ISSN 2616-5775 Vol. 6, Issue 12: 106-117, DOI: 10.25236/AJCIS.2023.061212

Sports Motion Feature Extraction and Automatic Recognition Algorithm Based on Video Image Technology

Peiqi He*

Qingdao Institute of Software, College of Computer Science and Technology, China University of Petroleum (East China), Qingdao, Shandong, 266000, China
2004030218@s.upc.edu.cn
**Corresponding author*

Abstract: *With the rapid development of computer technology, body motion recognition technology based on computer vision came into being. In the research of sports field, the coach can help the athletes to carry out auxiliary training and find out the advantages and disadvantages of the athletes through the analysis of the athletes' sports state, so as to guide athletes pertinently and improve their mastery of sports actions. However, the traditional motion recognition methods have low recognition accuracy and low recognition efficiency, and are difficult to be effectively used in sports training. Based on this, this paper studied the feature extraction and automatic recognition of sports actions based on video image technology, proposed the feature extraction method of sports actions based on bone key points and the K-Nearest-Neighbor (KNN) sports action recognition method combined with the improved Dynamic Time Warping (DTW) algorithm, and carried out experimental research on this method. The research showed that compared with the KNN-based sports action recognition method, the improved DTW-based KNN sports action recognition method had a 4.2% higher precision and a 2.4% higher recall. KNN sports movement recognition method based on improved DTW had good recognition effect.*

Keywords: *Sports Action, Feature Extraction, Action Recognition, Video Image Technology*

1. Introduction

Human motion recognition using computer vision technology has become a research hotspot. In video image retrieval, video classification and retrieval using computer vision technology can not only ensure classification accuracy, but also replace manual video annotation and classification, thus greatly reducing the time for users to find their own video. In the aspect of video surveillance, it can observe the behavior of people under surveillance through the recognition of human motion, so as to give early warning of abnormal behavior. In sports, people can analyze the athletes' movement state in the video to improve their training effect. Aiming at the problem that the recognition accuracy of traditional motion recognition methods is not as expected, this paper studies the feature extraction and automatic recognition of sports motion.

Many scholars have studied human motion feature extraction and motion recognition. Zhang Xiaojun analyzed the application of convolution neural network in motion recognition, and studied human motion recognition and intelligent wearable devices based on deep learning in motion [1]. Cust Emily E. systematically reviewed the literature on machine and depth learning for specific motion recognition using inertial measurement units and computer vision data input [2]. Yang Yun has manufactured a pressure-sensitive insole and intelligent ski stick based on customizable and flexible friction nano-generator to monitor athletes' sports [3]. Rahmad Nur Azmina believed that the crucial performance analysis in sports practice can improve the performance of athletes in the competition. Based on this, he has studied the video-based motion recognition in sports [4]. Dong Zhe proposed an automatic identification method for javelin thrower's throw angle, aiming at the problem that the regular statistical performance of body feature data before recognizing the throw angle is poor, which leads to the deviation of javelin flight trajectory judgment results [5]. Gurbuz Sevgi Zubeyde put deep learning in the context of data-driven motion classification methods, and compared its performance with other methods that use handmade features [6]. Jaouedi Neziha believed that human motion recognition is an important challenge in various applications including human-computer interaction and intelligent video surveillance. He also proposed a new method of human motion recognition based on hybrid deep learning model [7].

Wang Jucui proposed a B-spline active contour model based on dynamic programming method to solve the problems of traditional motion extraction methods, and proposed a method to use the model for face image processing, extracting computer tomography image data and building a 3D model [8]. Wang Peng studied deep learning as a data-driven technology for continuous human motion analysis and future human-computer cooperation demand prediction [9]. Hendry Danica provided a new method to measure the amount of dancer training by quantifying specific sports tasks. Such a system can be used to further understand the relationship between dancer's pain and training volume, and can also be used for athlete monitoring system [10]. Rangasamy Keerthana reviewed the traditional manual methods and deep learning methods in motion video analysis based on human activity recognition, and summarized the latest research on video-based human activity recognition in motion analysis [11]. Kong Longteng proposed a zoom and occlusion robust tracker for athlete tracking and motion recognition [12]. Ullah Mohib discussed the effectiveness of sparse representation obtained by learning a set of overcomplete bases in the context of video motion recognition [13]. In order to solve the problem of insufficient training data for human motion recognition, Li Xinyu proposed an example-based migration learning method, which has limited radar micro-Doppler signatures, reducing the burden of collecting and labeling a large number of radar samples [14]. The above scholars have studied human motion feature extraction and motion recognition, and put forward valuable suggestions.

In order to enable coaches to better understand the movements of athletes, targeted training plans are formulated to improve the competitive ability of athletes. This paper studies the feature extraction method and recognition method of sports action through video image technology, and proposes the feature extraction method of sports action based on the key points of human bone. The KNN sports action recognition method combined with DTW is proposed, and a comparative study between the KNN sports action recognition method combined with DTW and the traditional KNN sports action recognition method is carried out. Compared with other related studies, this paper uses a self-built video dataset, including hurdles, push-ups, squats and lunges and leg presses, which are common sports movements.

2. Problems and Method Categories of Human Action Behavior Recognition

2.1 Problems in Human Action Behavior Recognition

With the rapid development of computer and artificial intelligence technology, the motion recognition technology in images has been paid more and more attention by scholars. However, so far, there are still many technical problems in human motion recognition in video, mainly including the following problems, as shown in Figure 1.

