT model and laid the foundation for future work on HAR [13].

In summary, Kinect, developed by Microsoft, has played a great role in human motion detection and has been widely used. Significant achievements have also been made in the field of HMR based on this device. However, this somatosensory algorithm is not mature enough to adapt to changes in movements, resulting in poor adaptability and a poor experience for gymnasts. Therefore, this study proposes a Kinect based study on the recognition and detection of fitness qigong movements to address this issue. It is hoped to assist trainers in correcting incorrect movements and improve their efficiency in learning standard movements.

3. Design of Gymnastic Movement Posture Recognition and Detection Algorithm Based on Kinect.

Gymnastics requires extreme precision in the movements and postures of athletes. Traditional methods for recognizing and detecting gymnastics movements often rely on the subjective judgments of athletes, coaches, and referees. However, traditional computer vision methods still face significant challenges in recognizing and detecting complex gymnastics movements. Microsoft's Kinect can address the impact of factors such as perspective and ambient lighting on gymnastics movement recognition. Therefore, this study is based on Kinect's algorithm design for Gymnastics Posture (Gym-P) recognition and detection, which will provide more accuracy and easiness to operate technical support for gymnastics teaching and training.

3.1. **Design of pose recognition method on the basis of SK-m.** HMR is a popular issue, because it can implement various types of application fields like monitoring systems, games, and HCI [14]. The significant progress in image processing technology has brought enormous benefits to the invention of Kinect sensors for visual recognition tasks, i.e., object recognition, healthcare monitoring, and HMR [15]. This system proposes a method for calculating human motion using skeleton-distance features and SK-m. This method utilizes bone and joint features to effectively develop and recognize gymnastics movements. The structure of the recognition system for human body movements is shown in Figure 1.

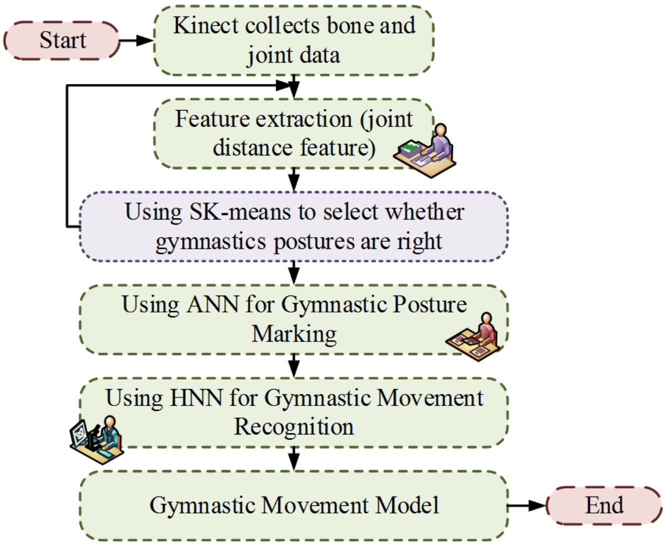


FIGURE 1. Flow chart of gymnastic movement recognition system

Use 3D bone and joint data as input for Kinect sensors, and perform characteristic analysis using joint distance features. Developing clustering through SK-m aims to optimize the effectiveness of Gym-P selection [16]. Then, Artificial Neural Networks (ANNs) are utilized to ensure the category labels for each Gym-P to improve performance and accuracy. Finally, a Hidden Markov Model (HMM) is combined to identify gymnastics movements from a known set of postures. The posture of human gymnastics movements in each frame has many positions of bone joints, and the mathematical expression is expressed as Equation (1).

$$H_t = \left\{P_t^1, P_t^2, \ldots, P_t^i\right\} \tag{1}$$

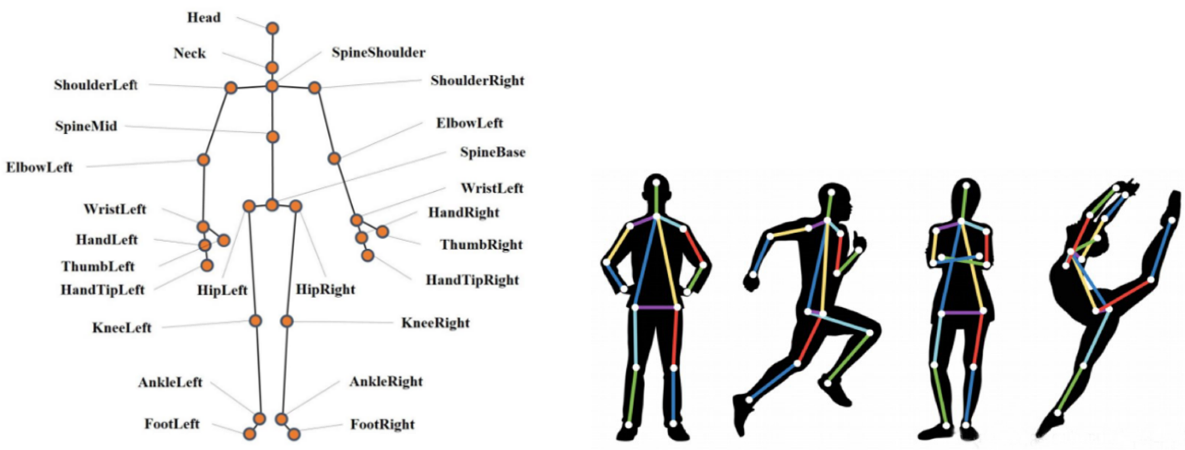
In Equation (1), P_t^i represents the position of the i-th joint at time t, and the 3D coordinates of each joint are x_t^i, y_t^i, z_t^i. The transformed joint coordinate expression is expressed as Equation (2).

