

Article

Evaluation of Human Action Based on Feature-Weighted Dynamic Time Warping

Mingdie Yan, Xia Liu ^{*}, Zhaoyang Li and Naiyu Guo

School of Intelligent Manufacturing, Jiangnan University, Wuhan 430056, China; yan_mingdie@163.com (M.Y.); lil_zhaoyang@163.com (Z.L.); guo_naiyu@163.com (N.G.)

* Correspondence: liuxia@jhu.edu.cn

Abstract: Action evaluation can automatically detect abnormal actions by evaluating the quality of human actions in specific postures, which is widely used in the field of rehabilitation medicine. This paper proposes an intelligent rehabilitation action evaluation system to evaluate the quality of patients' actions during rehabilitation training, which helps medical professionals to more effectively monitor and guide the process, thus improving rehabilitation effects. Firstly, we collected human skeletal key-point data based on a depth camera and processed these data with gap filling and filtering; then, the effective data segments were segmented from the whole action dataset, angle and distance features were extracted, and the feature matrix was obtained; then, we used the Euclidean Barycenter Dynamic Time Warping–Barycenter Averaging algorithm to produce action templates; finally, we proposed a Feature-Weighted Dynamic Time Warping algorithm to calculate the similarity between the detected action and the template action and established an action achievement score mechanism to evaluate the rehabilitation action. The experimental results show that compared with the action evaluation method based on feature-matrix DTW, the proposed method significantly improves the similarity between healthy people and patients, and the similarity improvement for patients is more significant. Based on similarity scores, the difference between the actions of healthy people and patients and the template actions is more than 80%, which shows that the method can evaluate action quality in people with different health conditions and effectively reduce error in action evaluation. The confidence level of the action achievement score mechanism reaches 99%, which meets the actual application requirements.



Citation: Yan, M.; Liu, X.; Li, Z.; Guo, N. Evaluation of Human Action Based on Feature-Weighted Dynamic Time Warping. *Appl. Sci.* **2024**, *14*, 11130. <https://doi.org/10.3390/app142311130>

Academic Editor: Arkady Voloshin

Received: 20 September 2024

Revised: 18 November 2024

Accepted: 22 November 2024

Published: 29 November 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: action evaluation; IRAES; action segmentation; feature weighted; DTW

1. Introduction

In recent years, people have been seriously affected in terms of physical health and quality of life due to motor dysfunction accompanied by cervical spondylosis, stroke, Alzheimer's disease, etc. Additionally, the proportion of individuals experiencing motor function disorders due to various accidents is also on the rise [1]. Research has shown that recovering the ability to engage in activities of daily living as early as possible is the main goal of rehabilitation for patients, and specific daily movement training, such as activities of daily living training, is the main means of improving the probability of limb recovery [2].

The traditional action training program has the disadvantages of over-reliance on doctors' professional ability, boring and cumbersome processes, and low training efficiency [3]. With the development of technologies such as artificial intelligence and big data, intelligent rehabilitation medical treatment has received widespread attention [4]. Intelligent rehabilitation medical treatment records and analyzes the evaluation process, simplifying training and saving human resources through an intelligent action evaluation system, which has the advantages of high convenience, efficiency, and accuracy [5]. Therefore, the automatic assessment of daily movements in rehabilitation training can standardize assessment standards and is of great significance in providing training guidance [6].

Research on human action evaluation can be categorized into two types based on the method of data acquisition: that on wearable sensors and that on non-wearable sensors [7]. The former uses inertial sensors and potentiometers in wearable devices, which are cost-effective and free from angle occlusion issues but may cause discomfort and restrict natural action [8]. The latter employs vision sensors such as Kinect and Xtion PRO to provide natural, convenient, and non-intrusive 3D key-point data acquisition, which is essential to detecting abnormal actions and evaluating action quality [9]. Due to the cost and data advantages of non-wearable sensors, research in this area is flourishing. Ma used Kinect sensors to obtain depth images, proposed the use of multiple Kinect detection of human key points to solve the action occlusion problem, and realized action quality evaluation for rehabilitation training [10]. Bruce compared human key-point data obtained from Kinect v2 and the Vicon optical action capture system to achieve human action evaluation, including anomaly detection and quality evaluation [11]. Wang used an Azure Kinect camera (Microsoft, Redmond, WA, USA) to capture a 3D model of the human body and selected actions such as standing, stepping, leg lifting, and squatting to calculate joint angles, enabling knee function evaluation and exercise evaluation [12].

In terms of model building and algorithm research, more and more research in the literature explores action evaluation methods based on machine learning. Ding quantitatively evaluated features through multiple supervised learning models and employed end-to-end recurrent neural networks to improve action classification and evaluation [13]. Wang used OpenPose to extract human key points and angle features from RGB images and successfully implemented tennis action evaluation [14]. Li addressed the need for the automation of action evaluation in rehabilitation medicine and used time-domain filtering and a convolutional neural network to classify fine actions, achieving efficient action evaluation [5].

Most current studies rely on comparisons with manual predefined templates when quantifying action quality, which suffer from inconsistent time-series length. Some research has proposed time alignment algorithms such as Dynamic Time Warping (DTW) and its variants to address this problem [15]. Zhou proposed standard component analysis and extended it to standard time warping to solve the temporal alignment problem of human action sequences [16]. Gong combined stream shape learning and a new robust similarity metric to solve the temporal synchronization problem by using dynamic folding and warping, extending previous temporal alignment studies [17]. Fan constructed action standard datasets and designed a distance function combined with a DTW algorithm for key-frame extraction and action evaluation [18]. While these methods succeed in resolving the inconsistency of time-series lengths, they produce errors in calculating angles and distances between key points and do not consider weight in angle and distance features.

To address the above problems, in the second part of this article, an intelligent rehabilitation action evaluation system (IRAES) framework based on the Feature-Weighted DTW (FW-DTW) algorithm is proposed. Sections 3–5 present data acquisition and processing methods, action segmentation and feature extraction strategies, and template creation and action evaluation algorithms, respectively. Section 6 provides the experimental results. Finally, Section 7 concludes the paper.

2. Intelligent Rehabilitation Action Evaluation System (IRAES) Framework

A. IRAES framework

Rehabilitation therapy usually requires regular action training at a specialized medical center, which is often costly and time-consuming. Therefore, it makes sense to set up an action evaluation system at home or in an outpatient clinic. Under the guidance of a professional doctor, the quality of action training is fed back through the network, and the action evaluation is carried out by referring to the action rating scale, which enables the score of action achievement. This study utilizes the intelligent rehabilitation action evaluation system (IRAES) framework, which combines the traditional patient–clinician

assessment model and the new home monitoring assessment model [19], as shown in Figure 1.

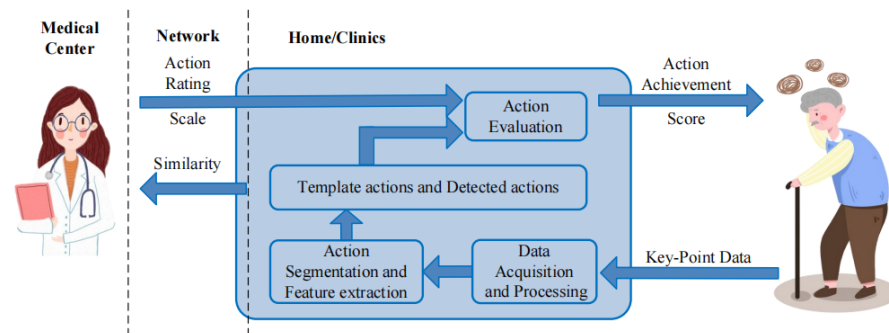


Figure 1. IRAES framework.