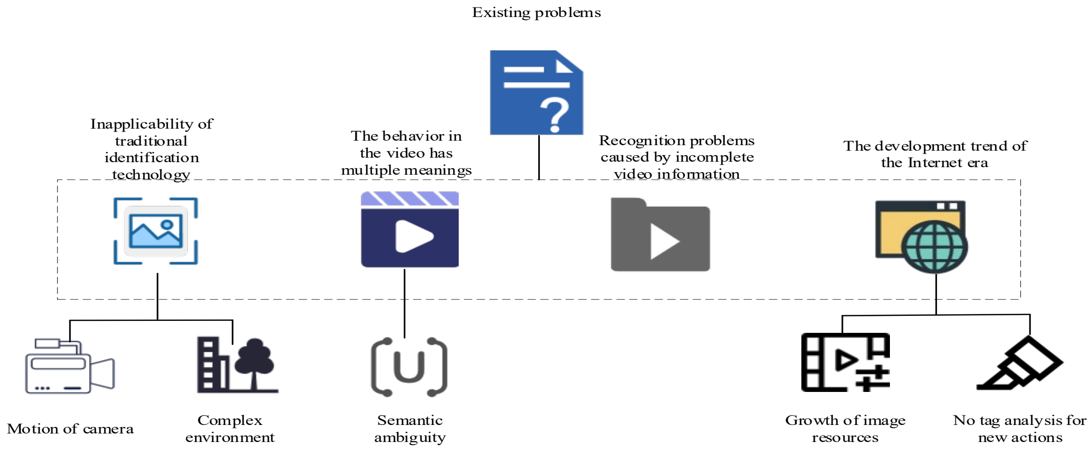


Figure 1: Problems in human action behavior recognition

Inapplicability of traditional recognition technology: human motion recognition technology is limited to fixed view angle, static background, etc., while the established motion database can only accommodate simple actions. With the development of various imaging devices, people began to establish multiple dynamic databases. Compared with the traditional camera, simple shooting object and fixed shooting mode, the movement in actual operation would be affected by camera movement, complex environment and other factors, and it is difficult to analyze and identify.

The behavior in the video has multiple meanings. In the video, the behavior is not a simple performance, but a complex behavior. For example, "hurdle" is actually a combination of basic "jump" and "run", which is difficult to define and distinguish. In fact, in real life, there are countless similar actions and complex actions. Because each action is difficult to classify, and the type of each action is very small, the number of specific actions is limited, which leads to many semantic ambiguity, making it difficult to identify and analyze. At present, many scholars have studied this problem and made a lot of improvements to it in order to improve the image recognition effect. However, there are many difficulties in visual recognition, especially in motion recognition.

Incomplete video information can also cause recognition problems, and it is also beneficial to recognize when video information is incomplete, as shown in Figure 2:

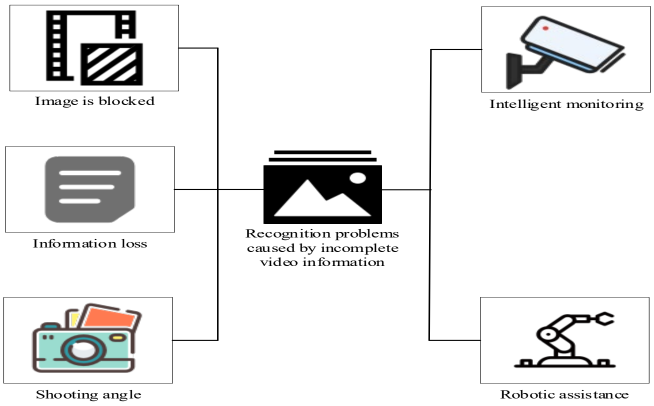


Figure 2: Recognition problems caused by incomplete video information

Most of the traditional video image recognition technologies are comprehensive analysis and recognition of the image. When the image is blocked or the information is lost in the process of transmission, the motion of the image is incomplete, which affects the correct recognition of the video. In incomplete images, if good identification effect can be achieved, this technology can be used to identify in densely populated areas, then can be used to early warning some behaviors that endanger people and the public, and can be applied to intelligent monitoring and robot assistance and other fields. Therefore, using incomplete images for behavior identification is also a meaningful research topic. In addition, because of the difference of shooting angle, it would also affect the recognition of motion. In general, motion recognition must make the motion image in the training library and the recognized human motion image at the same angle. Some motion recognition algorithms cannot effectively solve this problem. However, in reality, especially when the number of cameras is large, the shooting angle would vary greatly. In addition, the unfamiliar actions people found in many exercises are also a problem that cannot be ignored.

The development trend of the Internet era: with the development of the Internet, the rise of various media, and the full range of online live broadcasting, various video resources have shown a geometric progression of growth, while at the same time, people have also begun to appear a variety of entertainment activities. To conduct detailed analysis and identification of this behavior requires not only a lot of manpower and material resources, but also is difficult to achieve. At the same time, because the new movement is not marked and analyzed, and the new movement cannot be accurately identified and analyzed, the motion recognition technology in this area needs further research. If people want to obtain massive data, people must spend time and energy to find some special data from the massive data. The existing data retrieval technology has a good application in text and image retrieval, but there is not much research in specific video search. This is because the processing of video data is very complex and difficult to match, but it has many practical applications.

2.2 Categories of Human Action Behavior Recognition Methods

Extracting the motion characteristics of the objects in the video is a very important work. After extracting the video image characteristics, people would enter the next step. In video human motion recognition, the most representative methods are template matching and probabilistic statistical model.

ISSN 2616-5775 Vol. 6, Issue 12: 106-117, DOI: 10.25236/AJCIS.2023.061212

In addition, various technologies based on semantic description and syntax analysis are also widely used in the video field. With the continuous development of computer software and hardware technology, the deep learning method is also increasingly applied to video. As a means of behavior recognition, the human action behavior recognition method is shown in Figure 3.