$$P_t^{ki} = P_t^i - P_t^{hipcenter}, \quad 1 \leq i \leq N \tag{2}$$

In Equation (2), P_t^{ki} represents the same as P_t^i. N is the quantity of bone joints. The definition of the feature vectors for the Gym-P sequence of each skeleton frame f is expressed as Equation (3).

$$f = \left\{P_t^{k1}, P_t^{k2}, \ldots, P_t^{kN}\right\} \tag{3}$$

Use the K-means and use the squared Euclidean Distance (ED) measurement technique to reduce the similarity of Gym-P. Then use the K-means to reduce the repetition of Gym-P sequences and correctly determine the category markers for each Gym-P. Therefore, the depth profile and bone joints detected using Kinect sensors are shown in Figure 2.



(a) Kinect sensor bone joint names and positions

(b) Depth contours and bone joints detected through Kinect

FIGURE 2. Location diagram of activation function in neural network

The body movement relative to Kinect device related joints can be used to describe various movements of the human body. Skeletal joints allow for a constant distance between the position of sensors and the appearance of humans [17]. HMM for motion recognition can extract Gym-P by using the labeled ANN in a hidden state sequence. The component parameter is expressed as Equation (4).

$$\lambda = (\pi, A, B) \tag{4}$$

In Equation (4), π represents the probability set. The instantaneous conditions belong to different states, and the state at t is q_t. The expression of probability set $\pi = \pi_i$ is expressed as Equation (5).

$$\pi_i = P\left[S_i = q_t\right], \ 1 \leq i, t \leq N \tag{5}$$

In Equation (5), π_i represents the initial state of a state sequence with probability, which is assumed to be an equal possibility distribution of S_i. N represents different status. The transition-prob of state $A = \{a_{ij}\}$ from S_i to S_j is expressed in Equation (6).

$$a_{ij} = P\left[S_j = q_{t+1} \left| S_j = q_t\right.\right], \ 1 \leq i, j, t \leq N \tag{6}$$

In Equation (6), q_{t+1} represents the state at time $t+1$. Assuming that the amount of various observation symbols in each state is R, then $B = \{b_j(k)\}$ as expressed as Equation (7).

$$b_j(k) = P\left[u_k att \left| S_j = q_t\right.\right], \ 1 \leq j, t \leq N, 1 \leq k \leq R \tag{7}$$

In Equation (7), $b_j(k)$ represents the distribution probability of the observed symbols in state j. Train and test the corresponding HMM for each known Gym-P sequence, and recognize them grounded on the maximum posterior probability. The overview of the action recognition is shown in Figure 3.

This system deeply utilizes 3D bone joint data from Kinect sensors to recognize and analyze human movements. Firstly, feature extraction is performed on the data by examining the distance features between joints. Then, SK-m is used for feature clustering to improve the selection performance of Gym-P. Through multiple layers of complex calculations and recognition, capturing and recognizing human gymnastics movements demonstrates strong practicality and foresight.

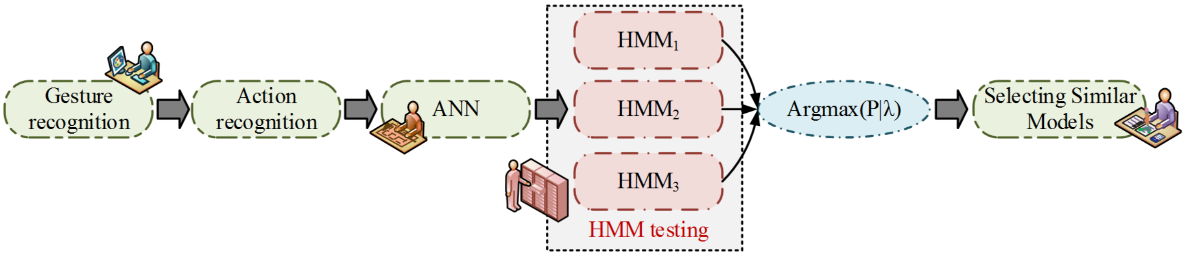


FIGURE 3. Gymnastic movement recognition process

3.2. **Design of recognition and detection algorithm based on action time regulation.** In order to construct a gymnastics movement model in space, it is necessary to transform the gymnastics movements into posture sequences represented [18]. To determine whether two sets of limb movements belong to the same gymnastics movement, it is necessary to evaluate based on the similarity of the waveform. Therefore, the dynamic time concept of DTW is taken to find the mapping connection among the points and two sets of gymnastics action sequences. The mapping relationship is shown in Figure 4.