B. System Component

In order to establish a link between rehabilitation training and action evaluation, this paper leverages the IRAES framework, adopts the FW-DTW algorithm to evaluate daily upper-limb actions, and proposes an action achievement score method to enhance evaluation accuracy. The whole system mainly includes modules such as data acquisition, data processing, action segmentation, feature extraction, action evaluation, and achievement score, as shown in Figure 2. The Azure Kinect sensor was used to collect videos of four daily actions, that is, drinking water, combing hair, touching shoulders, and touching pockets, to obtain the 3D coordinate data of the key points of the human skeleton. Data filtering and gap filling algorithms were used for data processing. The three-axis variance and difference of the main key points were calculated through a sliding window to find the peaks and valleys, and the start and end points of the action segments were determined for action segmentation. Human structure vectors were constructed, the joint angles and inter-joint distances were calculated, and features were normalized to form a feature matrix. The Euclidean Barycenter Dynamic Time Warping Barycenter Averaging (DBA) algorithm was used to generate an action template, while the DTW algorithm matched the feature matrix to the action template sequence. The similarity between the two sequences was calculated, and a score mechanism was established to evaluate action performance.

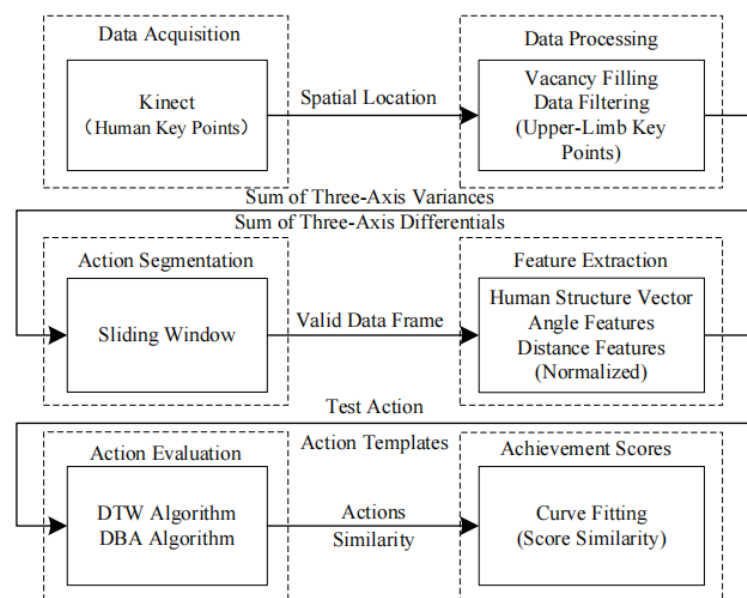


Figure 2. System composition.

3. Data Acquisition and Processing

A. Data acquisition

An Azure Kinect depth camera was used to collect 3D coordinate data of 32 key points of human skeleton, and the human key-point skeleton diagram composed of 32 key points is shown in Figure 3. This paper evaluates upper-limb action, including the nose, by selecting key points including the head, neck, shoulders, elbows, hands, wrists, thoracic spine, lumbar spine, spine base, and hips; these key points are highlighted with colored circles in Figure 4. Key points such as the eye, collarbone, fingertips, and lower limbs exert minimal impact on action evaluation. Since the hand and thumb are close, it is sufficient to choose one of the two, represented by a gray circle, which can reduce the data dimension and improve the efficiency of action evaluation. During the basic action, where the person remains stationary, the remaining six key points, shown in black circles, are selected as reference points for action evaluation.

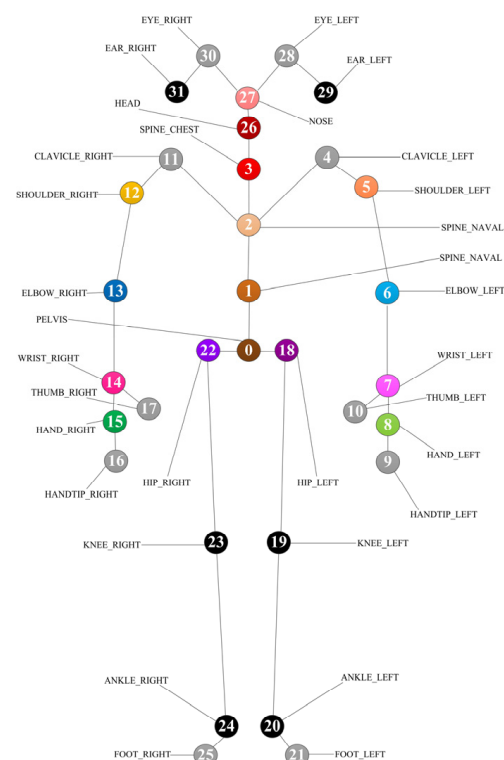


Figure 3. Skeleton diagram of human key points.

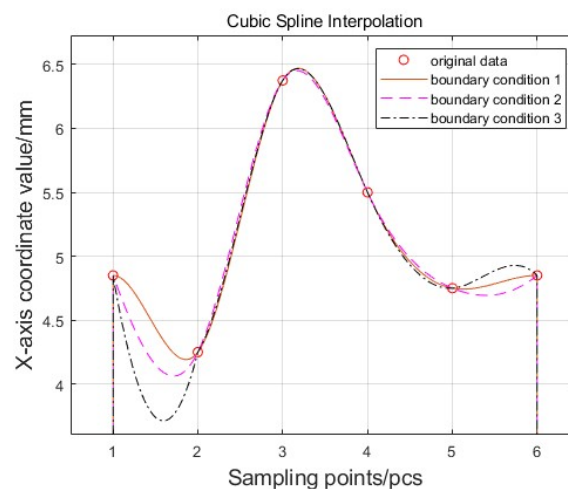


Figure 4. Cubic spline interpolation.

B. Data processing

During data collection, due to the occlusion of key points, ambient light, clothing worn, external interference, and equipment placement, there may be data gaps and noise, resulting in errors between experimental values and real values and loss of key information. In order to reduce the impact of these errors, gap filling and data filtering are required:

- (1) Gap filling: In this paper, segmented cubic spline interpolation is used to fill the gaps. The main idea is to divide the data series into several intervals and use cubic polynomials to fit the data in each interval, thus calculating the value and filling in the gaps. The fitted curve is represented by the segmented function $S(x)$; for example, $n + 1$ data points are given (x_i, y_i) , where $i = 0, 1, \dots, n, x_i \in [a, b]$. Segmentation is the division of the interval $[a, b]$ into n intervals $[(x_0, x_1), (x_1, x_2), \dots, (x_{n-1}, x_n)]$, where the left-end point $x_0 = a$, the right-end point $x_n = b$, and a cubic polynomial of the shape $y_i = a + bx_i + cx_i^2 + dx_i^3$ is constructed in each interval. $S(x)$ must meet the following conditions:
 - (1) If the interpolation condition is satisfied, every point in the interval passes through a cubic spline function, that is, $S(x_i) = y_i$, where $i = 0, 1, \dots, n$.
 - (2) In each interval $[x_i, x_{i+1}]$, there is a cubic polynomial.
 - (3) Curve smoothing: $S(x) \in C^2[a, b]$, $S(x)$, $S'(x)$, and $S''(x)$ are continuous over the interval $[a, b]$.

$S(x)$ can be solved by using three types of boundary conditions. Based on the cubic equation of $S(x)$, each interval contains four unknowns. For n intervals, this results in $4n$ unknowns, requiring an equal number of equations ($4n$) to solve. In total, $4n - 2$ equations can be obtained from the three satisfying conditions of the cubic spline function, and the remaining two equations can be obtained from the boundary conditions, which are of three types:

- (1) The second-order derivative of the specified endpoint, here M_0 and M_n , respectively. $S''(x_0) = M_0$, and $S''(x_n) = M_n$; in particular, when $M_0 = M_n = 0$, that is, $S''(x_0) = S''(x_n) = 0$, they are called natural boundary conditions.
- (2) The first-order derivative of the specified endpoint, here m_0 and m_n , that is, $S'(x_0) = m_0$ and $S'(x_n) = m_n$.
- (3) $S(x)$ is a periodic function with period $(b-a)$: $S^{(k)}(x_0 + 0) = S^{(k)}(x_n - 0)$, with $k = 0, 1, 2$.

