

Research Article

Research on Aerobics Training and Evaluation Method Based on Artificial Intelligence-Aided Modeling

Chen Chen 

College of Physical Education, Fuyang Normal University, Fuyang, Anhui, China

Correspondence should be addressed to Chen Chen; 201707033@fynu.edu.cn

Received 15 October 2021; Revised 28 October 2021; Accepted 5 November 2021; Published 18 November 2021

Academic Editor: Le Sun

Copyright © 2021 Chen Chen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Traditional aerobics training methods have the problems of lack of auxiliary teaching conditions and low-training efficiency. With the in-depth application of artificial intelligence and computer-aided training methods in the field of aerobics teaching and practice, this paper proposes a local space-time preserving Fisher vector (FV) coding method and monocular motion video automatic scoring technology. Firstly, the gradient direction histogram and optical flow histogram are extracted to describe the motion posture and motion characteristics of the human body in motion video. After normalization and data dimensionality reduction based on the principal component analysis, the human motion feature vector with discrimination ability is obtained. Then, the spatiotemporal pyramid method is used to embed spatiotemporal features in FV coding to improve the ability to identify the correctness and coordination of human behavior. Finally, the linear model of different action classifications is established to determine the action score. In the key frame extraction experiment of the aerobics action video, the ST-FMP model improves the recognition accuracy of uncertain human parts in the flexible hybrid joint human model by about 15 percentage points, and the key frame extraction accuracy reaches 81%, which is better than the traditional algorithm. This algorithm is not only sensitive to human motion characteristics and human posture but also suitable for sports video annotation evaluation, which has a certain reference significance for improving the level of aerobics training.

1. Introduction

Aerobics aims to show athletes' ability to perfectly complete difficult movements and continue complex and high-intensity sports under the accompaniment of music. How to carry out scientific training is the basis and key link of excellent aerobics, which has become an important scientific research topic [1]. Traditional aerobics training methods are based on the analysis of previous competitions and training cases, they lack auxiliary teaching and real-time analysis, and the training efficiency is low. Artificial intelligence and computer-aided system is a computer science and technology, including robot, language, image recognition, and expert system [2]. In recent years, artificial intelligence technology has been widely used in aerobics teaching and practice. This is of great significance to improve the training quality of aerobics athletes and help coaches formulate reasonable sports plans.

Some teachers' traditional training ideas are deeply rooted and difficult to accept the new training mode. In the concept of many aerobics teachers, aerobics training does not need the aid of multimedia teaching means. It only needs to demonstrate aerobics movements and then let students practice [3]. This traditional teaching concept mainly exists in some older aerobics teachers. Because of the deep-rooted teaching concept and poor adaptability, it is difficult to accept new ideas and keep pace with the times. There are loopholes in the Computer-Aided Instruction (CAI) system. Although a few schools in our country have begun to use the computer-aided teaching system to carry out aerobics teaching activities [4], but due to the limited scope of application, there are still some loopholes in the system, such as the lack of teaching feedback module in the established computer-aided teaching system, so there are still some problems in the information communication with students, which cannot let students learn well. It is difficult to

understand students' views and opinions on aerobics teaching. Some teachers have not fully played a guiding role in the dependence of the computer-aided instruction system. Since the CAI system is a kind of auxiliary means, the teaching theme should be the interactive process between teachers and students. However, with the help of the computer-aided instruction system, many teachers have developed a kind of dependence on it. The aerobics teaching mode has changed from the previous demonstration mode to the video demonstration mode and the students' self-practice mode. This dependence on the computer-aided instruction system not only fails to achieve the purpose of further promoting the effect of students' aerobics practice but also causes students to have confused psychology under the condition of teachers' more relaxed supervision, which will affect the teaching quality [5]. Therefore, although the computer-aided instruction system helps teachers to share a lot of work, the teacher's teaching guidance function is still irreplaceable. In the actual training process, teachers need to master the computer-aided instruction system in the teaching process of the component to avoid making it become the dominant aerobics teaching.

From the problems existing in the practice of intelligent computer-aided aerobics training, it is worth in-depth research on how to carry out aerobics training based on intelligent auxiliary software. Therefore, starting from the design of the computer-aided aerobics training practice system, aiming at the problems existing in the intelligent computer-aided aerobics training application system, this paper puts forward the aerobics feature extraction method under intelligent computer-aided modeling and gives the aerobics evaluation scheme. Firstly, in order to maintain the spatiotemporal continuity of the FMP model in motion video human body pose estimation, the ST-FMP model with spatiotemporal continuity is obtained by establishing time continuity constraints between the vertex pairs of human parts in adjacent frames, and then, the ST-FMP model is simplified by embedding spatiotemporal constraints between uncertain human parts in adjacent video frames. The N-best optimization algorithm is used to estimate the human posture parameters, which improves the efficiency of solving the human posture parameters in a motion video. Then, the relative position characteristics and motion direction of each part of the human body are used to describe the motion characteristics of each part of the human body, and the Laplace scoring algorithm is used for feature selection to form a locally discriminated human motion feature vector. Finally, the key frames of the motion video are determined by the ISODATA dynamic clustering algorithm. Finally, the effectiveness of intelligent assistant software in aerobics training is analyzed through experiments.

2. Related Works

An intelligent computer-aided aerobics training application system is the basis of realizing intelligent-assisted training. When constructing the application system of

intelligent computer-aided aerobics training, we need to determine the framework of the system first. The framework of the intelligent computer-aided aerobics training system is shown in Figure 1. The determination of the frame structure is the first step to establish the intelligent computer-aided instruction system, which needs to be carried out under the premise of making clear the purpose and task of aerobics training. Then, combined with the focus of the system, the overall framework is designed on the basis of the syllabus. In order to make aerobics teaching activities more effective, the design of the overall framework structure should be concise, intuitive, clear, and interrelated [6]. Combined with aerobics training materials, the design of the computer-aided teaching system is divided into four main modules: basic theory of action module, training site and evaluation principle module, action training and difficult motion playback module, and assessment and evaluation module.