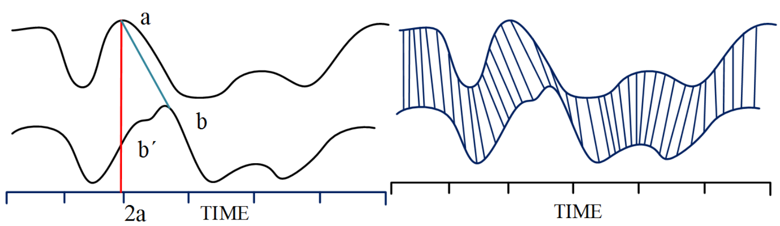


FIGURE 4. Mapping relationship between two sequences before and after regularization at a certain time point

The method of extracting motion features itself may show the chanllenges for alogning time series point, while the principle of DTW is primarily to handle the min distance between two sequences [19]. Therefore, this study constructs a matrix network with n rows and m columns through ED. The distance between two points i and j in the matrix is represented by $d(T_i, S_j)$ for similarity. Distance and similarity are inversely proportional. The calculation formula is expressed as Equation (8).

$$d(T_i, S_j) = \sqrt{\sum_{\omega=1}^{N} (T_{i\omega} - S_{j\omega})^2}, \quad 1 \le \omega \le N,\ N = 24 \tag{8}$$

Equation (8) represents the 24 dimensional pose feature vector at a certain point in time. N represents the dimension of the distance feature of Gym-P. $T_{i\omega}$ and $S_{j\omega}$ represent the distance feature values corresponding to the actions in frames i and j in different Gym-P sequences T and S. The mapping correlation between different T and S is expressed as Equation (9).

$$W = \{w_1, w_2, \ldots, w_{k'}, \ldots, w_K\},\ \max(m, n) \le K < m + n - 1 \tag{9}$$

In Equation (9), W is the planned path, and $w_k = (i, j)_k$ means the k-th point in it. The cumulative distance is the ED sum of points T_i and S_j set to $\gamma(i, j)$ and the distance of the nearest element that can reach that point. Among them, the value of $\gamma(i, j)$ is inversely proportional to the similarity. The formula for calculating the cumulative distance is expressed as Equation (10).

$$Y(i, j) = d(T_i, S_j) + \min\{Y(i-1, j-1), Y(i-1, j), Y(i, j-1)\} \tag{10}$$

In Equation (10), $(i+1, j), (i, j+1), (i+1, j+1)$ represent the three directional vectors through which point (i, j) passes. To calculate the likeness between two Gym-Ps, DTW is adopted for matching. Compare the input gymnastics action sequence with the action sequence in the standard template to obtain similarity. In the system architecture of the Kinect based gymnastics motion recognition system in this study, each module connects the previous and the following, presenting the most primitive motion data in the form of the final human body model animation [20]. The structure of the system is shown in Figure 5.

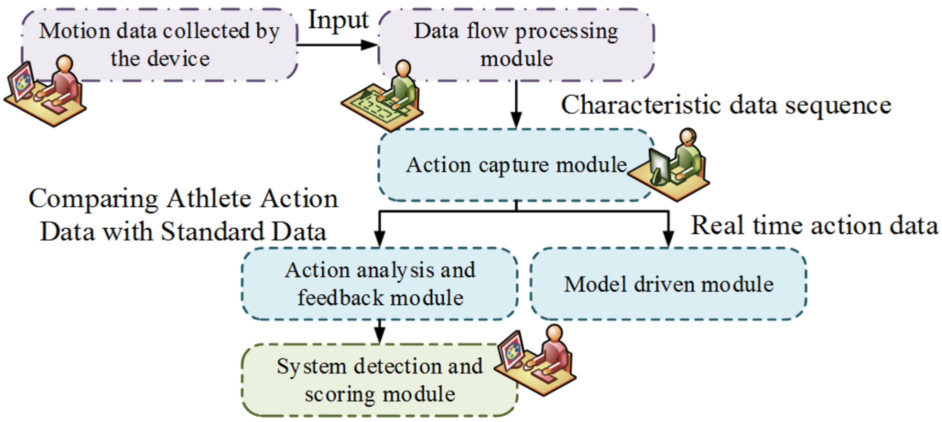


FIGURE 5. Relationship between information processing modules of gymnastic movement recognition system based on dynamic time regulation

Data processing converts the initial elements captured by Kinect, combined with time vectors, into an ordered frame sequence for use in the next step [21]. Assuming a set of different action sequences $M = \{S_1, S_2, \ldots, S_i, \ldots, S_M\}$, the calculation formula for the gymnastics action sequence to be tested is expressed as Equation (11).

$$Lable(T) = Lable(S_c),\ c = \arg\min \gamma(T_i, S_j),\ i = 1, 2, \ldots, M \tag{11}$$

In Equation (11), i and c represent the numbers of the i-th sequence in the database and the sample with the smallest distance in the template library. $\gamma(T_i, S_j)$ represents the similarity between the T and i action sequences. $Lable(S_c)$ represents the category of the action sequence corresponding to c. To avoid gymnastics actions that are not part of the gymnastics action template database being recognized as gymnastics actions, a threshold τ is pre-set, as expressed as Equation (12).

$$\tau = \max dtw(S_i, S_j),\ i = 1, 2, \ldots, M,\ j = 1, 2, \ldots, M,\ i \neq j \tag{12}$$

In Equation (12), $dtw(S_i, S_j)$ represents the DTW distance between each gymnastics action sequence in the template. Combine the DTW algorithm to match the corresponding frames of gymnastics action sequences, and design a scoring function module for gymnastics trainers. By calculating the angle difference between the gymnastics action sequence and the standard action sequence, the results are obtained as the basis for detection and the final result for scoring [22]. The design steps of the scoring algorithm are shown in Figure 6.