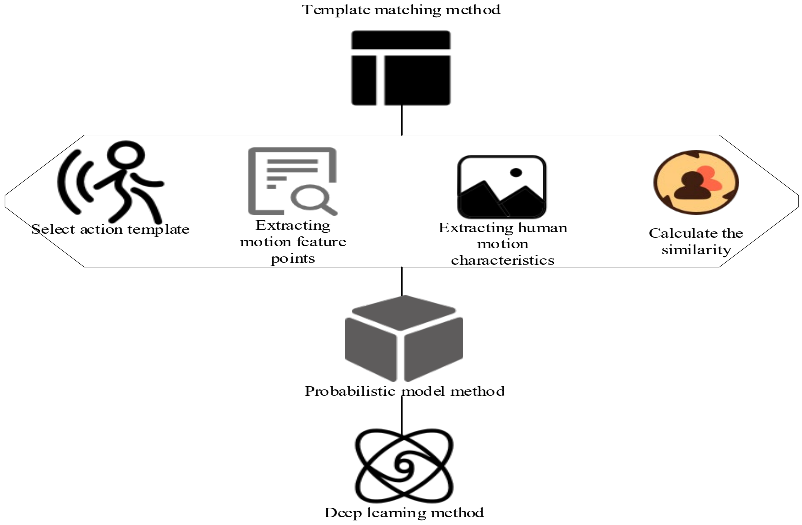


Figure 3: Classification of human action behavior recognition methods

Template matching method: the template matching technology uses the pre-stored action template to determine the type of action according to the similarity comparison. In this process, people can choose different goals. Template matching can be divided into two categories: one is regular template matching, and the other is dynamic timing adjustment.

First of all, the motion feature points are extracted from the video by selecting the action template. According to the template, the extracted features can be low-level or high-level. Then, the same algorithm is used to extract the human motion features from the recognized image, and then the similarity between the template and the detected object is calculated by calculating the distance between the template movement and the recognized object. When the similarity reaches a certain level, the identified behaviors would be classified. In pattern recognition, the most common method is single target matching. This method is simple, fast and efficient. However, in this algorithm, how to choose the time interval of action video is a big problem. If the running time of the model is very different between the template and the case to be tested, the number of templates required would be reduced and the required effect would not be achieved. To obtain good recognition, it is necessary to ensure that there are enough templates.

Probability model method: probability model method is the most widely used method in video image recognition. At present, there are two main statistical analysis methods: generating model method and discriminant method. Using statistical theory and generating model, people can analyze the motion characteristics and motion state characteristics of video images. The generative model has great flexibility.

At present, several commonly used models include Dynamic Bayesian Network (DBN), Hidden Markov Model (HMM), etc. HMM not only has the advantages of time difference correction, but also has many advantages such as recognition and learning. It has good application value in time-varying signal processing. HMM needs to focus all features on a node, and HMM needs a large number of learning samples. The HMM model cannot examine the characteristics of context, so its independence assumption would also restrict its selection. In fact, the current behavior state and long-term observation are complementary. In addition, this method can also meet the requirements of observation and state intersection. DBN can appear at any time and can be used to describe various states, observation states and the relationship between various states.

In video, people also generally use discrimination. When constructing the discriminant model, only the posterior density is considered, not the density based on class conditions. Therefore, it is not difficult. In addition, when estimating model parameters, people only need to approximate the regression function, such as Support Vector Machine (SVM).

Deep learning method: deep learning is similar to human cognitive characteristics, and can extract

ISSN 2616-5775 Vol. 6, Issue 12: 106-117, DOI: 10.25236/AJCIS.2023.061212

some special features from images. After sufficient sampling training, some images with special meanings can be obtained, and the recognition of behavioral targets can also get good results [15].

3. Method of Sports Movement Feature Extraction and Automatic Recognition

3.1 Physical Movement Feature Extraction Based on Bone Key Points

The European distance conversion technology is used to process the moving binary image, eliminate the noise in the image, and obtain an ideal human skeleton image per unit pixel width. Then the actual position of human joint points can be accurately determined by analyzing and searching the status of each joint point pixel.

In essence, distance conversion is to convert a binary image once in a two-dimensional space, while a binary image can be considered to contain only object pixels and background pixels. The gray level of the target pixel is 1, while the gray level of the background pixel is 0. Using distance conversion technology, a gray image can be obtained. In a distance image, the distance between each pixel and its nearest background pixel is the gray level of the pixel.

The expression formulas of Euclidean distance transformation are:

$$Y = \{(a,b) | I_{ab} = 0\} \tag{1}$$

$$G = \{(p,q) | I_{pq} = 1\} \tag{2}$$

Among them, Y is the background pixel set.

The shortest European distance is:

$$e_{pq} = \min\{E[(p,q),(a,b)],(a,b) \in Y\} \tag{3}$$

Among them, $E[(p,q),(a,b)] = \sqrt{(p-a)^2 + (q-b)^2}$.

The Euclidean distance transform image can be obtained by calculation.

The background difference method is used to judge pixel points:

$$(a, b) = \begin{cases} Former\ attractions, \left|G_s(a, b) - Y(a, b)_s\right| \geq S \\ Background\ point, \left|G_s(a, b) - Y(a, b)_s\right| < S \end{cases} \tag{4}$$

Gaussian background modeling technology is used to model the background of motion sequence. Through morphological processing of the obtained image, and through expansion, corrosion, opening and closing operations, noise is eliminated, so as to obtain a smoother and more realistic human contour.