Spline curves under three different boundary conditions are used, where boundary condition 1 is the second-order derivative of the specified endpoint, boundary condition 2 is the first-order derivative of the specified endpoint, and boundary condition 3 is a periodic function within the interval; the results are shown in Figure 4.

From the figure, the spline curves under all three boundary conditions exhibit noticeable variations at the endpoints, while the curves remain almost unchanged in the middle section. Boundary condition 1 (natural boundary condition), where the second derivative at the endpoints is set to zero, results in the smoothest and most natural transition at the endpoints. This condition avoids unnecessary oscillations and ensures a smooth, continuous curve that better reflects the natural motion trends, especially in applications like human body key-point data interpolation. Compared with first-order derivatives or periodic boundary conditions, especially for motion data, the velocity and acceleration at the endpoints usually have physical significance. Setting the second derivative at the endpoints (i.e., acceleration) allows the interpolation curve to better align with the actual physical motion laws. Additionally, experiments demonstrated that boundary condition 1 yields more accurate and realistic results in handling key-point drift and gap filling during fast movement, without introducing errors or unnatural changes. Its simplicity and computational efficiency further make it an ideal choice, especially when dealing with large data. Therefore, boundary condition 1 is chosen for segmented cubic spline interpolation in this study.

- (2) Data filtering: this paper adopts the method of moving average to remove noise from the key-point data. The filter window size is set to 1, 3, 5, and 7, and a certain key-point data in the X-axis direction are mean-filtered; the filtering effect is shown in Figure 5.

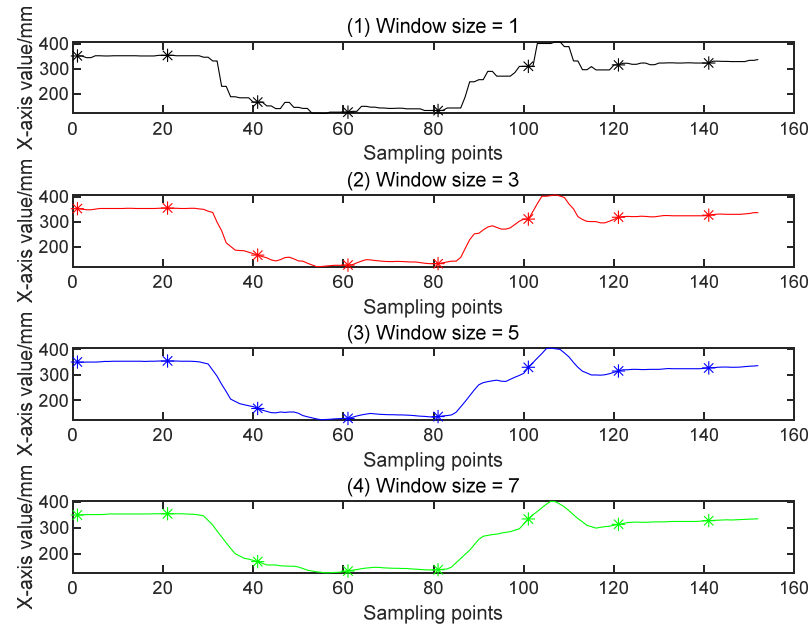


Figure 5. Moving average.

As shown in the figure, with the increase in the window size, the data noise points are gradually reduced, and the smoothness of the curve is gradually increased, but when the window size is 7, the original data are changed to a large extent, and the accuracy of the experimental results cannot be guaranteed. In order to eliminate the influence of data noise, drift, etc., without changing the accuracy of the original data, it is appropriate to select the window size of 5 for the moving average.

4. Action Segmentation and Feature Extraction Strategy

A. Action Segmentation

The collected action data included invalid actions, such as the preparation process at the beginning of the action and the static process at the end; the invalid information should be removed, and the effective action data should be segmented before action feature extraction to improve the efficiency and accuracy of action evaluation. The action segmentation process is as follows:

- ① Set the length of the window to w and the fixed sliding step to l , where the units are all the number of samples;
- ② Select hand key points, and calculate the sum of the X, Y, and Z three-axis differences E_k within each sliding window with the following equation:

$$E_k = \sum_{i=1}^w \left(|P_{k+i}^x - P_{k+i-1}^x| + |P_{k+i}^y - P_{k+i-1}^y| + |P_{k+i}^z - P_{k+i-1}^z| \right) \quad (1)$$

where E_k is the sum of the X-, Y-, and Z-axis differences of the hand key points at moment k ; P_{k+i}^x , P_{k+i}^y , and P_{k+i}^z are the hand key-point X-, Y-, and Z-coordinates of the i th point in the window at moment k , respectively; and P_{k+i-1}^x , P_{k+i-1}^y , and P_{k+i-1}^z are the X-, Y-, and Z-coordinates of the hand key points of the $i-1$ th point in the window at moment k , respectively.

③ Calculate the sum of the X-, Y-, and Z-axis variances, V_k , of the key points of the hand within the sliding window with the following equation:

$$V_k = \frac{\sum_{i=0}^{w-1} (P_{k+i}^x - \bar{P}^x)^2 + \sum_{i=0}^{w-1} (P_{k+i}^y - \bar{P}^y)^2 + \sum_{i=0}^{w-1} (P_{k+i}^z - \bar{P}^z)^2}{w} \quad (2)$$

where V_k is the sum of the X, Y, and Z three-axis variances of the hand key points at moment k , and $\bar{P}^x$, $\bar{P}^y$, and $\bar{P}^z$ are the mean values of X, Y, and Z in the window, respectively.

④ Draw the V -value curve graph, and look for the peaks in the V -value curve that meet the trajectory threshold condition of the peaks, constituting the trajectory sequence $V^p = \{V_1^p, \dots, V_n^p\}$, where $V_i^p > t_v$, $i, \{1, \dots, n\}$ and t_v is the V -value trajectory threshold, the setting of which can remove the interference generated by key-point judder.

⑤ Find the starting point of the action, and traverse the elements in the trajectory sequence V^p from front to back. If the horizontal coordinate of the current V_k^p value corresponds to the difference and $E_k > t_E$, then subtract the current sampling point from the start-point window length offset value m_s , that is, obtain the start point of the action; if the current V_k^p value of the horizontal coordinates corresponds to the difference and $E_k < t_E$, then the sampling point is judged to be an interference point, and continue to judge the next waveform. t_E is the threshold value of difference, and E is used to eliminate the interference generated by the judder of the key points of the hand.

⑥ Find the end point of the action, and traverse the elements in the trajectory sequence V^p from back to front. The judgement step is the same as in point ⑤; find the sampling point that meets the condition, add the sampling point to the end-point window length offset value m_e , and obtain the end point of the action.

B. Feature extraction

In each frame of the human body key-point data, 16 main key points are selected and combined in pairs to form 15 vectors representing the human body structure. In addition, two auxiliary vectors for calculating angles are added, which are from the left shoulder pointing to the left hip and from the right shoulder pointing to the right hip, for a total of 17 vectors. The projection of these vectors on the XOY plane is shown in Figure 6.

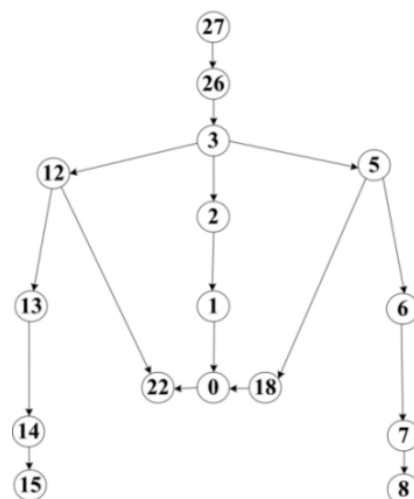


Figure 6. Vector diagram of human body structure.