The basic theory module of action mainly includes the common research methods of aerobics technical movements, sports teaching theory, training methods, and other related units. The teaching site and evaluation principle module mainly includes aerobics formation layout, clothing specifications and related decoration, judgment standard, referee's responsibility, analysis of common judgment errors of aerobics referee, and troubleshooting unit in the form of human-computer interaction. The motion teaching and key and difficult point motion playback module mainly includes the technical action analysis unit, essential analysis unit, basic teaching method unit, error prone action analysis unit, standard action demonstration unit, key and difficult motion analysis unit, and action video decomposition teaching unit.

The CAI system contains rich teaching resources, including text, animation, video, sound, and pictures. Therefore, in the process of establishing the computer-aided instruction system, the collection of relevant materials is very important. In order to enrich the teaching resources of the CAI system as much as possible, in the preparation stage of materials, we need to follow the following procedures:

- (i) Ask experienced aerobics teachers to determine the scope of CAI materials and select qualified materials according to specific teaching objectives and tasks. After determining the material, you can write the design text, record the sound, and choose the appropriate background music.
- (ii) Shooting video teaching resources, including standard action video and wrong action video, for students to learn and learn from.
- (iii) Take a video of referees in aerobics competitions. Using a JVC digital camera, the organization, division of labor, and gesture process of aerobics competition referees are completely photographed. Finally, with the help of 3D animation technology and image processing software, the materials are assembled into teaching resources with teaching value and inserted into the corresponding plate after the completion of the computer-aided instruction system.

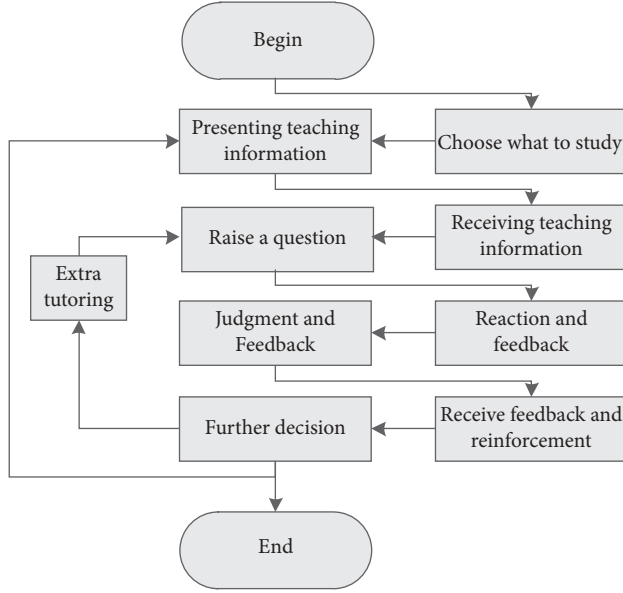


FIGURE 1: Framework of intelligent computer-aided aerobics training system.

Finally, a computer-aided instruction system can be established. Authorware 6.0 is a kind of multimedia visualization development work, which is based on the flowchart, so it can be selected as the operation platform of aerobics calculation and auxiliary teaching system [7]. Based on the Authorware 6.0 development platform, the calisthenics intelligent computer-aided teaching system can easily realize the classification of aerobics teaching modules and store the corresponding teaching resources. Then, the navigation icon is used to convert different modules, the framework is used to turn pages of different teaching contents, the scroll function of the text is used to realize the continuous reading of a long text, and the “hot area response” and “button response” in the interactive icon help students to convert between image, text, and video more easily and use teaching resources conveniently.

No matter students or teachers, as long as they click the corresponding module, they can get the required teaching resources on the corresponding window and easily realize the jump between different modules. At the same time, the human-computer interaction mode can reflect the problems existing in the process of students’ aerobics learning more carefully and then continuously improve, greatly improving the effect of aerobics teaching. In addition, the real-time analysis and assessment mode in the comprehensive evaluation module can also exercise the students’ observation ability, analysis ability, and comprehensive control ability, improve the interest of aerobics teaching, attract students to explore constantly, and then make progress in the process of exploration. The establishment of the intelligent computer-aided instruction system overcomes the problems of monotonous teaching mode and low-teaching effect in the past, stimulates students’ learning enthusiasm, and makes the teaching quality achieve a qualitative leap.

3. Aerobics Feature Extraction Based on Intelligent Computer-Aided Modeling

3.1. Human Pose Estimation Model with Spatiotemporal Features. With the application of intelligent computers aided in aerobics training, it is urgent to develop an adaptive key frame extraction technology which can accurately reflect the characteristics of human body movements. Considering the good performance of the pose recognition method of the rigid human body model [8], this paper proposes to embed the temporal and spatial characteristics of human body parts in the flexible hybrid articulated human body model (FMP), so as to improve the robustness of human body and motion recognition and determine the key frames of an aerobics athletes’ action video by using the human posture parameters and motion characteristics.

In the FMP model, different affine deformations (such as rotation or bending) of each part of the human body are called the mixed type of the part, which is referred to as the mixed type [9]. Because the same human body part corresponds to several mixing types, the human posture parameters in an image I are determined by the position information of each part and their mixed types. In general, a K -relation graph $G = (V, E)$ is used to describe a pose in I . The vertex set V represents the human body parts (such as the head, upper limbs, and trunk), and the edge set $E \subseteq V \times V$ represents the consistent constraint relationship between different human positions. According to the definition of FMP, the problem of parameter estimation of the human posture p in an image I can be formalized as the problem of cost minimization, which is shown as follows:

$$C(I, p) \propto \sum_{u \in V} \varphi_u(I, p^u) + \sum_{(u,v) \in E} \phi_{u,v}(p^u - p^v), \quad (1)$$

where $\varphi_u(I, p^u)$ is an appearance model, which represents the cost of recognizing the human body part u at the position p^u of the image I ; $\phi_{u,v}(p^u - p^v)$ is a deformation model (usually assumed to be a spring energy model), which represents the deformation cost between two human parts u and v .