On this basis, an image refinement algorithm based on Euclidean distance transform is used to segment the original binary image into a skeleton of unit pixel width. On this basis, a convolution kernel of matrix is constructed and solved. The calculation formula of Euclidean distance transformation value is:

$$S(a,b) = \min_{(a',b') \in Y} \sqrt{(a-a')^2 + (b-b')^2} \tag{5}$$

The Euclidean distance transform image containing the distance is output, and each pixel is scanned horizontally with the linear filtering method. Each row on the vertical axis b is marked, and the pixel at the maximum point is converted with the Euclidean distance:

$$N(b) = \max_{b'=b} S(a',b') \tag{6}$$

The curve representing the structure of human bones can be obtained from Formula (6). This method adopts quadratic polynomial integral fitting and rough mesh method for noise elimination,

ISSN 2616-5775 Vol. 6, Issue 12: 106-117, DOI: 10.25236/AJCIS.2023.061212

which effectively reduces the redundant points on the skeleton and maintains the basic features of the skeleton, thus obtaining the human skeleton image per pixel wide. After obtaining the binary image, it is converted into a two-dimensional array $I[N,M]=[I_{ab}]$.

The joint points of head, hand, foot, shoulder, knee and hip are extracted. *I* is the pixel detected in the bone image; g(I) is the gray level; $M_p(p=1,2,...,8)$ is the adjacent area of the pixel. If the number of non-zero pixels in the eight adjacent regions of *I* is M(I), then there is:

$$M(I)=g(M_1)+g(M_2)+g(M_3)+...+g(M_8) \tag{7}$$

The recognition and marking condition of head, hand and foot joints is as follows:

$$M(I)=1 \tag{8}$$

The recognition and marking condition of hip joint points is:

$$M(I)=3 \tag{9}$$

Recognition and identification conditions of knee joint points: in the case of no serious torsion of the human trunk, transverse scanning shall be carried out according to the proportion between knee and height in human anatomy to find a pixel point with one gray scale. After finding that the body of the human body has undergone severe distortion, the curvature of the bump points on the human bone curve is measured to determine the knee. At any point *I* on the lower limb bone curve of the human body, it is represented by I_{+n}, I_{-n}. From this point, *n* steps are moved along the horizontal and vertical directions of the lower limb bones. The normalized curvature at point *I* is:

$$K_i(n)=\begin{cases} U_i(n), \text{If I point is convex} \\ -U_i(n), \text{If I point is concave} \end{cases} \tag{10}$$

Among them, $U_i(n)=\dfrac{|II_{+n}|^2+|II_{-n}|^2-|I_{+n}I_{-n}|^2}{4|II_{+n}||II_{-n}|}+\dfrac{1}{2}$.

With the hip joint as the center point, the transverse and longitudinal distances of the other six points except the head are calculated from the transverse and longitudinal directions, and the distance feature vector is determined:

$$e_m=\left[e_{lh_a}, e_{lh_b}, e_{rh_a}, e_{rh_b}, e_{lk_a}, e_{lk_b}, e_{rk_a}, e_{rk_b}, ...\right] \tag{11}$$

The European distances from neck to hip joint are standardized as a standard:

$$e_{mn}=\sqrt{(M_a-N_a)^2+(M_b-N_b)^2} \tag{12}$$

$$e'_m=\frac{e_m}{e_{mn}} \tag{13}$$

$$E=\left[e'_1, e'_2, e'_3, ..., e'_m\right] \tag{14}$$

The motion characteristics are described by the speed of the checkpoint, and the Euclidean distance between two frames is calculated by the plane coordinates of the two checkpoints. The difference between frames is taken as the time of change, and the ratio of distance to time is obtained, which is the joint speed characteristics. The two-dimensional coordinates of the two frames on the left are $lh_q(lh_a, lh_b), lh_{q-1}(lh_{a-1}, lh_{b-1})$, while the q-th frame rate on the left is:

$$w_{lh_q}=\frac{\left\|lh_q-lh_{q-1}\right\|_2}{s} \tag{15}$$

ISSN 2616-5775 Vol. 6, Issue 12: 106-117, DOI: 10.25236/AJCIS.2023.061212

The principal component analysis method is used to combine two complementary single features to fully and completely describe the action, and the principal component analysis method is used to reduce the feature dimension:

A set of data samples is set as A:

$$A = \begin{bmatrix} a_{11}\ a_{12} \dots\ a_{1p} \\ \dots \\ a_{q1}\ a_{q2} \dots\ a_{qp} \end{bmatrix} \tag{16}$$

The calculation formula of correlation coefficient and matrix is:

$$U = \left(u_{ij}\right)_{p*p}, u_{ij} = \frac{\sum_{k=1}^{m} \overline{b}_{ki} \overline{b}_{kj}}{(m-1)} \tag{17}$$

Among them, u_{ij} represents the correlation coefficient of the j-th index of the i-th sample.