Angle values are formed between structure vectors, and the changes in angle values reflect different action trends, so angle features can be extracted to evaluate actions. Through analysis, it is found that when different people perform the same action, the angle changes formed by key points such as shoulders, elbows, wrists, and hips are basically the

same. Therefore, in this paper, we select four angle features consisting of these key points to evaluate the actions, $\theta^k = (\theta_1^k, \dots, \theta_4^k)$, as shown in Table 1.

Table 1. Angle features.

Angle (Rad)	Composition of Key Points	Angle (Rad)	Composition of Key Points
θ_1^k	Left elbow–left shoulder–left hip	θ_3^k	Right elbow–right shoulder–right hip
θ_2^k	Left shoulder–left elbow–left wrist	θ_4^k	Right shoulder–right elbow–right wrist

Angle features can only reflect the action trend, and the distance features between different key points are also needed to supplement the description of action details for similar actions. Through analysis, it is found that the distance features between the hand and the nose, the head, the right shoulder, and the right hip are more obvious and can allow for distinguishing different details of the same action, so these eight distance features are selected and normalized in this paper, $d^k = (d_1^k, \dots, d_8^k)$, as shown in Table 2.

Table 2. Distance features.

Normalized Distance (mm)	Combination of Key Points	Normalized Distance (mm)	Combination of Key Points
d_1^k	Left hand–head	d_5^k	Right hand–head
d_2^k	Left hand–nose	d_6^k	Right hand–nose
d_3^k	Left hand–right shoulder	d_7^k	Right hand–left shoulder
d_4^k	Left hand–left hip	d_8^k	Right hand–right hip

Combining angle features and distance features to form a feature matrix can provide multi-dimensional action information for action evaluation and reduce the error of action evaluation. Assuming that the angle feature of the m th action sequence is $\mathbf{A}_m = (\theta^1, \dots, \theta^k)^T$, where k is the number of samples, the distance feature is $\mathbf{D}_m = (d^1, \dots, d^k)^T$, where k is the number of samples, the combination of angle and distance features forms the action feature matrix $\mathbf{F}_m = (\mathbf{A}_m, \mathbf{D}_m)$, as shown in the following equation:

$$\mathbf{F}_m = \begin{bmatrix} \theta^1 d^1 \\ \vdots \\ \theta^k d^k \end{bmatrix} \quad (3)$$

5. Template Creation and Action Evaluation

A. The overall framework of the action evaluation process

In action evaluation, in order to more objectively evaluate action achievement in the subject, it is necessary to produce action templates. In this paper, we use the DBA algorithm to create a unique action template for each type of action, which reduces the amount of computation and eliminates the chance of selecting action templates; then, we use the DTW algorithm to measure the similarity between the template action and the action to be tested, and we establish a score mechanism for action achievement based on the action rating scale [20] to realize the action evaluation; the flowchart is shown in Figure 7.

B. DTW algorithm

DTW is an algorithm that can compute the similarity of two time series of different lengths. It evaluates the similarity of two action sequences by dynamically adjusting the length of the action sequence to be measured and calculating the cumulative shortest distance from the template action sequence.

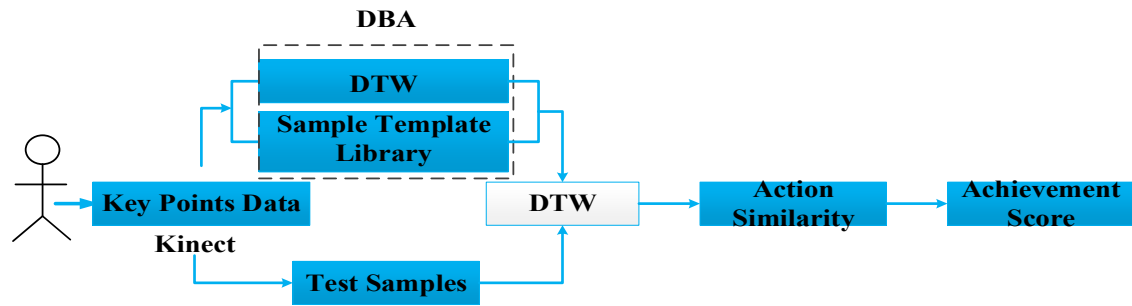


Figure 7. Overall framework of action evaluation process.

Let us suppose that there are two time series $x(i)$ and $y(j)$, where $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, n$. The warping path is denoted by W to represent the alignment or mapping of the time series $x(i)$ and $y(j)$, $W = \{w_k\}, k = 1, 2, \dots, p$, and p denotes the length of W . The warping path should satisfy three constraints of boundary condition, continuity, and monotonicity.

Construct an $m \times n$ lattice matrix with $x(i)$ and $y(i)$ sequence lengths with m and n as rows and columns, respectively, as shown in Figure 8. The DTW algorithm accumulates $d(i, j)$ to find the grid point with the shortest cumulative distance $\gamma(i, j)$ to plan the optimal path. The calculation of the cumulative distance $\gamma(i, j)$ of any point in the grid is shown in the following equation:

$$\gamma(i, j) = \begin{cases} d(i, j) & i = 1, j = 1 \\ \gamma(i-1, j) + d(i, j) & i > 1, j = 1 \\ \gamma(i, j-1) + d(i, j) & i = 1, j > 1 \\ \min(\gamma(i-1, j), \gamma(i, j-1), \gamma(i-1, j-1)) + d(i, j) & i > 1, j > 1 \end{cases} \quad (4)$$

where $d(i, j)$ is the Euclidean distance between elements in the $x(i)$ sequence and elements in the $y(i)$ sequence, $\gamma(i, j)$ is the cumulative distance of the grid at $d(i, j)$ from $(1, 1)$ to (i, j) , and $\min(\gamma(i-1, j), \gamma(i, j-1), \gamma(i-1, j-1))$ denotes the selection of the point with the smallest cumulative distance.

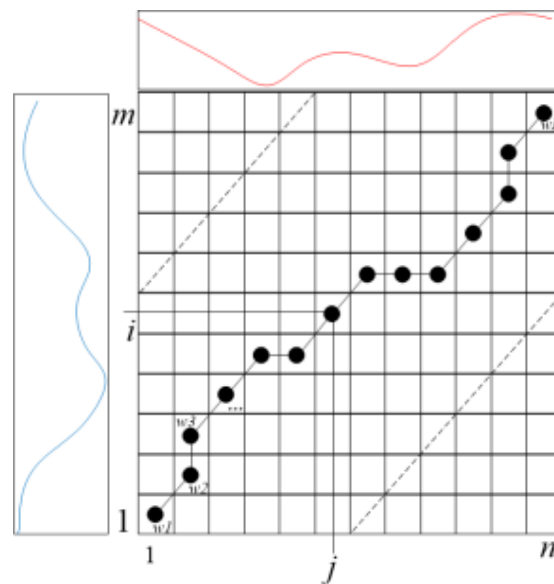


Figure 8. DTW warping path diagram.

In this paper, the tested action feature matrix F_m is replaced by $x(i)$, and the template action feature matrix F'_m is replaced by $y(j)$. The calculation of $d(i, j)$ in Equation (4) is

shown in Equation (5), since the feature matrix is a combination of angle features and distance features:

$$d(i, j) = \sqrt{\sum_{k=1}^l (F_m(i, k) - F'_m(j, k))^2} \quad (5)$$

where $F_m(i, k)$ is the k th feature value of the i th frame feature vector in F_m , $F'_m(j, k)$ is the k th feature value of the j th frame feature vector in the action feature matrix F_m , and l is the number of feature values.