When using the FMP model to estimate the body posture in an Aerobics Athletes’ action video, in order to maintain the continuity of human posture parameters in time and space, a temporal feature edge is added between p_t^u and p_{t+1}^u in adjacent frames I_t and I_{t+1} to form a flexible hybrid articulated human model (ST-FMP) with spatiotemporal characteristics [4]. In ST-FMP, the continuous error of the human posture defined by a time-series feature edge is calculated by an optical flow difference between p_t^u and p_{t+1}^u :

$$\theta(p_t, p_{t+1}, I_t, I_{t+1}) = \sum_{u \in V} \frac{\|p_{t+1}^u - p_t^u - f(p_t^u)\|_2^2}{2}, \quad (2)$$

where $f(p_t^u)$ is the optical flow value from I_t to I_{t+1} estimated at put. Suppose that the frame image set of an aerobics athlete’s action video is $I = \{I_1, I_2, \dots, I_T\}$. The estimated attitude parameter sequence is $P = \{p_1, p_2, \dots, p_T\}$. Then, using the ST-FMP model, we can obtain the cost of I as follows:

$$C(I_T, p_T) + \sum_{t=1}^{T-1} C(I_t, p_t) + \lambda_1 \theta(p_t, p_{t+1}, I_t, I_{t+1}), \quad (3)$$

where $C(I_T, p_T)$ represents the cost of the estimating human posture p_T from the image I_T obtained according to formula (1), t is the frame number of an aerobics athletes' action video, λ_1 is the normalized constant, and $\theta(\cdot)$ is the spatiotemporal continuity error expressed in equation (2).

3.2. ST-FMP Solution Based on Uncertain Position Optimization. FMP is a rigid human body model, which can be represented by a Markov random field (MRF). The parameters of human body parts can be determined by the machine learning method. When FMP is used to estimate human posture parameters in a single frame image, MRF is regarded as a tree or star graph structure and is solved by confidence propagation BP. With the introduction of time constraint, a large number of loops will be generated in ST-FMP, which needs to be solved by minimizing equation (3) with an approximate algorithm such as cyclic confidence propagation (LBP) [10]. However, the LBP algorithm is a graph maximum clique problem, which has exponential complexity and low-time efficiency in a long video pose estimation. Therefore, this paper designs a two-stage ST-FMP algorithm based on uncertain human body parts.

3.2.1. Generating the Candidate Human Pose Sets. Firstly, the N-best algorithm and formula (1) are used to generate K_b human pose sets from a single frame image in $O(K_b^2 T)$ time without considering the space-time continuity constraint of human posture [11]. Due to motion blur and self-occlusion, some human body parts in K_b , such as the elbow (LE, RE), wrist (LW, RW), knee (LK, RK), and ankle (LA, RA), identified by eight white points in Figure 2(a), are difficult to estimate accurately. Therefore, in this paper, the uncertain part of I_{t+1} and \tilde{p}_{t+1} is introduced into the attitude estimation of I_T . With the help of local time continuity of human body parts (represented by four dotted lines in Figure 2(a)), the accuracy of the human posture estimation is improved as follows:

$$C(I_t, p_t) + \tilde{C}(I_{t+1}, \tilde{p}_{t+1}) + \tilde{\lambda}_1 \sum_{u \in W} \|\tilde{p}_{t+1}^u - p_t^u - f(p_t^u)\|_2^2, \quad (4)$$

where $W \subset V$ represents the set of uncertain parts and $\tilde{\lambda}_1$ is the normalized constant; except that $\tilde{C}(\cdot)$ only calculates the appearance cost and deformation cost of uncertain parts in W , the meaning of $\tilde{C}(\cdot)$ in formula (1) is the same.

3.2.2. Determining the Optimal Solution. After obtaining the K_b attitude parameters of each frame image, the dynamic programming algorithm is used to determine the optimal attitude parameters of each frame in $O(K_b^2 T)$ time by minimizing formula (3) as follows:

$$\min_{p_t \in P_t, \forall t} C(I_T, p_T) + \sum_{t=1}^{T-1} C(I_t, p_t) + \lambda_1 \theta(p_t, p_{t+1}, I_t, I_{t+1}), \quad (5)$$

Since the two steps of the ST-FMP method can be completed in $O(K_b^2 T)$, the two-stage ST-FMP algorithm based on uncertain human body parts significantly reduces the time complexity of the LBP-based ST-FMP algorithm.

3.3. Description of Motion Characteristics of Human Body Parts. At present, the human motion model used for motion capture and motion recognition cannot be used directly because of the lack of accurate motion parameters such as joint angular velocity and displacement velocity. Therefore, this paper designs a human motion feature description model based on relative position characteristics and motion direction of human body parts [12]. If the upper left corner of a single frame image is taken as the coordinate origin, the width and height directions are x and y axes, and pixels are taken as the unit to establish the coordinate system, it is assumed that $C^u = (C_x^u, C_y^u)$ is the center position coordinate of the human body part p_u , and (C_x, C_y) is the position coordinate of human center of gravity. In this paper, the position information $l_u = (x^u, y^u)$ of a human body part is defined as the relative position between (C_x, C_y) and (C_x^u, C_y^u) , which is shown as follows:

$$l_u = (x^u, y^u) = (C_x^u - C_x, C_y^u - C_y). \quad (6)$$

At the same time, the motion direction V_t^u of p_u at time t is defined as the composite vector of the motion directions of all moving points in p_u shown as follows:

$$V_t^u = \frac{1}{|\xi_t^u|} \sum_{v^{u,i} \in \xi_t^u} v^{u,i}, \quad (7)$$

where $v_t^{u,i}$ are the movement directions of the i_{th} pixel in p_u at time t and ξ_t^u is the collection of all moving points in p_u at time t . In this paper, we estimate $v_t^{u,i}$ by comparing the dense optical flow of adjacent frames I_{t+1} and I_{t+1} . With the help of formulas (6) and (7), the feature matrix composed of motion direction and relative position information of human body parts can be used to represent the human motion features in a frame image. Suppose that $j_u = (x^u, y^u, V_t^u) \in R^3$ is the motion characteristic of p_u at time t , and x is the motion feature of the image I_i containing d ($d = 26$) human parts at t time, then the motion characteristics of an aerobics athlete action video with the frame number T can be expressed as $f = [J_1, J_1, \dots, J_T] \in R^{3 \times d \times T}$.