Among different Gym-Ps, there are sequences of Gym-Ps to be tested and standard movements [23]. By calculating the differences in the angles of gymnastics movements, the detection of movements is achieved, represented by the Manhattan distance, as shown in Equation (13).

$$D_{Angle}(A,B) = \frac{1}{n} \cdot \sum_{i=1}^{n} D^{i}_{Angle}\left(B'_i, B_i\right) \tag{13}$$

In Equation (13), n represents the sequence of actions. A, B represent the pose sequence and standard action sequence of the gymnastics movements to be tested, respectively. D_{Angle} represents the Manhattan distance. The calculation formula for the speed difference is expressed as Equation (14).

$$D_{Speed}(A,B) = \frac{1}{n} \cdot \sum_{i=1}^{n} D^{i}_{Speed}\left(B'_i, B_i\right) \tag{14}$$

In Equation (14), D_{Speed} represents the Manhattan distance. Based on the design of gymnastics motion recognition methods, the corresponding modules of the system are further elaborated, and a gymnastics motion recognition algorithm is designed using dynamic time regulation [24]. Gymnastic movement differences are calculated in both angle and speed dimensions to achieve movement detection, and further refinement is made to the data processing of the entire system.

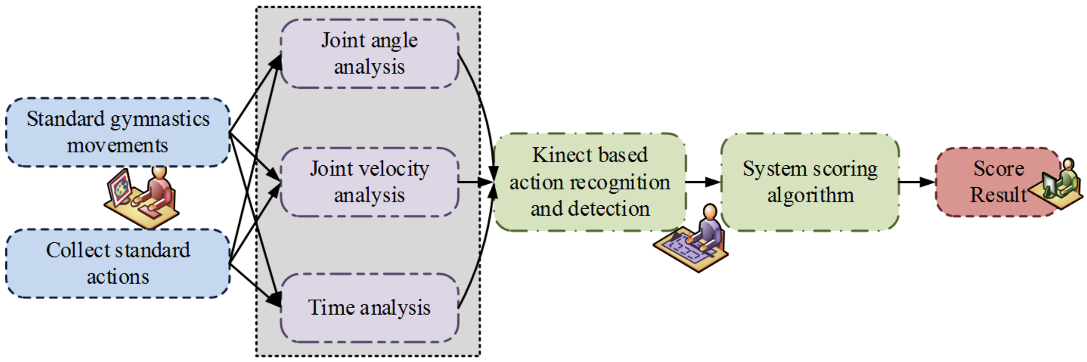


Figure 6. Gymnastic movement detection system and scoring module steps based on dynamic time regulation

4. **Experimental Analysis of Gymnastic Movement Recognition and Detection Performance Based on Kinect.** Exploring Kinect based methods for recognizing and detecting gymnastics movements, first analyze the limitations of these methods when dealing with gymnastics. Secondly, to further improve the recognition and detection performance of gymnastics movements, a new method based on deep dynamic images was designed. It uses an autoencoder to learn attitude feature representation and uses a Long Short-Term Memory (LSTM) network to model the timing information of motion actions. The effectiveness of the proposed method has been demonstrated through comparative experiments with other methods. Table 1 shows the parameters of the operating environment.

In Table 1, a machine equipped with Kinect for Windows V1.0 and Intel i5 processors is shown, with 12GB of RAM and NVIDIA GeForce GT750M graphics card. It runs Windows 8.1 Professional Edition, uses 3ds Max 2015 modeling, Unity 5.6 as the system platform, developed with Visual Studio 2012, and comes with Kinect Studio V1.0 driver. Table 2 compares the UTKinect training set with NSK-m and SK-m.