C. Template creation

In this paper, the DBA algorithm is used to obtain the average time series of multiple action template time series of each type of action to obtain a unique template of each type of action, which reduces the matching calculation time between the action time series to be measured and multiple template action time series, improves the efficiency of action evaluation, and eliminates the contingency of selecting action templates.

The input of DBA is a set of time series, and the output is the average sequence of the set of series. The purpose of DBA is to calculate an average sequence that minimizes the sum of squares of the DTW distances to all sequence in the series set. DTW alignment is performed on each time series to obtain the aligned sequence, that is, each sequence is mapped to the same timeline. Calculating the barycenter point of each sequence, that is, the points on each sequence are mapped to the average sequence to obtain the initial value of the average sequence. Then, DTW alignment is performed on the average sequence to obtain the aligned average sequence, and the center of gravity of the average sequence is calculated, that is, the points are mapped on each sequence to the aligned average sequence to obtain the new value of the average sequence. If the average sequence value changes, the average time series needs to be updated repeatedly until the value of the average series converges. DBA is an iterative algorithm, and the flowchart is shown in Figure 9.

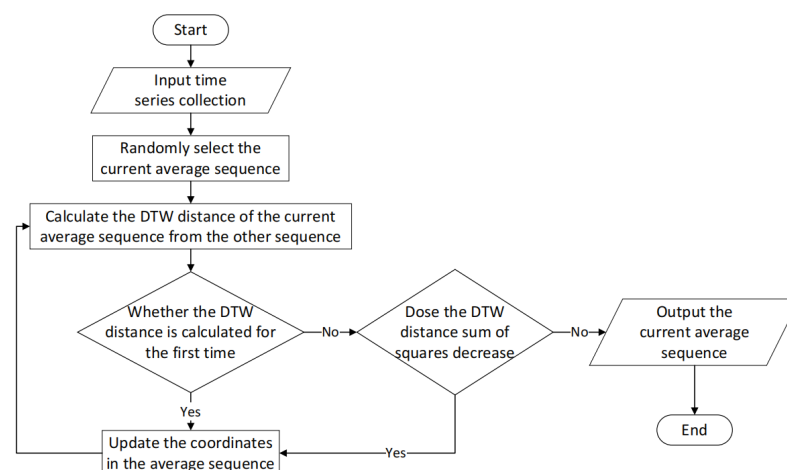


Figure 9. Flowchart of DBA algorithm.

D. Action Evaluation

- (1) DTW algorithm improvement: this paper improves the Euclidean distance calculation method in the DTW algorithm.

Euclidean distance calculation method based on feature weights: Due to data point drift when selecting key points during movement, it has a greater impact on angle features than distance features, and the error of action evaluation will be larger, especially when the actions are closer to each other. Therefore, this paper proposes the Euclidean distance calculation method based on feature weights, setting the weight share of angle features and

distance features in action features as w_1 and w_2 , respectively, and the sum of w_1 and w_2 is 1. Equation (5) is changed to the following equation:

$$d(i, j) = w_1 \times \sqrt{\sum_{k_1=1}^{l_1} (F_m(i, k_1) - F'_m(j, k_1))^2} + w_2 \times \sqrt{\sum_{k_2=1}^{l_2} (F_m(i, k_2) - F'_m(j, k_2))^2} \quad (6)$$

where $F_m(i, k_1)$ is the k_1 th feature value of the angle feature vector of the i th frame in the action feature matrix F_m , $F'_m(j, k_1)$ is the k_1 th feature value of the angle feature vector of the j th frame in the action feature matrix F'_m , and l_1 is the number of angle feature values; $F_m(i, k_2)$ is the k_2 th feature value of the distance feature vector of the i th frame in the action feature matrix F_m , $F'_m(j, k_2)$ is the k_2 th feature value of the distance feature vector of the j th frame in the action feature matrix F'_m and l_2 is the number of distance feature values.

- (2) Similarity calculation: Action evaluation requires the actions to be quantified, evaluated, and compared to other actions by similarity calculations. Similarity measures can be used to compare how similar different people are when performing particular actions. It can reflect an individual's action control and determine their level of performance when executing a specific action.

In this paper, DTW is used to calculate the distance between the action sequence to be tested and the template sequence, which is relative and cannot clearly reflect the degree of similarity between the two sequences. Therefore, after the cumulative shortest distance is obtained by DTW calculation, the similarity between the action sequence to be tested and the template sequence needs to be calculated in order to determine the degree of achievement of the action [21]. The specific calculation formula is shown in the following equation:

$$s = \frac{1}{1 + \frac{\gamma}{2k}} \quad (7)$$

where s denotes the similarity between the action sequence to be tested and the template sequence with a value in the range of $[0, 1]$, γ is the Euclidean distance between the action sequence to be tested and the template sequence, and k is the larger number of elements in the action sequence to be tested and the template sequence.

- (3) Achievement score: In rehabilitation medicine, it is crucial to observe a patient's motor performance and score his or her action achievement according to specific scoring criteria to evaluate the patient's performance in rehabilitation training. Scoring criteria usually include requirements such as correct action form, speed, balance, and coordination. Assessors can use a scoring system to determine the patient's level of rehabilitation and subsequently adjust the treatment plan to track the patient's progress based on the score results. This is an important aspect of rehabilitation medicine that aims to help patients return to a normal level of living and improve their quality of life. In this paper, the action performance is linked to the action score, and a six-score scale is used to score action performance under the guidance of a professional doctor and with reference to the relevant literature [22] and the Action Rating Scale [23]. This article uses a five-point scale.

In rehabilitation action evaluation, a more standardized score mechanism is important for rehabilitation trainers and can provide more reference for subsequent treatment. When using the FW-DTW algorithm to calculate action similarity, the closer it is to 1, the higher the degree of achievement of the action sequence to be tested relative to the template action. Therefore, this paper links action similarity with action score, sets the correspondence between action score and action similarity with reference to the action rating scale, and establishes a five-point action achievement score mechanism, where scores of 0 to 4 corresponds to different levels of similarity, as shown in Table 3.

Table 3. Five-point action achievement score mechanism.

Score	Similarity	Action
4	0.8~1.0	Completion of action independently
3	0.6~0.8	Completion of action slowly but independently
2	0.4~0.6	No guardian or assistive devices required to assist in movement
1	0.2~0.4	Assistance of guardian or assistive device required
0	0.0~0.2	Unable to perform action