According to the above definition, J_i is the 78-dimensional motion vector and f is the $78 \times T$ vector matrix. Experiments show that the time complexity of the key frame extraction in $78 \times T$ high-dimensional motion feature space is high, and a large amount of data redundancy and noise information will directly affect the accuracy of the key frame extraction. In order to improve the expression ability of local features of motion vectors, Laplacian scorching (LS) [13] is used to reduce the dimension of motion vectors to

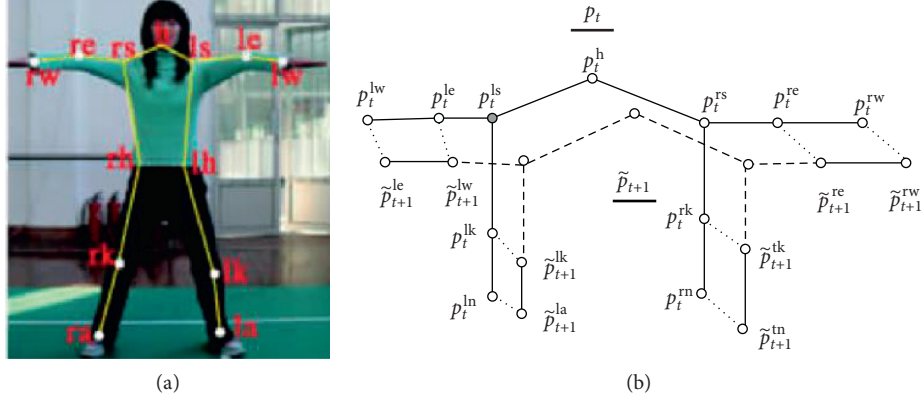


FIGURE 2: Local action time continuity of uncertain parts. (a) Nondeterministic position and (b) time continuity constraint.

determine more discriminative human motion features. Firstly, a k -nearest neighbor graph G_k is constructed; then, the similarity of two connected nodes in G_k is calculated by using the thermal kernel function to obtain the Laplacian score L_r of the r_{th} motion feature; finally, the first n ($1 \leq n \leq 3d$) motion features with smaller L_r are determined as the body posture feature vector of the aerobics athlete's action video.

4. Aerobics Evaluation Based on Intelligent Computer-Aided Modeling

Usually, the differences between two aerobics athletes in an action video about human body movements are mainly manifested in the differences of the body posture, shape, and movement speed. Therefore, this paper uses HOF and HOG to extract the human motion features in the aerobics athletes' action video and completes the coding of human motion features in the video by fisher vector technology.

4.1. Human Posture Shape Feature Extraction. The posture feature of the body movement in an aerobics video is a static local topological structure of human body parts. The existing research results show that, even if the position of the gradient direction cannot be accurately obtained, the local dense gradient feature (HOG feature) can accurately represent the static features such as the human shape and position [14]. Therefore, the posture characteristics of aerobics athletes in an action video can be described by HOG characteristics. Figure 3 shows the HOG characteristics of an aerobics athlete after visualization.

Firstly, the frame image of 240×320 pixels is divided into different cells according to the size of 40×40 pixels. Then, the $[1, 0, 1]$ template is used to convolute the horizontal and vertical directions of the image, and the gradient size and direction of each pixel are calculated as follows:

$$M(x, y) = \sqrt{I_x^2 + I_y^2}, \quad (8)$$

$$\theta(x, y) = \tan^{-1} \frac{I_y}{I_x} \in [0^\circ, 180^\circ),$$

where I_x and I_y represent the horizontal and vertical gradient values, respectively; $M(x, y)$ represents the gradient size; and $\theta(x, y)$ represents the gradient direction.

Then, the gradient direction from 0° to 180° is divided into 16 bins. At the same time, every four cells are combined with a block of 80×80 pixels. In the same block, the Euclidean norm of histogram is used to normalize each cell. Since each unit contains 16-dimensional feature vectors, each block has only one $4 \times 16 = 64$ -dimensional feature vector. The local features of the human posture and shape in each frame image are represented by hog feature vectors with a dimension of $6 \times 8 \times 16 = 768$.

4.2. Human Motion Feature Extraction. According to the principle of the optical flow field, the gray change of a human body image in adjacent video frames can better reflect the motion characteristics of the human body [15]. Therefore, in this paper, the Lucas Kanade algorithm (LK algorithm for short) [16] is used to estimate the optical flow vector of each point in the image, and the motion information of the human body is determined by the optical flow histogram (HOF) feature [17]. When the LK algorithm is used, it is assumed that the optical flow remains unchanged in a small local area of the adjacent frame images. Then, the velocity characteristics of the human motion from time t to time $t + \Delta t$ can be determined by the least-squares method. According to the LK algorithm, after obtaining the optical flow of each pixel in the frame image, this paper first divides a 240×320 -pixel frame image into $6 \times 8 = 48$ cells and then discretizes the optical flow vector in each cell into 16 bin directions and normalizes them to obtain the $6 \times 8 \times 16 = 768$ -dimensional HOF feature vector.

4.3. Automatic Scoring of Aerobics Athletes' Action Video. In order to complete the automatic scoring of aerobics athletes' action videos, the aerobics athletes' action videos in the training dataset are divided into three categories according to the manual scoring results and the action completion quality. The STLPFV algorithm is used to obtain the action features of each segment of aerobics athletes' action videos, and then the k -neighbor algorithm is used to