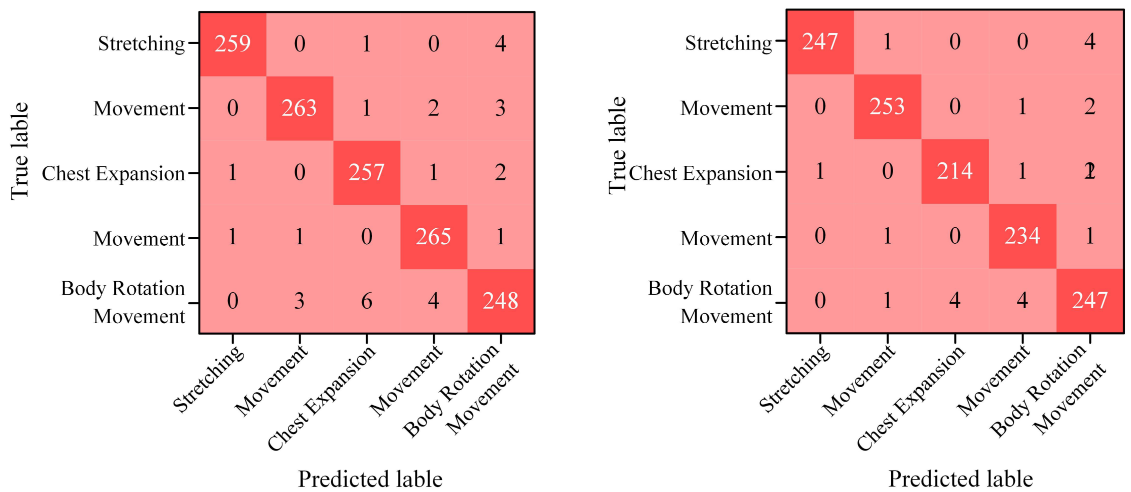
Table 2 shows the classification accuracy of various movements under different methods. For example, the accuracy of “stretching exercise” is 92% under Non-Static K-means (NSK-m), while it reaches 100% under SK-m. The accuracy of “body rotation motion” in NSK-m and SK-m is 37% and 95%, respectively. These data show that using this method

TABLE 1. Parameters of the operating environment

Category	Details
Kinect Setup	Kinect for Windows V1.0, including main machine, power cord, and data cable
Processor (CPU)	Intel®Core^{TM}i5-3230M@3.20GHz
Main Memory (RAM)	12GB
Graphics Adapter (GPU)	NVIDIA GeForce GT750M
Motherboard Interface	USB3.0
Operating System	Windows 8.1 Professional Edition
Modeling Software	3ds Max 2015
System Platform	Unity 5.6 (64-bit)
Development Tool	Visual Studio 2012
Device Driver	Kinect Studio V1.0

TABLE 2. Comparative data of UTKinect set and NSK-m and SK-m

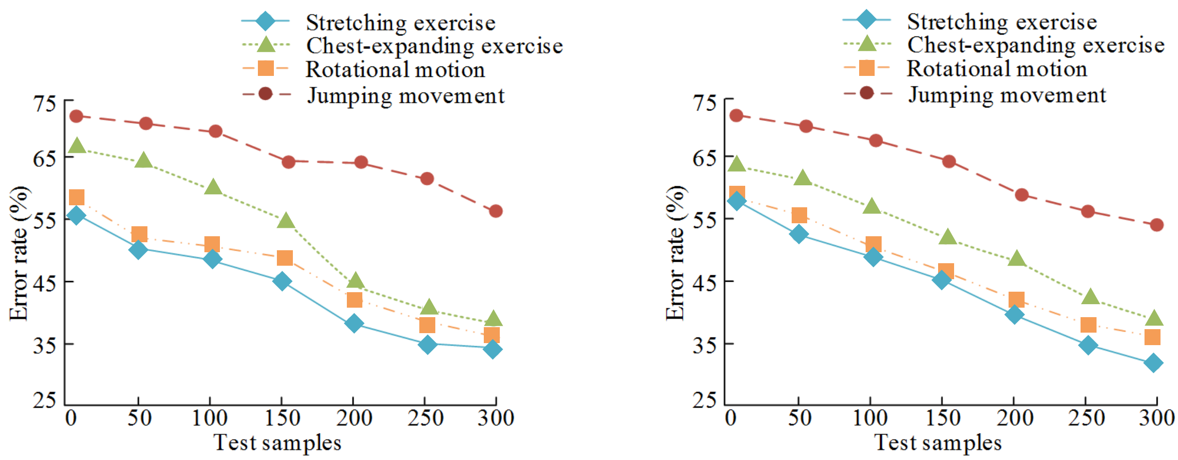
Action type	NSK-m (%)	SK-m (%)	NSK-m2 (%)	SK-m2 (%)
Stretching Movement	92	100	93	100
Chest Expansion Movement	54	100	56	100
Body Rotation Movement	37	95	40	96
Jumping Movement	82	95	85	97
Additional Movement	75	90	78	92
Extra Movement 1	70	85	73	88
Extra Movement 2	65	80	68	83



(a) Research method prediction results (b) Actual predicted results

FIGURE 7. Comparison of accuracy between actual actions and system recognition results

for action classification can achieve high accuracy, whether it is the original dataset or the preprocessed dataset. The preprocessed dataset has more accurate training results. By comparing actual actions and system recognition results, the accuracy of the system is evaluated and Figure 7 is obtained.



(a) No static

(b) Static state

FIGURE 8. The variation curves of four types of actions on the UTKincet under static and non-static conditions

In Figure 7, the recognition accuracy of the original dataset for stretching and chest expansion movements is 247 and 253, respectively. The recognition accuracy of these two actions in the preprocessed dataset has been improved to 259 and 263, respectively. This indicates that preprocessing the dataset can improve the recognition accuracy of the model. In addition, whether using the original dataset or the preprocessed dataset, the convergence speed of this method during the training process is fast, indicating that the model has good applicability. By comparing the actual action and system recognition results, the training results of the preprocessed dataset are more accurate, verifying that the preprocessing method can further improve the accuracy of action classification. After smoothing the cross entropy loss function in the table, the loss function of the dataset is shown in Figure 8.