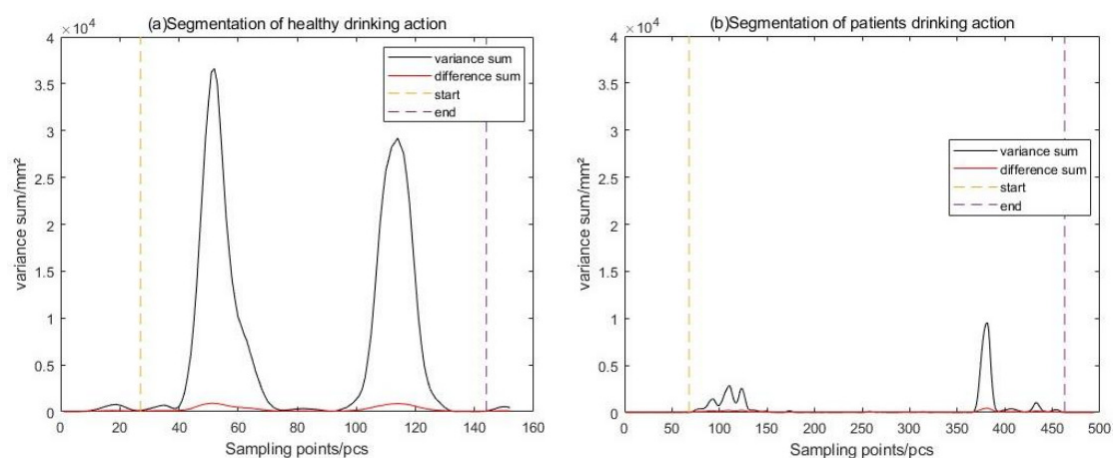
6. Numerical Experiments

A. Preliminary

- (1) Experimental environment: In this paper, we used a computer device with CPU model Intel i5-10400 (Intel, Santa Clara, CA, USA), a graphics card RTX 2070, and 16G RAM; we used Azure Kinect sensors (Microsoft, Redmond, WA, USA) to collect motion data, C++ to write the program and Visual Studio 2017 to compile it. Since Azure Kinect is no longer commercially available, it can be replaced with a depth camera.
- (2) Data acquisition: Self-constructed healthy person action datasets, which include four types of daily activity actions, i.e., drinking water, combing hair, touching the opposite shoulder, and touching the back pocket, were used; each action type involved 10 healthy subjects to conduct the experiment, where each person completed the action process once in a natural state, resulting in a total of 40 sets of action data. Among the 10 groups of data for each type of action, 4 groups were randomly selected to generate templates for each type of action, and the remaining 6 groups were used as test groups. In the self-constructed patient action datasets, each type of action included experimental data completed by three patients, each of whom completed the action once in their natural state, and a total of 12 sets of action data were obtained.

B. Action segmentation

By setting the window length $w = 10$ and the sliding step $l = 1$, we calculated the three-axis difference sum and variance sum of the hand joints through the sliding window. We set the V-value curve peak–peak threshold t_v as the average of the sequence data, the E-value curve data threshold $t_E = 10$, the window length offset value $m_s = 1.5w$, and $m_e = 3w$ for the segmentation of the four kinds of actions of healthy people and patients. Figure 10 shows the results for the drinking action as an example.

**Figure 10.** Segmentation of drinking action in healthy people (a) and patients (b).

In the healthy person's action, there are two peaks in the drinking action, where the first peak is shifted forward from the start-point window value and the last peak is shifted

backward from the end-point window value, to obtain the start point and end point of the action, respectively. There are two to three trajectory peaks and some small trajectory peaks in the patients' actions, which are caused by abnormal hand actions and the limb fluttering of the patient, indicating that it is more difficult for the patient to carry out actions with a large spatial span. The results show that the method used in this experiment can better segment the start and end points of the action and has good practicality and universality.

C. Feature extraction

- (1) Angle features: Based on the key-point vectors of the human body, we could obtain angle features of healthy people and patients. Figure 11 shows the results for the drinking action as an example.

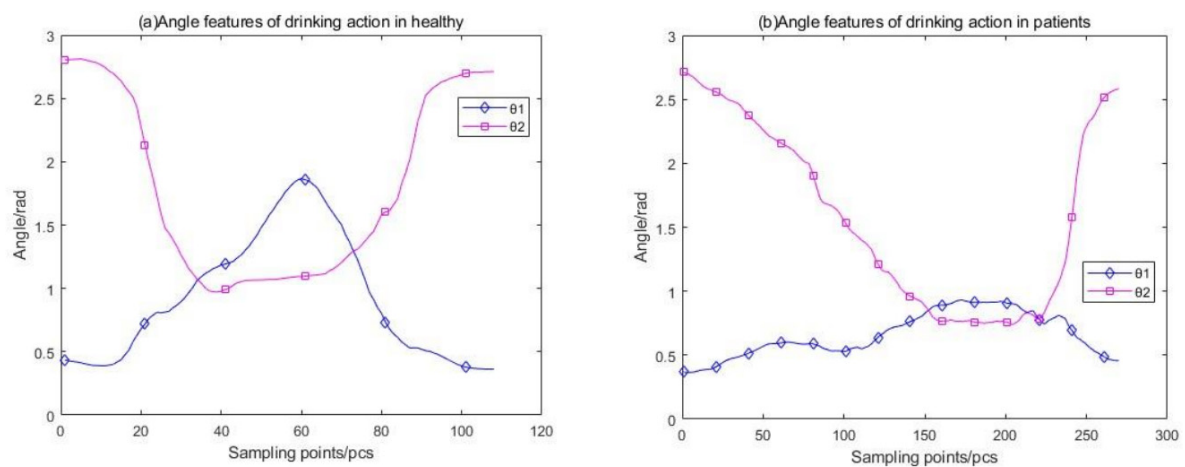


Figure 11. Angle features of drinking water action in healthy people (a) and patients (b).

As can be seen from the figure, the patients performed the drinking action with varying degrees of differences in the angles of the vectors compared with the healthy individuals, which indicates that the patients performed the action with significant dyskinesia, especially in terms of large-scale action and elbow bending.

- (2) Distance features: Based on the distance of the human body reference points in the Y-axis direction, we normalized the distance and obtained the distance features for healthy people and patients. Figure 12 shows the results for the drinking action as an example.

By comparing the distance characteristic plots of healthy people and patients and analyzing the changes in distance during the drinking action, it was found that the patients performed the action with greater fluctuations in the distance characteristic curves, shorter periods of calmness, and obvious hand shaking, which prevented them from completing the action.

D. Action Evaluation

- (1) Similarity calculation: The feature matrix DTW algorithm and the FW-DTW algorithm proposed in this paper were applied to healthy people and patients, respectively, and the similarity of their actions was calculated and averaged to compare the differences between the two methods in healthy individuals and patients; the results are shown in Table 4. After a lot of experimental testing, the weighting ratio of the angle features in the action features was set to $w_1 = 0.2$, and the weighting ratio of the distance features in the action features was set to $w_2 = 0.8$.

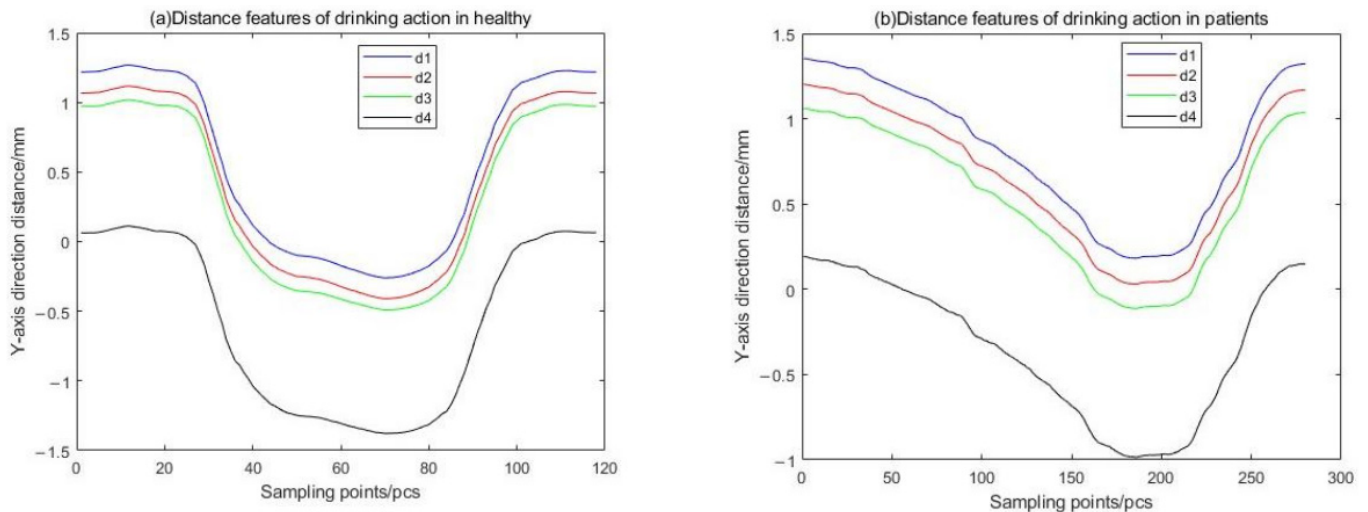


Figure 12. Distance features of drinking water action in healthy people (a) and patients (b).