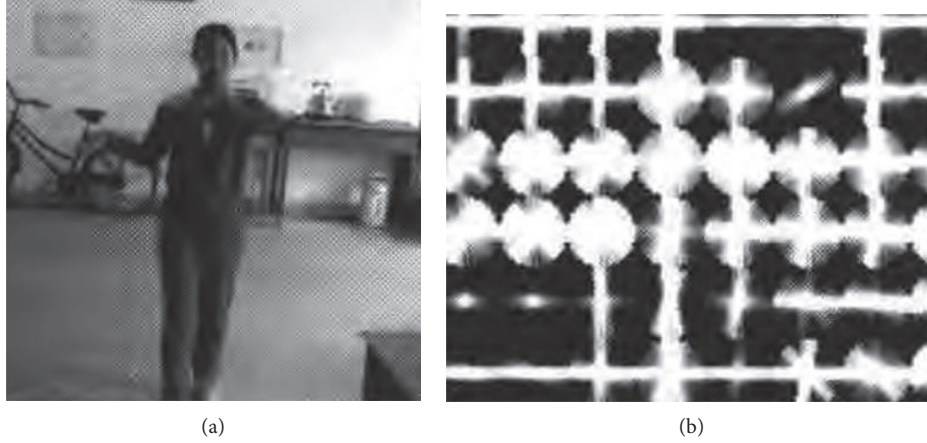


FIGURE 3: HOG characteristics of aerobics athletes' action visualization. (a) The action of aerobics athlete and (b) action visualization HOG feature.

complete the motion feature modeling. When the aerobics player's action video is scored automatically, the FV code of the target aerobics athlete's action video is generated by the STLPFV algorithm [18]. Finally, the KNN distance between the action video of target aerobics athletes and the action video of aerobics athletes with three different scores is classified, and the least-squares method is used for automatic scoring [19].

$C = \{\bar{c}_l, \bar{c}_m, \bar{c}_h\}$ is the classification center set of three kinds of actions obtained by the k -nearest neighbor algorithm. If a triple $KNN(s_i) = p_i = [p_i^{(l)}, p_i^{(m)}, p_i^{(h)}]$ is used to represent the KNN distance vector between the target aerobics athletes' action video s_i and C (where $p_i^{(l)}, p_i^{(m)}, p_i^{(h)}$, respectively, represent the probability values of s_i belonging to three different score action videos), then the weight of s_i on C is as follows:

$$w_i = p_i^{(l)}\bar{c}_l + p_i^{(m)}\bar{c}_m + p_i^{(h)}\bar{c}_h. \quad (9)$$

Weight w_i obtained by formula (9) needs to be calibrated as the corresponding action score. In this paper, the least-square method and action video training set are used to establish the linear relationship model between weight w_i and action score. The specific method is to assume that $Y^{(\alpha)} = \{y_i^{(\alpha)} | i = 1, 2, \dots, n\}$ is the score set of all kinds of actions in C (where n represents the number of action videos and $\alpha \in \{l, m, h\}$ represents three types of action videos with different scores, i.e., low score, medium score, and high score), which is shown as follows:

$$(\bar{p}^{(\alpha)}, w_i) = \arg \min_{y_i^{(\alpha)} \in Y^{(\alpha)}} |y_i^{(\alpha)} - (\rho_0^{(\alpha)} w_i + \rho_1^{(\alpha)})|^2, \quad (10)$$

where $\bar{p}^{(\alpha)} = \{\bar{p}_0^{(\alpha)}, \bar{p}_1^{(\alpha)}\}$ represents the coefficient value vector obtained by solving formula (10) of an action video training set. At this time, the target aerobics athletes' action video s_i corresponding score is defined as follows:

$$S_i^*(\bar{p}^{(\alpha)}, w_i) = \bar{p}_0^{(\alpha)} w_i + \bar{p}_1^{(\alpha)}. \quad (11)$$

Accordingly, the flow of the automatic scoring algorithm of an aerobics athletes' action video based on the STLPFV method is shown in Figure 4. It includes the following parts:

- (1) The STLPFV method is used to calculate the human motion characteristics in the motion video s_i of aerobics athletes, and the FV code is obtained.
- (2) The KNN distance between s_i and C is calculated by the k -neighbor algorithm, and the probability $KNN(s_i)$ that s_i belongs to all kinds of action scores in C is obtained. Weight w_i is calculated by formula (11).
- (3) According to the KNN distance, s_i is classified as the first kind of a motion video.
- (4) The action score s_i^* corresponding to s_i is calculated by equation (11).

5. Experiment and Analysis

This paper takes the key frame of aerobics action video as an example to carry out the simulation experiment. The experimental results are compared with the results of the manual extraction and the extraction results of the latest aerobics athletes' action video key frame algorithm [20].

5.1. Training and Evaluation Criteria of Experimental Data Samples and Characteristics. At first, three students were invited to do 120 s public aerobics twice, and a video with 640×480 resolution was recorded by an ordinary webcam at the sampling frequency of 20 frames/S. Then, 300 images were selected from each of the six sets of videos as experimental data from the 10th rhythm. Finally, in 1800 frames of images, according to Figure 2(a), 13 joint positions of each aerobics movement are manually marked. In the simulation experiment, 900 images of the first pass and the second pass are selected as the training sample dataset and the test sample dataset.

In the process of training, in order to enrich the human motion characteristics of the positive samples, the rotation (according to the four angles of -15° , -7.5° , 7.5° , and 15°) and

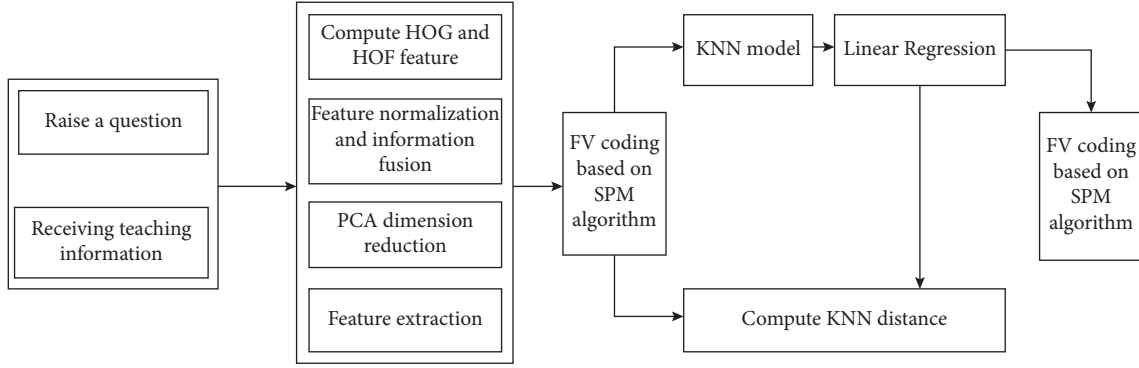


FIGURE 4: The automatic scoring process of aerobics athletes in STLPFV.

mirror image operation are used to expand 900 training samples into 10×900 positive samples, which are recorded as Φ_p^* . Then, by adding key points at the midpoint of two adjacent joint positions in Figure 2(a), the coordinates of 26 flexible human body parts in each image in Φ_p^* are calculated. Finally, the HOG features of each human body part are calculated by using a 5×5 -pixel HOG unit, and parameters such as $\varphi(I, l_u)$ and $\phi(l_u, -l_v)$ in formula (3) are determined by the linear support vector machine.