In Figure 8, the error rates of four different types of motion states continue to decrease in both static and non-static states. This downward trend is very significant. In Figure 8(a), when the number of tests is 0, the error rates for the four different motion states are 73.68%, 67.94%, 57.33%, and 56.29%, respectively. However, when the number of tests increased, especially after increasing to 300, the highest error rate decreased to 57.89%, and the lowest error rate had already decreased to 34.12%. This reflects that as the number of tests increases, the accuracy of the model's recognition of motion states continues to improve. In Figure 8(b), when the number of tests is 150, the error rates for the four motion states are 67.34%, 56.71%, 47.38%, and 46.11%, respectively, increasing the number of tests to 300. The error rates of the four motion states showed a significant decrease again. The highest error rate this time has been reduced to 56.65%, while the lowest error rate has further decreased to 34.56%. This further proves that as the number of tests increases, the accuracy of successfully identifying four motion states continues to improve. Organize and analyze the balance of the LSTM based test model, as shown in Figure 9.

In Figure 9, the horizontal axis represents the iteration number of training, while the vertical axis represents the values of the model's balance indicators (such as F1 score) during training. There are significant differences in the balance performance of these three networks. The balance index of the optimal LSTM with Attention is 0.85, with a blackish green line of 0.78 and a dark blue line of 0.65. The speed of improving the balance of

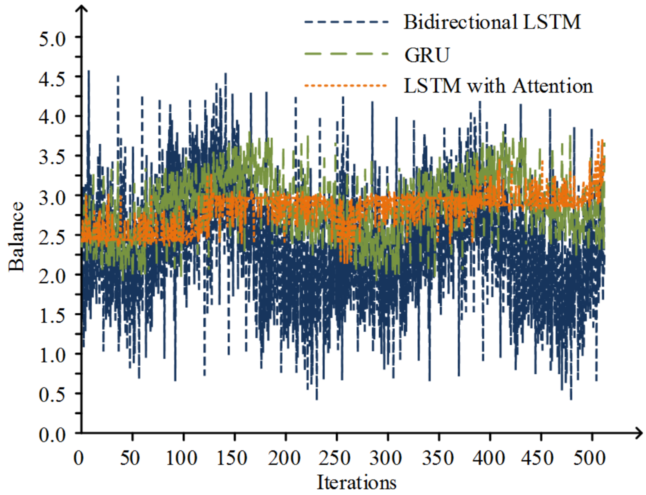


Figure 9. (color online) Balance analysis of test model based on LSTM network

Table 3. Comparison of UTKinect and Non-Static and SK-m

Action type	Accuracy of the 1st experiment	Accuracy of the 2nd experiment	Accuracy of the 3rd experiment	Accuracy of the 4th experiment	Accuracy of the 5th experiment	Accuracy of the 6th experiment	Accuracy of the 7th experiment	Accuracy of the 8th experiment	Accuracy of the 9th experiment	Accuracy of the 10th experiment	Average accuracy
Extend	83%	83%	67%	100%	83%	83%	83%	100%	100%	83%	87%
Chest enlargement	17%	17%	17%	17%	17%	0	17%	50%	17%	0	17%
Body rotation	67%	67%	100%	50%	67%	67%	50%	67%	50%	83%	67%
Jump	83%	83%	83%	83%	83%	83%	83%	83%	83%	83%	83%

these three neural networks also varies greatly. When the number of iterations is 500, the balance index of GRU is 0.73, the bidirectional neural network is 0.68, and the LSTM with Attention is 0.92. This indicates that among these three types of networks, LSTM with Attention has a faster speed of improving balance. While maintaining a higher balance index, its balance improvement speed is also faster.

In Table 3, the K-means has a higher recognition accuracy for NSK-m compared to the training set of the UTKinect dataset. For static and non-static training of the UTKinect, the accuracy of NSK-m is higher than that of SK-m by over 90 actions, and only 46 actions are lower than that of SK-m. This indicates that NSK-m can improve recognition accuracy by approximately 51%. When the amount of training data increases from a thousand to ten thousand, the advantage of NSK-m becomes more apparent. This is related to the K-means being more suitable for large-scale datasets.