Table 4. Similarity of actions between healthy people and patients.

Action	Feature Matrix DTW		FW-DTW	
	Healthy People	Patients	Healthy People	Patients
Drinking water	0.9993	0.0980	0.9996	0.1801
Combing hair	0.9994	0.1133	0.9997	0.1747
Touching the opposite shoulder	0.9968	0.1267	0.9985	0.2286
Touching the back pocket	0.9978	0.2734	0.9991	0.5005

By comparing the similarity averages in Table 4, it can be found that the FW-DTW algorithm improved the similarity scores of drinking water and combing hair by 0.0003 in healthy individuals and touching the opposite shoulder and touching the back pocket in patients by 0.0017 and 0.0013, respectively. For patients, the similarity scores of drinking water, combing hair, touching the opposite shoulder, and touching the back pocket increased by 0.0821, 0.0614, 0.1019, and 0.2271, respectively. This is because the feature matrix-based algorithm integrates angle and distance features, while the feature weight-based algorithm reduces the error caused by the angle features and thus improves the similarity, proving that the feature weight-based DTW algorithm makes a significant contribution to improving the similarity of actions. Additionally, by comparing the average similarity values for each type of action between healthy individuals and patients, it was found that the similarity between the template actions and those performed by healthy individuals differed by more than 80% from those performed by patients. This indicates that the proposed evaluation method can effectively assess the action quality of individuals with varying health conditions, identifying those with limb disabilities.

- (2) Achievement score: In rehabilitation action evaluation, for the four actions of drinking water, combing hair, touching the opposite shoulder, and touching the back pocket, we used the FW-DTW algorithm to score the action similarity data of six healthy people and three patients. The action score–similarity comparison table is shown in Table 5.

Mathematical fitting was used, and we found that the mathematical relationship of score–similarity is a linear relation.

A polynomial was fitted to the score–similarity, and the results are shown in Figure 13, where the confidence level of the fitted parameters reaches 99%, which can meet the needs of practical applications.

Table 5. Action score–similarity control table.

Serial Number	Drinking Water		Combing Hair		Touching the Opposite Shoulder		Touching the Back Pocket	
	Similarity Score		Similarity Score		Similarity Score		Similarity Score	
1	0.9999	4	0.9998	4	0.9995	4	0.9995	4
2	0.9997	4	0.9998	4	0.9992	4	0.9992	4
3	0.9997	4	0.9997	4	0.9989	4	0.9992	4
4	0.9996	4	0.9997	4	0.9984	4	0.9990	4
5	0.9995	4	0.9997	4	0.9983	4	0.9990	4
6	0.9992	4	0.9996	4	0.9967	4	0.9986	4
7	0.1872	0	0.1981	0	0.2925	1	0.5466	2
8	0.1828	0	0.1874	0	0.2213	1	0.4984	2
9	0.1704	0	0.1387	0	0.1721	0	0.4564	2

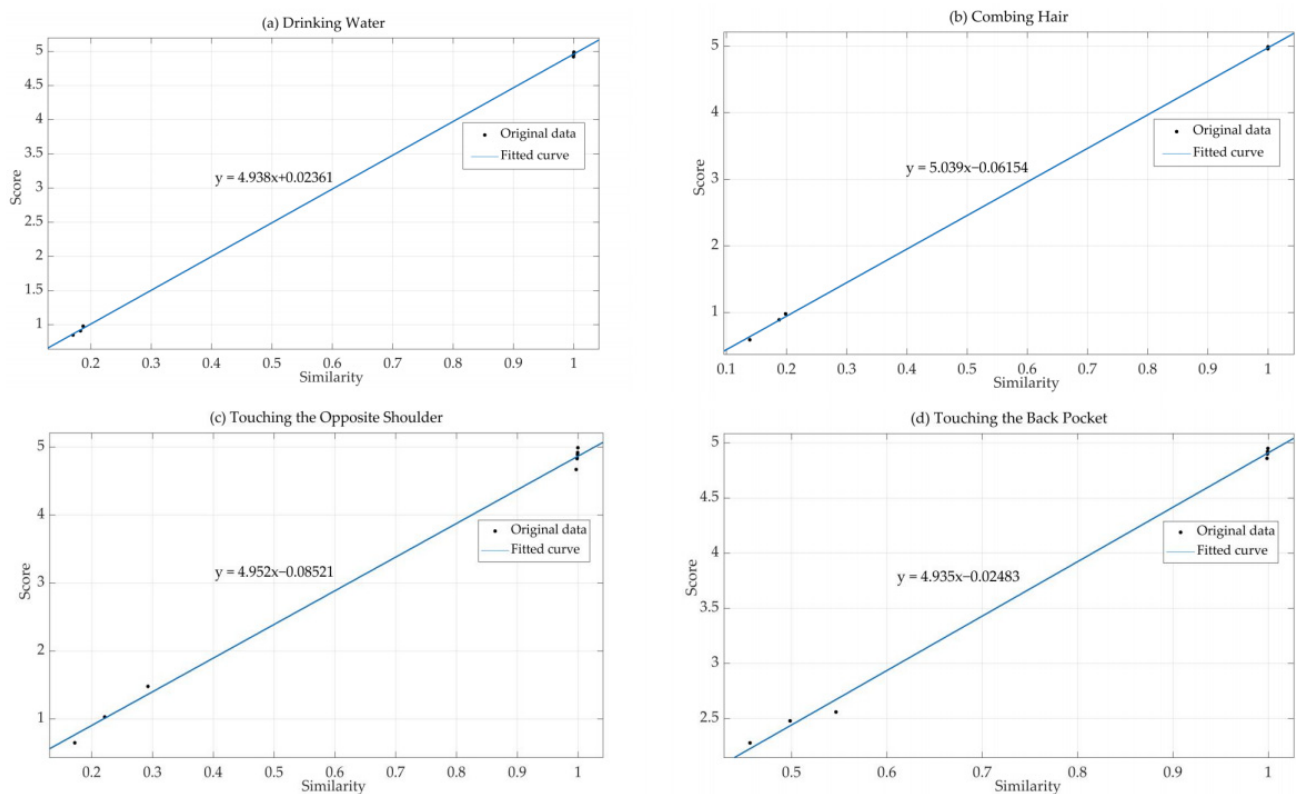


Figure 13. Score–similarity fitting results ((a–d) are the four actions of drinking water, combing hair, touching the opposite shoulder, and touching the back pocket, respectively).

7. Conclusions

The intelligent rehabilitation action evaluation system (IRAES) framework is proposed in this article; it improves the accuracy of action similarity assessment by improving the distance calculation method in the DTW algorithm and provides a novel approach for the development of intelligent rehabilitation systems. Compared with traditional rehabilitation action evaluation methods, DTW-based evaluation using feature matrix, and machine learning-based evaluation, the IRAES makes the following three main contributions:

(1) Firstly, this paper proposes an intelligent rehabilitation action evaluation system framework (IRAES) based on the Kinect sensor, addressing the inefficiency, cumbersome, and subjective bias of traditional assessment methods. The system uses Kinect to capture the three-dimensional coordinate data of key points of the human skeleton and displays the patient's movement assessment results in real time, which helps doctors adjust the rehabilitation plan in time. Compared with manual observation, this system provides a more objective, accurate, and real-time assessment of action quality. In particular, the DTW algorithm based on feature weights can adaptively adjust the alignment of action

sequences, effectively balance the weights of angle and distance features, reduce errors, and improve the accuracy and robustness of action evaluation.

(2) Secondly, in view of the limitations of existing studies that primarily focus on data from healthy individuals, this paper constructed a comprehensive dataset containing action data of healthy people and patients for the IRAES. Compared with studies that only involve simple actions such as standing, sitting, and walking or only focus on the data of healthy people, this paper better aligns with the clinical application requirements in terms of action types and population coverage. By analyzing the similarity differences in four actions between healthy individuals and patients, limb disability can be tested, providing references for personalized rehabilitation plans. In addition, this paper proposes a DBA-based action template creation algorithm, which generates action templates from multiple action samples, significantly reducing computational load and time, thereby enhancing overall efficiency. These innovative designs in dataset construction and template generation algorithms support the real-time applicability of the intelligent rehabilitation action evaluation system.