All the key frames in the test sample set are extracted manually, and the common key frame extraction accuracy is taken as the evaluation standard of algorithm performance [21], which is shown as follows:

$$\text{Accuracy} = \sum_{i=1}^m \frac{\delta(f_i, r_i)}{n}, \quad (12)$$

where n and m represent the number of key frames extracted manually; f_i and r_i represent the key frames extracted by algorithm, respectively; and $\delta(\cdot)$ is the similarity function between f_i and r_i . When f_i and r_i are the same, $\delta(\cdot)$ value is 1; otherwise, it is 0.

5.2. Comparison of the Effectiveness of Spatiotemporal Feature Embedding in Uncertain Parts. In order to detect the accuracy of the manikin embedded with a uncertain part of the temporal and spatial features in the video, three different ST-FMP models are used to realize the human body parts, and the comparative experiments of elbow and knee parts are carried out according to different error pixel thresholds. The experimental results are shown in Figure 5.

As can be seen from Figure 5, compared with the FMP model, the accuracy of the uncertain part can be significantly improved within a certain pixel error range by using the ST-FMP algorithm to estimate the human posture in the aerobics athlete's action video. Taking the error threshold of 20 pixels as an example, the accuracy of the elbow and knee obtained by the ST-FMP algorithm is about 15% and 19% higher than that of the FMP model, respectively. However, when the pixel error threshold is large (e.g., greater than 40 pixels) or small (such as greater than 10 pixels), the accuracy difference is not significant.

According to Figure 5, only when the time continuity of the uncertain part of the upper limb (lower limb) is

maintained, the recognition accuracy of the elbow and wrist (knee and ankle) will be higher than that of the FMP model, but lower than that of the ST-FMP algorithm. The experimental results show that the ST-FMP algorithm can significantly improve the recognition performance of nondeterministic parts in aerobics athletes' action videos by optimizing the recognition results of human body parts by local time continuity constraints.

This paper also compares the performance of the FMP model, ST-FMP model, and their different implementations in the key frame extraction of the aerobics athletes' action video. The experimental results are shown in Figure 6.

According to the results shown in Figure 6, two conclusions are obtained as follows:

- (1) When the accuracy error is less than 30 pixels, the accuracy of the ST-FMP algorithm is higher and more stable, which is about 11% higher than that of the FMP algorithm. In addition, when the time continuity constraint of the upper limb (lower limb) is added, the extraction accuracy of FMP model key frames will be improved by about 3%.
- (2) When the error accuracy is greater than 35 pixels, the performance of the ST-FMP algorithm is still better than that of the FMP model, but the accuracy rate is reduced by about 15%. At the same time, when the accuracy error is 30 pixels and the key frames are extracted by the ST-FMP algorithm with the different number of motion features, the accuracy curve of the algorithm does not fluctuate violently, and the performance is stable in the range of 15–60 motion feature numbers, as shown in Figure 7. The experimental results show that the ST-FMP algorithm is sensitive to human posture estimation results and the local topological structure of the human body, and spatiotemporal constraints of uncertain parts play an important role in key frame extraction performance.

5.3. Performance Comparison of Key Frame Algorithms. In order to compare the performance of the key frame algorithm, the simulation experimental results of the ST-FMP algorithm are compared with the operation results of the KFE algorithm based on priori in reference [22] and the

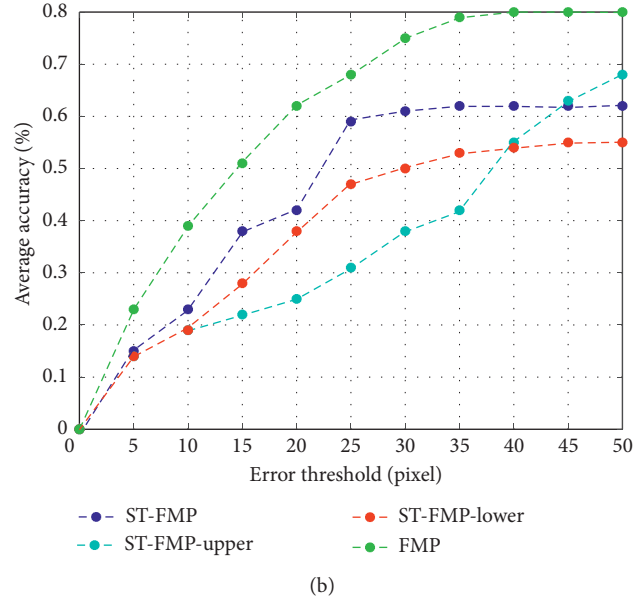
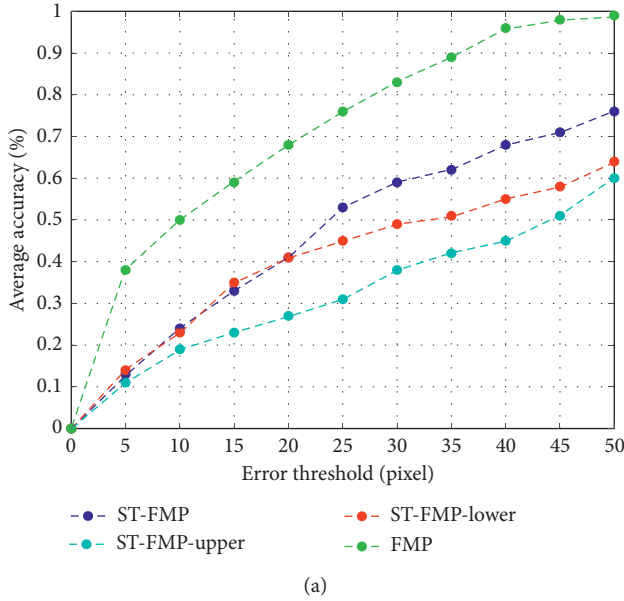


FIGURE 5: Comparison of recognition accuracy of human body parts under different ST-FMP. (a) Elbow joint and (b) knee joint.