For the calculation of the eigenvalues and eigenvectors of *U*, the eigenvalues of *U* are $\kappa_1, \kappa_2, \dots, \kappa_p$, and the eigenvector after orthogonalization is:

$$x_1 = \begin{bmatrix} x_{11} \\ \dots \\ x_{i1} \end{bmatrix}, \dots, x_i = \begin{bmatrix} x_{1i} \\ \dots \\ x_{ii} \end{bmatrix} \tag{18}$$

The exponential variables represented by each column in *A* are integrated into vectors:

$$Wxu = \left[Wxu_1, Wxu_2, \dots, Wxu_i\right]^S \tag{19}$$

The p-th principal component in *A* is:

$$G_p = \left(x_p\right)^S Wxu = x_{1p} * Wxu_1 + \dots + x_{ip} * Wxu_i \tag{20}$$

The cumulative contribution rate is calculated. When it reaches 85%, it is regarded as the main component. The calculation formula of the cumulative contribution rate is:

$$x_i = \frac{\sum_{c=1}^{i} \kappa_c}{\sum_{c=1}^{m} \kappa_c} \tag{21}$$

3.2 Sports Movement Recognition Based on KNN

KNN algorithm flow is data acquisition. A tested dataset $S = \{(a_1, b_1), (a_2, b_2), \dots, (a_M, b_M)\}$ is entered. a_M belongs to a feature vector of sports action and b_M belongs to sports action type. The distance between a series of sports action sequences to be measured and all known sports action sequences is calculated. The most recent *k* instances are selected. The occurrence times of the first *k* points are determined. The category with the largest number of occurrences in the previous *k* points is used as the prediction classification of the current point.

KNN usually uses Euclidean distance as the measurement method, and its calculation method is:

$$e(X, Y) = \sqrt{\sum_{c=1}^{m} \left(a_p - b_p\right)^2} \tag{22}$$

However, because Euclidean distance cannot solve the problem of unequal length of time series, DTW is introduced to solve this problem.

Two sports action sequences X and Y are set:

$$X = \left(X_1, X_2, \ldots, X_p, \ldots, X_m\right) \tag{23}$$

$$Y = \left(Y_1, Y_2, \ldots, Y_q, \ldots, Y_n\right) \tag{24}$$

Among them, *X* is the sports action test sequence; *m* and *n* are the number of frames of the sports action sequence; X_p and Y_q are the feature vectors of frame *p* and frame *q*.

If *m* equals *n*, the cumulative distance of two sports action sequences can be directly calculated. If *m* and *n* are not equal, the sequence of the two motions would be lengthened or shortened to form the m*n matrix. (p, q) is the distance from X_p to Y_q, and the Euclidean distance is used to measure the distance of similarity:

$$e\left(X_p, Y_q\right) = \sqrt{\sum_{pv=1}^{N}\left(X_{pv} - Y_{qv}\right)^2} \tag{25}$$

Among them, X_{pv}, Y_{qv} are the characteristic value of sports action sequence *X* and *Y*.

The straight line through which the points aligned between the two sports action sequences *X* and *Y* pass is called the regular path, that is, the best route from (1,1) to (m, n). *U* represents the planned route, and the c-th element of *U* is defined as $U_c = \left(p, q\right)_c$:

$$U = \left\{U_1, U_2, \ldots, U_c\right\} \quad \max\left(m, n\right) \le c \le m + n - 1 \tag{26}$$

In dynamic programming, boundary constraints, continuity constraints and monotonicity constraints must be satisfied. The boundary limit is a dynamic regular path from the lower left corner to the upper right. Continuity restriction: the movement points in the sports action sequence must appear on the regular path, and the two points are continuous. Single tonality limit: if $U_{c-1} = \left(a, b\right)$ is set, the dynamic planning point is $U_c = \left(a', b'\right)$, and $\left(a' - a\right)\rangle\rangle 0, \left(b' - b\right)\rangle\rangle 0$ must be met to ensure the constraint of the planned route on the time axis.

According to the restriction conditions, the point (p, q) has (p+1, q), (p, q+1), and (p+1, q+1) connected with it. The Euclidean distance of $\varepsilon\left(p, q\right)$ is X_p and Y_q plus the distance from the nearest element to the point to this point is called the cumulative distance, which is inversely proportional to the similarity. The cumulative distance is expressed by the following formula:

$$\varepsilon\left(p, q\right) = e\left(X_p, Y_q\right) + \min\left\{\varepsilon\left(p-1, q-1\right), \varepsilon\left(p, q-1\right), \varepsilon\left(p-1, q\right)\right\} \tag{27}$$

To improve the formula and add the optimization coefficient, there are:

$$\varepsilon\left(p, q\right) = e\left(X_p, Y_q\right) + \min\left\{\sigma\varepsilon\left(p-1, q-1\right), \varsigma\varepsilon\left(p, q-1\right), \varepsilon\left(p-1, q\right)\right\} \tag{28}$$

Among them, the value of σ, ς is greater than the decimal times of $\sqrt{2}$.

4. Evaluation on the Effect of Sports Movement Feature Recognition Method

Table 1: Number of video and image frames

	Number of videos	The number of video key image frames
Hurdling video set	12	174
Push up video set	15	189
Squat video set	18	241
Video set of lunge leg press	21	268