(3) Finally, in order to solve problems such as human body differences, environmental interference, motion changes, and sensor errors, this article takes innovative measures in data processing, action segmentation, and feature extraction. We use segmented cubic spline interpolation and moving average to eliminate environmental interference and propose an action segmentation strategy based on action feature analysis to manage random pauses and repeated actions effectively. By constructing human body key-point vectors and extracting angle and distance features, the impact of individual differences and sensing errors is reduced. In order to achieve intelligent and adaptive evaluation, we designed an intelligent index system for rehabilitation movement evaluation, established a five-point score mechanism based on movement similarity, and achieved quantification and adaptability of movement quality assessment through polynomial fitting. Faced with the challenges of algorithm complexity and modeling difficulties, this article selected the DTW algorithm with low complexity and strong adaptability to achieve efficient and accurate rehabilitation action assessment, which has important reference value for the development of intelligent rehabilitation systems.

The experimental results demonstrate the following findings: Firstly, compared with action evaluation based on the feature matrix DTW algorithm, action evaluation based on the FW-DTW algorithm significantly improves the similarity of four types of actions, thereby enhancing the overall performance of action evaluation. Second, the difference in similarity between patients' action data and the template action data is large enough to evaluate patients' limb disability status. Finally, the confidence level of the established score mechanism reached 99%, confirming its applicability for rehabilitation action evaluation.

Author Contributions: Conceptualization, M.Y. and X.L.; methodology, M.Y.; software, N.G.; validation, Z.L., M.Y. and N.G.; formal analysis, X.L.; investigation, Z.L.; resources, Z.L.; data curation, X.L.; writing—original draft preparation, M.Y.; writing—review and editing, M.Y.; visualization, Z.L.; supervision, N.G.; project administration, X.L.; funding acquisition, X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially supported by a key research and development program project of Hubei province (No. 2020BCB054).

Institutional Review Board Statement: This study was approved by the Ethics Committee of Wuhan Wuchang Hospital (No. 2022001, 2 March 2022) and complied with the Declaration of Helsinki.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author due to privacy.

Acknowledgments: We thank all the participants for their time, effort and valuable feedback.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Hu, B. Research on Recognition Algorithms of Rehabilitation Posture and Action Based on Multi-Rules Learning. Master's Thesis, Yanshan University, Qinhuangdao, China, 2019.
2. Yan, H.; Chen, G.; Cui, L.Y.; Zhang, L.Y.; Hu, B.C. Online human rehabilitation action recognition based on monocular vision. *Comput. Appl. Softw.* **2021**, *38*, 171–178.
3. Chen, X. Research on Action Recognition and Its Application in Upper Limb Rehabilitation Training. Master's Thesis, Qufu Normal University, Qufu, China, 2021.
4. Ma, Y.T.; Wang, S.; Liu, Y.F. Research on Human Action Recognition Method by Fusing Multimodal Data. *Comput. Eng.* **2022**, *48*, 180–188.
5. Li, R.M. Research on Fine Classification and Evaluation of Human Action Based on Visual Data. Ph.D. Thesis, Xi'an Institute of Optics & Precision Mechanics, Chinese Academy of Sciences, Xi'an, China, 2020.
6. Ren, Z.Y. Research on Recognition of Upper Limb Rehabilitation Exercises After Stroke Based on Convolutional Neural Network. Master's Thesis, University of Electronic Science and Technology of China, Chengdu, China, 2021.
7. Mao, K.; Wu, X.M.; Wang, C.S. Evaluation of upper limb rehabilitation movement based on improved DTW. *Sci. Technol. Vis.* **2021**, *16*, 119–121. [[CrossRef](#)]
8. Liu, Y.T. Human Motion Capture and Action Recognition Based on Wearable Sensors. Master's Thesis, Harbin Institute of Technology, Harbin, China, 2020.
9. Elkholy, A.; Hussein, M.E.; Gomaa, W.; Damen, D.; Saba, E. Efficient and Robust Skeleton-Based Quality Assessment and Abnormality Detection in Human Action Performance. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 280–291. [[CrossRef](#)]
10. Ma, Y.X. System of Rehabilitation Exercise Quality Evaluation Based on Kinect. Master's Thesis, University of Chinese Academy of Sciences, Shenzhen, China, 2021.
11. Bruce, X.B.; Liu, Y.; Chan, K.C.; Yang, Q.; Wang, X. Skeleton-based human action evaluation using graph convolutional network for monitoring Alzheimer's progression. *Pattern Recognit.* **2021**, *119*, 108095.
12. Wang, G.J.; Cheng, M.; Wang, X.S.; Fan, Y.; Chen, X.; Yao, L.L.; Zhang, H.Y.; Ma, Z.C. Design and Evaluation of an Exergame System of Knee with the Azure Kinect. In Proceedings of the International Conference of Pioneering Computer Scientists, Engineers and Educators (ICPCSEE, originally ICYCSEE), Taiyuan, China, 17–20 September 2021; pp. 97–99.
13. Ding, K.L. Quantitation and Screening of Bradykinesia in Parkinson's Disease using Motion Capture. Master's Thesis, Nankai University, Tianjing, China, 2021.
14. Wang, Y.P. Algorithm and Implementation of Real-Time Human Action Evaluation in Tennis Training Robot. Master's Thesis, Beijing University of Posts and Telecommunications, Beijing, China, 2018.
15. Rao, C.; Gritai, A.; Shah, M.; Syeda-Mahmood, T. View-invariant Alignment and Matching of Video Sequences. In Proceedings of the Ninth IEEE International Conference on Computer Vision, Nice, France, 13–16 October 2003; pp. 939–945.
16. Zhou, F.; Torre FD, L. Canonical time warping for alignment of human behavior. In *Advances in Neural Information Processing Systems 22: 23rd Annual Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 7–10 December 2009*; Curran Associates: Red Hook, NY, USA, 2009; pp. 1–9.
17. Gong, D.; Medioni, G. Dynamic manifold warping for view invariant action recognition. In Proceedings of the IEEE International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 571–578.
18. Fan, Y.Q. Human Pose Estimation Based on Action Recognition and Action Evaluation. Master's Thesis, Beijing University of Posts and Telecommunications, Beijing, China, 2021.
19. Wang, R.Z.; Medioni, G.; Winstein, C.J.; Blanco, C. Home Monitoring Musculo-Skeletal Disorders with a Single 3D Sensor. In Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition Workshops, Portland, OR, USA, 23–28 June 2013; pp. 521–528.
20. Chen, Y.S. Research on Action Analysis Based on Human Skeleton Information. Master's Thesis, Hangzhou Dianzi University, Hangzhou, China, 2021.
21. Liu, X.; Yan, M.D.; Li, Y.Z.; Li, X. Evaluation System of Push-up Action Based on Kinect. In Proceedings of the 2022 14th International Conference on Machine Learning and Computing, Guangzhou, China, 18–21 February 2022; pp. 308–312.
22. Zhang, X.; Li, Z.; Li, S.; Xing, Z.; Wang, J. Effect of rehabilitation training guided by Pro-kin balance system on proprioception and balance function of the affected knee after anterior cruciate ligament reconstruction. *Chin. J. Tissue Eng. Res.* **2024**, *28*, 1259.
23. Choi, H.S.; Cho, S.H. Effects of Multimodal Rehabilitation on the Activities of Daily Living, Quality of Life, and Burden of Care for Patients with Parkinson's Disease: A Randomized Control Study. *Healthcare* **2022**, *10*, 1888. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.