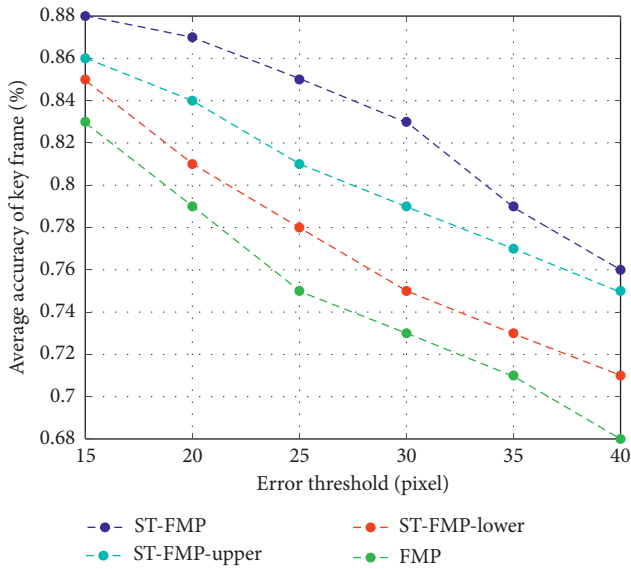


FIGURE 6: Experimental results of key frame extraction based on FMP and ST-FMP.

motion block-based key frame extraction algorithm (referred to as the motion block algorithm) in reference [23], as shown in Table 1.

The experimental results in Table 1 show that the accuracy and recall rate of the ST-FMP algorithm are better than the other two algorithms. First of all, as can be seen from Table 1, the accuracy rate of the ST-FMP algorithm is about 18% and 26% higher than KFE algorithm and motion block algorithm, respectively. The KFE algorithm uses predefined motion directions of 16 blocks to represent human motion features, while the ST-FMP algorithm uses the first 15 LS human motion pose eigenvalues of each action video to represent the human motion. Therefore, the ST-FMP algorithm uses less redundant

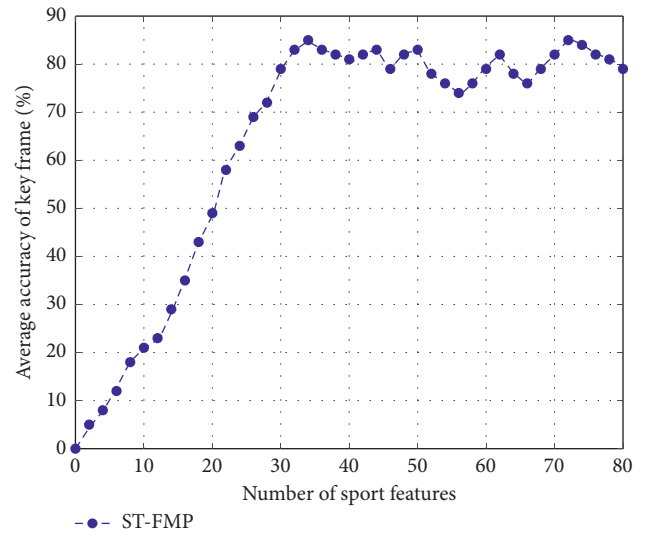


FIGURE 7: Key frame accuracy comparisons of different motion feature numbers.

motion feature vectors, less noise, and accurate expression of local motion of human body parts, which is conducive to improving the accuracy of key frame and action recognition.

Secondly, it can be seen from Table 1 that the recall rate of the ST-FMP algorithm is significantly better than the other two algorithms, with an average of 23 and 13 percentage points higher, respectively. The KFE algorithm and motion block algorithm belong to key frame technology based on the difference of image bottom features. They select key frames by comparing the motion changes in different regions of the image. The ST-FMP algorithm describes the local motion characteristics of human body parts, which is a semantic model in essence. It can analyze and understand the human actions in the aerobics athletes' action video from

TABLE 1: Performance comparison of different key frame extraction algorithms.

Algorithms	Accuracy (%)	Recall (%)
ST-FMP	0.8148	0.8287
KFE	0.6358	0.5946
Sport block	0.5583	0.6982

the higher level aspects such as the participation the parts of the human body movement and its movement change trend. By using semantic rules such as the human posture similarity to select key frames, it can obtain more accurate and consistent with people's cognitive process key frame results.

The experimental results show that the ST-FMP algorithm can better express the local topology of the human body and support the key frame selection based on semantic rules. It is not only closer to the manual extraction results but also more suitable for the key frames based on feature extraction. At the same time, because the ST-FMP divides human body parts into different flexible parts, identifies the human posture through the local topology of flexible parts, and reduces continuous error of the human posture estimation by using temporal feature edge constraint, so it has strong robustness in complex scenes.

6. Conclusion

Aiming at the problems of lack of auxiliary teaching conditions and low-training efficiency in traditional aerobics training methods, this paper proposes a local space-time preserving Fisher vector (FV) coding method and monocular motion video automatic scoring technology. The experimental results show that the ST-FMP model significantly improves the recognition accuracy and attitude estimation performance by maintaining the spatiotemporal continuity of uncertain parts of the human body and embedding the spatiotemporal characteristics of local actions, and the key frame set obtained is more in line with people's cognitive process. The algorithm proposed in this paper still has some problems to be improved and optimized. For example, the intelligent computer-aided training system lacks the function of real-time tracking and monitoring in the process of aerobics training. Therefore, how to monitor and evaluate the level of aerobics training in real time according to the human movement characteristics and human posture sensitivity in aerobics training is the next problem to be solved [24].