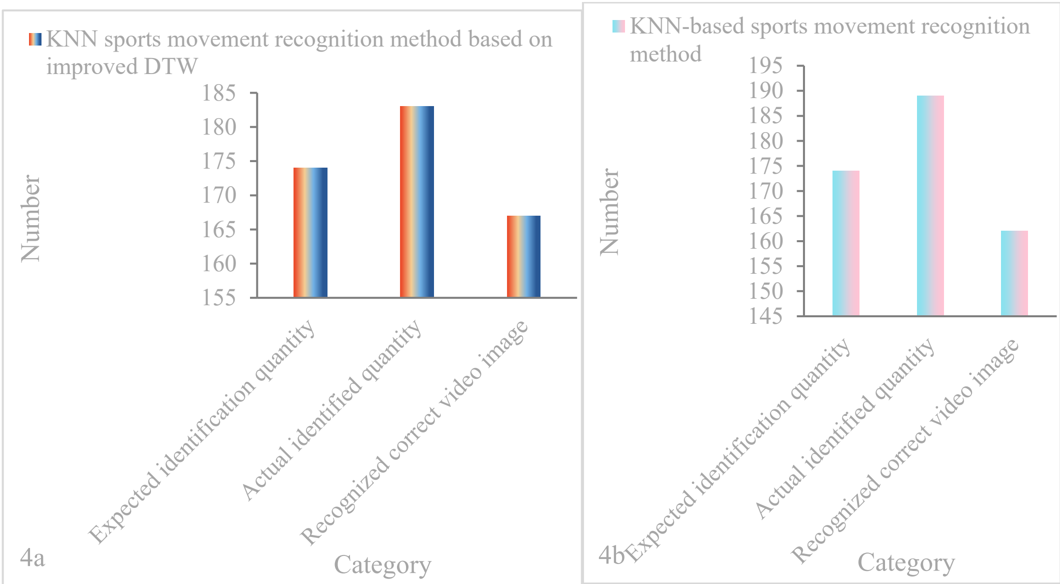
This article recorded several sports action video sets at Y School in Z City, including students' hurdle sports video sets, push-ups sports video sets, squats video sets and lunge leg press video sets. The recognition effect of these four video data sets was tested by using the sports motion feature

ISSN 2616-5775 Vol. 6, Issue 12: 106-117, DOI: 10.25236/AJCIS.2023.061212

extraction based on bone key points and the improved KNN sports motion recognition method of DTW. The video key image frames of each sport were obtained. The number of video and image frames was shown in Table 1.

4.1 Recognition Effect of Hurdle Movement

The same number of video image frames of other sports types were added to the hurdle movement video set. The KNN sports action recognition method combined with improved DTW and the KNN-based sports action recognition method were used to identify the hurdle movement in the hurdle movement video set, as shown in Figure 4.



4a: Hurdle motion recognition under KNN sports motion recognition method combined with improved DTW

4b: Hurdle motion recognition based on KNN sports motion recognition method

Figure 4: The recognition effect of the hurdle action

As shown in Figure 4, Figure 4a shows the hurdle motion recognition results under the KNN sports motion recognition method combined with the improved DTW, and Figure 4b shows the hurdle motion recognition results under the KNN sports motion recognition method. It can be seen from Figure 4a that in the hurdle movement video set, the number of correct video images obtained by combining the improved DTW KNN sports motion recognition method was 167, and the number of recognized motion video images was 183. The precision and recall of this method for hurdle movement recognition were about 91.3% and 96% respectively. It can be seen from Figure 4b that in the hurdle motion video set, the number of correct video images obtained by the KNN-based sports motion recognition method was 162, and the number of recognized motion video images was 189. The precision and recall of this method for hurdle motion recognition were about 85.7% and 93.1%, respectively.

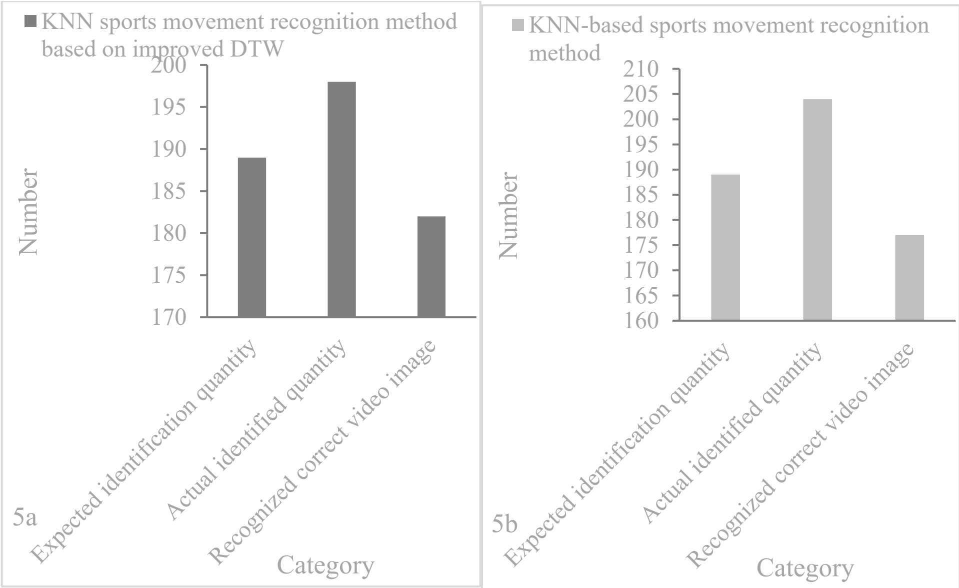
4.2 Recognition Effect of Push-up Sports Movement

The same number of video image frames of other sports types were added to the push-up motion video set. The KNN sports motion recognition method combined with the improved DTW and the KNN-based sports motion recognition method were respectively used to identify the push-up motion in the push-up motion video set, as shown in Figure 5.

As shown in Figure 5, Figure 5a shows the push-up motion recognition results under the KNN sports motion recognition method combined with the improved DTW, and Figure 5b shows the push-up motion recognition results under the KNN sports motion recognition method. As can be seen from Figure 5a, in the push-up motion video set, the number of correct video images obtained by combining the improved DTW KNN sports motion recognition method was 182, and the number of recognized motion video images was 198. The accuracy and recall of the push-up motion recognition by combining the improved DTW KNN sports motion recognition method were about 91.9% and 96.3%,

ISSN 2616-5775 Vol. 6, Issue 12: 106-117, DOI: 10.25236/AJCIS.2023.061212

respectively. It can be seen from Figure 5b that in the push-up motion video set, the number of correct video images obtained by the KNN-based sports motion recognition method was 177, and the number of recognized sports video images was 204. The accuracy and recall of the KNN-based sports motion recognition method for push-up motion recognition were about 86.8% and 93.7%, respectively.