5. **Conclusion.**

With the integration and progress of computer technology and kinesthetic games, more and more platforms and technologies related to kinesthetic games, as well as recognition and control technologies, are widely applied in people's daily lives. However, traditional motion recognition systems have issues such as high cost and risk, as well as the inability to label incorrect movements of athletes and urge them to correct them. Based on this, this study proposed a Kinect based study on the recognition and detection of fitness qigong movements in response to the problem of gymnastics movement recognition systems. Its aim was to simplify the gymnastics training process and help athletes efficiently learn various gymnastics events. The results of testing and evaluating the motion recognition and detection system demonstrated the classification accuracy of various movements under different methods. The accuracy of stretching exercise was 92%

under NSK-m, while it reached 100% under SK-m. The accuracy of body rotation motion in NSK-m and SK-m was 37% and 95%, respectively. In both static and non-static states, the error rates of four different types of motion states continued to decrease. This downward trend was very significant. When there was 0 test, the error rates for the four different motion states were 73.68%, 67.94%, 57.33%, and 56.29%, respectively. When the tests were 150, the error rates were 67.34%, 56.71%, 47.38%, and 46.11%, respectively. Increasing the tests to 300, the error rates of the four motion states showed a significant decrease again. The highest error rate this time had been reduced to 56.65%, while the lowest error rate had further decreased to 34.56%. This further proved that as the tests increased, the accuracy of successfully identifying four motion states continued to improve. Therefore, this technology efficiently assisted gymnasts in learning standard movements and provided targeted reminders and corrections. It also improved the effectiveness and efficiency of action learning, verifying the applicability and effectiveness of the system. However, in human bone modeling, when each limb action exceeds a certain range threshold, there will be overlapping phenomena at the bone joints. Therefore, further research and exploration are needed for the optimization of the system.

Acknowledgment. The research is supported by Fujian Province Higher Education Teaching Research Project, Innovation and Practice of Cultivating Applied Talents in Physical Education Based on OBE Concept (No. FBJY20230213).

REFERENCES

[1] A. Bawa, K. Banitsas and M. Abbod, A review on the use of Microsoft Kinect for gait abnormality and postural disorder assessment, *Journal of Healthcare Engineering*, vol.10, no.1, pp.1-19, 2021.

[2] T. M. Guess, R. Bliss, J. B. Hall and A. M. Kiselica, Comparison of Azure Kinect overground gait spatiotemporal parameters to marker based optical motion capture, *Gait & Posture*, vol.96, no.7, pp.130-136, 2022.

[3] T. Mahmood and Z. Ali, Analysis of Maclaurin symmetric mean operators for managing complex interval-valued q-Rung orthopair fuzzy setting and their applications, *Journal of Computational and Cognitive Engineering*, vol.2, no.2, pp.98-115, 2023.

[4] I-J. Lee, Kinect-for-windows with augmented reality in an interactive roleplay system for children with an autism spectrum disorder, *Interactive Learning Environments*, vol.29, no.4, pp.688-704, 2021.

[5] O. Elharrouss, N. Almaadeed, S. Al-Maadeed, A. Bouridane and A. Beghbadi, A combined multiple action recognition and summarization for surveillance video sequences, *Applied Intelligence*, vol.51, no.21, pp.690-712, 2021.

[6] M. Antico, N. Balletti, G. Laudato, A. Lazich, M. Notarantonio, R. Oliveto, S. Ricciardi, S. Scalabrino and J. Simeone, Postural control assessment via Microsoft Azure Kinect DK: An evaluation study, *Computer Methods and Programs in Biomedicine*, vol.209, no.9, 106324, 2021.

[7] W. Song, S. Shi, Y. Wu and G. An, Dual-attention guided network for facial action unit detection, *IET Image Processing*, vol.16, no.8, pp.2157-2170, 2022.

[8] S. Cerfoglio, C. Ferraris, L. Vismara, G. Amprimo, L. Priano, G. Pettiti, M. Galli, A. Mauro and V. Cimolin, Kinect-based assessment of lower limbs during gait in post-stroke hemiplegic patients: A narrative review, *Sensors*, vol.22, no.13, pp.10-49, 2022.

[9] R. Kumar and S. Kumar, Survey on artificial intelligence-based human action recognition in video sequences, *Optical Engineering*, vol.62, no.2, pp.2-21, 2023.

[10] Z. Zhong, Z. Hou, J. Liang, E. Lin and H. Shi, Multimodal cooperative self-attention network for action recognition, *IET Image Processing*, vol.17, no.6, pp.1775-1783, 2023.

[11] P. Pareek and A. Thakkar, A survey on video-based human action recognition: Recent updates, datasets, challenges, and applications, *Artificial Intelligence Review*, vol.54, no.1, pp.2259-2322, 2021.

[12] K. Muhammad, A. Ullah, A. S. Imran, M. Sajjad, M. S. Kiran, G. Sannino and V. H. C. Albuquerque, Human action recognition using attention based LSTM network with dilated CNN features, *Future Generation Computer Systems*, vol.125, no.10, pp.820-830, 2021.

[13] V. Mazzia, S. Angarano, F. Salvetti, F. Angelini and M. Chiaberge, Action transformer: A self-attention model for short-time pose-based human action recognition, *Pattern Recognition*, vol.124, no.4, 108487, 2022.

[14] M. A. Khan, Y. D. Zhang, S. A. Khan, M. Attique, A. Rehman and S. Seo, A resource conscious human action recognition framework using 26-layered deep convolutional neural network, *Multimedia Tools and Applications*, vol.80, no.1, pp.35827-35849, 2021.

[15] F. Afza, M. A. Khan, M. Sharif, S. Kadry, G. Manogaran, T. Saba, I. Ashraf and R. Damaseviclus, A framework of human action recognition using length control features fusion and weighted entropy-variances based feature selection, *Image and Vision Computing*, vol.106, no.1, pp.90-104, 2021.