Data Availability

The dataset can be accessed upon request to the corresponding author.

Conflicts of Interest

The author declares no conflicts of interest.

Acknowledgments

This work was supported by the Fuyang Normal University Industry-University-Research Fund Project: Research on the

preparation and implementation of special skills training for physical education teachers.

References

- [1] N. Rinaldo, E. Bacchi, G. Coratella et al., "Effects of combined aerobic-strength training vs fitness education program in COPD patients," *International Journal of Sports Medicine*, vol. 38, no. 13, pp. 1001–1008, 2017.
- [2] C. Xu, "A novel recommendation method based on social network using matrix factorization technique," *Information Processing & Management*, vol. 54, no. 3, pp. 463–474, 2018.
- [3] J. Zhao, "Application design of intelligent computer aided instruction system in aerobics teaching," in *Proceedings of the International Conference on Cyber Security Intelligence and Analytics*, pp. 512–517, Springer, Haikou, China, February 2020.
- [4] S. Qian, "The damping loss prevention research on aerobics special shoe materials based on intelligent analysis," *Science Heritage Journal*, vol. 3, no. 1, pp. 20–23, 2019.
- [5] Y. Yu and A. Qi, "Teaching system of smart learning environment for aerobics course," *International Journal of Emerging Technologies in Learning (ijET)*, vol. 13, no. 05, pp. 165–176, 2018.
- [6] Q. Huang and F. Wang, "Design method of aerobics teaching assistant platform based on 5G technology," in *Proceedings of the International Conference on Application of Intelligent Systems in Multi-Modal Information Analytics*, pp. 75–81, Springer, Changzhou, China, March 2020.
- [7] L. Jia and L. Li, "Research on core strength training of aerobics based on artificial intelligence and sensor network," *EURASIP Journal on Wireless Communications and Networking*, vol. 1, pp. 1–16, 2020.
- [8] C. D. Wei, L. Chao, W. Shun, W. X. Ling, and W. W. Fang, "Research and application of multimedia digital platform in the teaching of college physical education course," *Journal of Intelligent and Fuzzy Systems*, vol. 34, no. 2, pp. 893–901, 2018.
- [9] Z. L. Kozina, T. V. Shepelenko, M. Cieslicka et al., "The teams' formation in sport aerobics on the basis of application of multidimensional analysis methods," *Physical education of students*, vol. 21, no. 6, pp. 270–279, 2017.
- [10] Y. Fan, J. Chen, S. Li, L. Zhao, Y. Zhang, and C. Liu, "A novel fuzzy synthetic discrimination system for teaching ability of physical education majors based on analytic hierarchy process," *Educational Sciences: Theory and Practice*, vol. 18, no. 5, 2018.
- [11] J. Zhao and G. Li, "Application of brain science in physical education and teaching," *Educational Sciences: Theory and Practice*, vol. 18, no. 5, 2018.
- [12] B. SueSee, K. Edwards, S. Pill, and T. Cuddihy, "Observed teaching styles of senior physical education teachers in Australia," *Curriculum Perspectives*, vol. 39, no. 1, pp. 47–57, 2019.
- [13] J. Chen and Y. Yang, "Research about bilingual education issues in business intelligence teaching," in *Proceedings of the 2018 International Conference on Education, Economics and Social Science (ICEESS 2018)*, October 2018.
- [14] B. SueSee, S. Pill, and K. Edwards, "Interrogating assumptions of a curriculum: queensland senior physical education syllabus," *The Physical Educator*, vol. 75, no. 5, pp. 850–879, 2018.
- [15] K. Karatrantou, V. Gerodimos, K. Häkkinen, and A. Zafeiridis, "Health-promoting effects of serial vs. integrated combined strength and aerobic training," *International Journal of Sports Medicine*, vol. 38, no. 1, pp. 55–64, 2017.

- [16] B. Yuhong, J. Radio, and T. Stations, "Application of virtual reality technology in broadcasting and hosting major practice teaching," *Electronic Measurement Technology*, vol. 7, 2017.
- [17] Z. Huysmans, D. Clement, M. A. Whitley, M. Gonzalez, and T. Sheehy, "Putting kids first: an exploration of the Teaching Personal and Social Responsibility model to youth development in Eswatini," *Journal of Sport for Development*, vol. 7, no. 13, pp. 15–32, 2019.
- [18] T. I. Mokra, A. Y. Osipov, M. D. Kudryavtsev, R. S. Nagovitsyn, and K. K. Markov, "Practice of Kangoo Jumps Fitness to improve female students' cardiorespiratory fitness," *Physical education of students*, vol. 23, no. 4, pp. 191–197, 2019.
- [19] Z. Chen, Y. Zhang, C. Wu, and B. Ran, "Understanding individualization driving states via latent Dirichlet allocation model," *IEEE Intelligent Transportation Systems Magazine*, vol. 11, no. 2, pp. 41–53, 2019.
- [20] Z. J. Hai, L. L. Yuan, and G. T. Xiao, "Improving memory and executive function of the elderly with mild cognitive impairment-A smart health care approach," *Journal of Medical Imaging and Health Informatics*, vol. 10, no. 1, pp. 44–48, 2020.
- [21] B. Guo, "Analysis on influencing factors of dance teaching effect in colleges based on data analysis and decision tree model," *International Journal of Emerging Technologies in Learning (ijET)*, vol. 15, no. 9, pp. 245–257, 2020.
- [22] D. D. Palma, A. Ascione, and S. Napolitano, "Education to school inclusion through sport," *Sport Science*, vol. 11, pp. 42–46, 2018.
- [23] I. S. Mbambo, "Teachers as change agents: primary teachers' perceptions and management of bullying in a selected school at kwamashu area," Doctoral dissertation, University of KwaZulu-Natal, Durban, South Africa, 2016.
- [24] Z. Xie, R. Lin, J. Wang, W. Hu, and L. Miao, "Vicarious learning: how entrepreneurs enhance a firm's international competitiveness through learning from interlocking director network partners," *Frontiers in Psychology*, vol. 11, 2020.