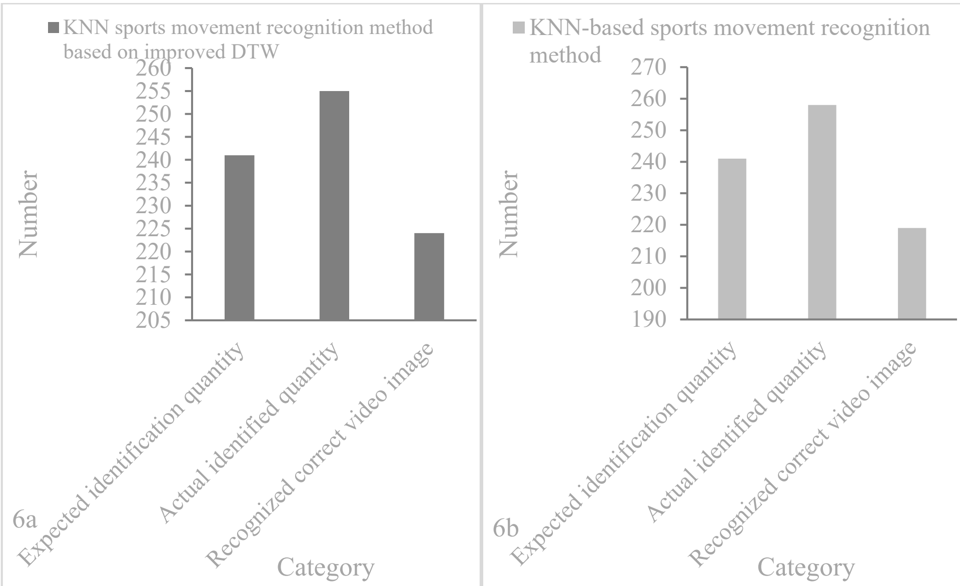
5a: Push-up motion recognition under KNN sports motion recognition method combined with improved DTW

5b: Push-up motion recognition based on KNN sports motion recognition method

Figure 5: Recognition of push-up movements

4.3 Recognition Effect of Squatting Sports

The same number of video image frames of other sports types were added to the squat motion video set. The KNN sports motion recognition method combined with the improved DTW and the KNN-based sports motion recognition method were used to identify the squat motion in the squat motion video set, as shown in Figure 6.



6a: Squatting motion recognition based on improved DTW KNN sports motion recognition method

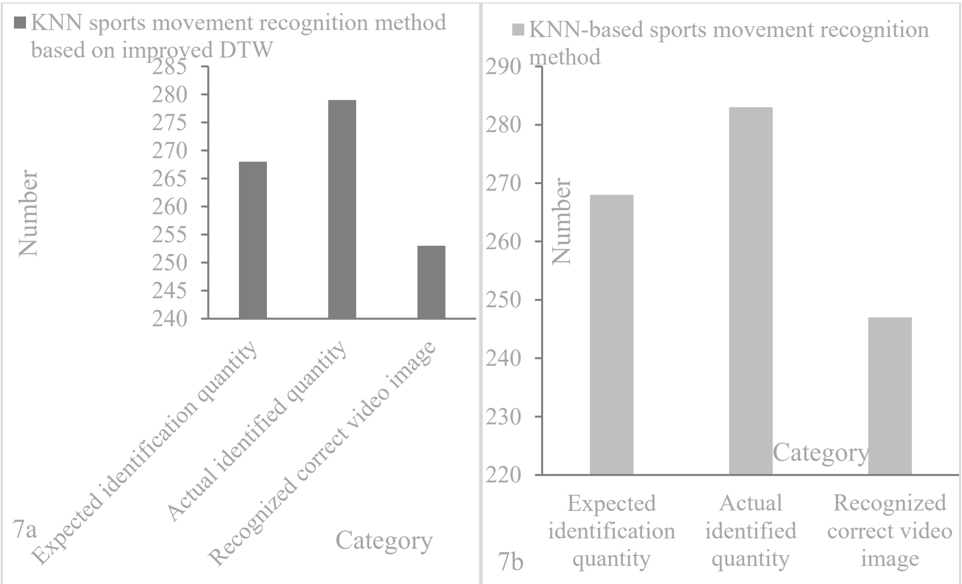
6b: Squatting motion recognition based on KNN

Figure 6: Recognition of squat sports movements

As shown in Figure 6, Figure 6a shows the squat motion recognition results under the KNN sports motion recognition method combined with the improved DTW, and Figure 6b shows the squat motion recognition results under the KNN sports motion recognition method. As can be seen from Figure 6a, in the squatting motion video set, the number of correct video images obtained by combining the improved DTW KNN sports motion recognition method was 224, and the number of recognized motion video images was 255. The accuracy and recall of squatting motion recognition by combining the improved DTW KNN sports motion recognition method were about 87.8% and 92.9%, respectively. It can be seen from Figure 6b that in the squatting motion video set, the number of correct video images obtained by the KNN-based sports motion recognition method was 219, and the number of recognized sports video images was 258. The precision and recall of the KNN-based sports motion recognition method for squatting motion recognition were about 84.9% and 90.9%, respectively.

4.4 Recognition Effect of Lunge Leg Press Movement

The same number of video image frames of other sports types were added to the lunge leg press video set. The KNN sports action recognition method combined with improved DTW and the KNN-based sports action recognition method were used to identify the lunge leg press movement in the lunge leg press video set, as shown in Figure 7.