[16] L. F. Yeung, Z. Yang, K. C. C. Cheng, D. Du and R. K. Y. Tong, Effects of camera viewing angles on tracking kinematic gait patterns using Azure Kinect, Kinect v2 and Orbbec Astra Pro v2, *Gait & Posture*, vol.87, no.6, pp.19-26, 2021.

[17] U. Özsoy, Y. Yıldırım, S. Karaşin, R. Sekerci and L. B. Suzen, Reliability and agreement of Azure Kinect and Kinect v2 depth sensors in the shoulder joint range of motion estimation, *Journal of Shoulder and Elbow Surgery*, vol.31, no.10, pp.2049-2056, 2022.

[18] G. Varol, I. Laptev, C. Schmid and A. Zisserman, Synthetic humans for action recognition from unseen viewpoints, *International Journal of Computer Vision*, vol.129, no.7, pp.2264-2287, 2021.

[19] C. Plizzari, M. Cannici and M. Matteucci, Skeleton-based action recognition via spatial and temporal transformer networks, *Computer Vision and Image Understanding*, vol.208, no.7, 103219, 2021.

[20] R. Hanif, S. Mustafa and S. Iqbal, A study of time series forecasting enrollments using fuzzy interval partitioning method, *Journal of Computational and Cognitive Engineering*, vol.2, no.2, pp.143-149, 2023.

[21] M. Yang and W. X. Wang, Recognition and movement of polystyrene nanoplastics in fish cells, *Environmental Pollution*, vol.316, no.2, pp.2-9, 2023.

[22] C. J. Little, M. Rizzuto, T. M. Luhring, J. D. Monk, R. J. Nowicki, R. E. Paseka et al., Movement with meaning: Integrating information into meta-ecology, *Oikos*, vol.22, no.8, pp.2-18, 2022.

[23] R. Peng, Y. Tian and S. Han, Multiuser wireless hand gesture recognition by spatial beamforming, *IEEE Transactions on Vehicular Technology*, vol.72, no.1, pp.805-820, 2023.

[24] Y. Wang, F. Chang, Y. Wu, Z. Hu, L. Li, P. Lang et al., Multi-Kinects fusion for full-body tracking in virtual reality-aided assembly simulation, *International Journal of Distributed Sensor Networks*, vol.18, no.5, pp.625-636, 2022.

Author Biography

Zhehua Fan obtained a Bachelor's degree in Sports Biology from Beijing Sport University in 1997 and a Master's degree in Sports Education and Training from Fujian Normal University in 2007. His research areas include sports human science, sports education and training, and sports rehabilitation. He is national level swimming referee, national level fitness qigong referee, national level social sports instructor, certified as a "sports prescription practitioner" by the Chinese Sports Science Society, responsible for teaching courses such as Sports Physiology, Sports Nutrition, Exercise Prescriptions, Practical Massage, and Fitness Qigong. The course of Sports Physiology was awarded the first tier offline course at the school level in 2020; In July 2016, he won the third place in the Six Character Formula collective project and the seventh place in the Five Animal Play men's individual project in the National Higher Education Fitness Qigong Competition; In August 2020, the Global Tai Chi Network Competition won the third prize in the 24 style Tai Chi. In August 2021, he was appointed as the Executive Director of Fujian Fitness Qigong Association. Recognized as a master's supervisor in December 2022, in January 2023, he was appointed as a member of the Expert Advisory Committee for the "Healthy Putian" Action. Guide students to participate in the Foot Fighter and Fitness Qigong events of the Fujian Provincial Games (College Student Group) and win 4 gold, 3 silver, and 6 bronze medals.

From August 1997 to December 1998, he participated in social practice at the Land Management Institute of Zhuangbian Town, Hanjiang District, Putian City; From January 1999 to March 2002, he taught in the Department of Physical Education at Putian Vocational College; From April 2002 to August 2005, he taught at the School of Physical Education of Putian University; From September 2005 to December 2007, he taught at the School of Physical Education of Putian University (studying as an in-service graduate student at Fujian Normal University); From January 2008 to July 2011, he taught at the School of Physical Education of Putian University; From August 2011 to present, he served as an associate professor at the School of Physical Education of Putian University (part-time research secretary of the School of Physical Education of Putian University from 2012 to 2015; part-time staff member of the Office of the Sports Committee of Putian University from 2017 to 2019).

He published over 10 papers, including 9 independently published academic papers; participated in 4 provincial and ministerial level projects, and 10 municipal and ministerial level projects, including leading 1 Fujian Provincial Social Science Fund project; participated in the writing of two textbooks and independently completed one monograph.

Keshuang Sun obtained a Bachelor's degree in Physical Education Teaching from Chengdu Sport University in 2006 and a Master's degree in Education Training from Fujian Normal University in 2009. Since 2009, he has been working in the Physical Education Teaching Research Department of Fuzhou University, as an Associate Professor. His research interests include physical education teaching and training, sports injuries and sports rehabilitation. He has published 8 academic papers, authored 1 textbook, and led and participated in 3 research projects.