7a: Recognition of lunge leg press motion under KNN sports motion recognition method combined with improved DTW

7b: Recognition of lunge leg press movement based on KNN sports movement recognition method

Figure 7: The recognition effect of lunge leg pressing sports movements

As shown in Figure 7, Figure 7a shows the recognition results of the lunge leg press movement under the KNN sports action recognition method combined with the improved DTW, and Figure 7b shows the recognition results of the lunge leg press movement under the KNN sports action recognition method. As can be seen from Figure 7a, in the lunge leg press motion video set, the number of correct video images obtained by combining the improved DTW KNN sports motion recognition method was 253, and the number of recognized sports video images was 279. The accuracy and recall of the lunge leg press motion recognition by combining the improved DTW KNN sports motion recognition method were about 90.7% and 94.4%, respectively. It can be seen from Figure 7b that in the video set of the lunge leg press movement, the number of correct video images obtained by the KNN-based sports action recognition method was 247, and the number of recognized sports video images was 283. The accuracy and recall of the KNN-based sports action recognition method for the recognition of the lunge leg press movement were about 87.3% and 92.2%, respectively.

From the comprehensive results of the two methods in the hurdles video set, the push-up video set, the squat video set and the lunge leg press video set, the accuracy and recall of KNN sports motion recognition method combined with improved DTW in sports motion recognition were about 90.4% and 94.9% respectively. The precision and recall of KNN-based sports movement recognition method in sports movement recognition were about 86.2% and 92.5% respectively. The precision of KNN sports action recognition method combined with improved DTW was 4.2% higher than that of KNN sports action recognition method. The precision of KNN sports action recognition method combined with improved DTW was 2.4% higher than that of KNN-based sports action recognition method.

5. Conclusions

This paper first outlined the problems in human action behavior recognition, including the inapplicability of traditional recognition technology, and the multiple meanings of the behavior in the video, and described the template matching method, probabilistic model method and other human action behavior recognition methods. Then, this paper proposed a method of sports action feature extraction based on bone key points, proposed a KNN sports action recognition method combined with DTW, and improved DTW. Finally, this paper studied the effects of the improved DTW KNN sports action recognition method and the KNN-based sports action recognition method in the hurdles video set, the push-up video set, the squat video set and the lunge leg press video set, calculated the precision and recall of the recognition results under the two recognition methods, and drew the conclusion that the improved DTW KNN sports action recognition method had more advantages. Although this study can provide some reference for relevant research, there are still some problems. The data set used in the test in this paper only selected hurdles, push-ups, squats and lunges, and tried to avoid the situation of limb occlusion when collecting the data set. In the future research, people would consider the recognition of sports actions when the limb is blocked, and would increase the types of sports actions.

References

[1] Zhang, Xiaojun. "Application of human motion recognition utilizing deep learning and smart wearable device in sports." International Journal of System Assurance Engineering and Management 12.4 (2021): 835-843.
[2] Cust, Emily E. "Machine and deep learning for sport-specific movement recognition: a systematic review of model development and performance." Journal of sports sciences 37.5 (2019): 568-600.
[3] Yang, Yun. "Human movement monitoring and behavior recognition for intelligent sports using customizable and flexible triboelectric nanogenerator." Science China Technological Sciences 65.4 (2022): 826-836.
[4] Rahmad, Nur Azmina. "A survey of video based action recognition in sports." Indonesian Journal of Electrical Engineering and Computer Science 11.3 (2018): 987-993.
[5] Dong, Zhe, and Wang, Xiongying. "Automatic recognition of javelin athletes' throwing angle based on recognisable statistical characteristics." International Journal of Biometrics 14.3-4 (2022): 408-420.
[6] Gurbuz, Sevgi Zubeyde, and Moeness G. Amin. "Radar-based human-motion recognition with deep learning: Promising applications for indoor monitoring." IEEE Signal Processing Magazine 36.4 (2019): 16-28.
[7] Jaouedi, Neziha, Noureddine Boujnah, and Med Salim Bouhlel. "A new hybrid deep learning model for human action recognition." Journal of King Saud University-Computer and Information Sciences 32.4 (2020): 447-453.
[8] Wang, Jucui. "Research on contour feature extraction method of multiple sports images based on nonlinear mechanics." Nonlinear Engineering 11.1 (2022): 347-354.
[9] Wang, Peng. "Deep learning-based human motion recognition for predictive context-aware human-robot collaboration." CIRP annals 67.1 (2018): 17-20.
[10] Hendry, Danica. "Development of a human activity recognition system for ballet tasks." Sports medicine-open 6.1 (2020): 1-10.
[11] Rangasamy, Keerthana. "Deep learning in sport video analysis: a review." TELKOMNIKA (Telecommunication Computing Electronics and Control) 18.4 (2020): 1926-1933.
[12] Kong, Longteng. "A joint framework for athlete tracking and action recognition in sports videos." IEEE transactions on circuits and systems for video technology 30.2 (2019): 532-548.
[13] Ullah, Mohib. "Attention-based LSTM network for action recognition in sports." Electronic Imaging 2021.6 (2021): 302.1-302.6.
[14] Li, Xinyu. "Human motion recognition with limited radar micro-Doppler signatures." IEEE Transactions on Geoscience and Remote Sensing 59.8 (2020): 6586-6599.
[15] Gao, Peijun, Zhao, Dan, and Chen, Xuanang. "Multi-dimensional data modelling of video image action recognition and motion capture in deep learning framework." IET Image Processing 14.7 (2020): 1257-1264.