Hindawi
Mobile Information Systems
Volume 2022, Article ID 1940218, 10 pages
https://doi.org/10.1155/2022/1940218

Research Article

Stage Performance Characteristics of Minority Dance Based on Human Motion Recognition

Yuzhu Shi

Qilu Normal University, Jinan, 250200, China

Correspondence should be addressed to Yuzhu Shi; 1490140306@xs.hnit.edu.cn

Received 14 March 2022; Revised 30 March 2022; Accepted 1 April 2022; Published 29 June 2022

Academic Editor: Liping Zhang

Due to the different development history of different nationalities, there is a big gap between traditional cultures. Especially, as an important embodiment and component of traditional culture, ethnic minority dance has unique characteristics. With the development of machine vision technology, human motion recognition has gradually become a hot research direction, but its application research in the field of dance motion recognition is still in its infancy. In this paper, the characteristics of stage performances of ethnic minority dances in the field of intelligent auxiliary training based on human motion recognition technology are introduced. The human body regions in the frame are obtained by using human posture, and 3D_SIFT and optical flow features are extracted from each region. Then, we use the extracted key frames and DTW (dynamic time warping) algorithm to realize the motion recognition of the motion capture data and carry out simulation experiments. The test results show that the algorithm can identify the minority dance video database, effectively identify the dance movements, and realize the movement correction of dancers.

1. Introduction

Minority culture is an important part of Chinese traditional culture, and it adds to the vibrancy of the culture. Minority dance culture, in particular, is an important means of expressing ethnic minorities' social forms, thoughts, and feelings. Minority dance's distinct national style may also demonstrate that it has strong national characteristics. Ethnic style is also a distinguishing feature of ethnic dances. Because each country's dance is distinct, no matter what happens, the basic shape of the body will remain unchanged. This paper uses computer motion capture technology to carry out all-round three-dimensional digital protection of Korean minority dance postures based on dance characteristics, opening up a new world for the protection of intangible cultural heritage. The importance of preserving and revitalizing national culture cannot be overstated.

Human motion recognition technology is widely used in a variety of fields, but dance motion recognition research is still in its early stages [1]. It is a brand-new dance education supplement. Dance movements can be accurately recognized and irregular movements of dancers can be distinguished and corrected using human body motion recognition technology applied to dance image motion recognition [2]. To extract key frames, Wang et al. used K-means clustering and then interpolation. Candidate frames are defined as frames larger than this value, and key frame sequences are obtained by clustering unrepresented candidate frames [3]. The disadvantage is that there is an excessive amount of calculation. Liu et al. created a high-dimensional space, treated each frame's motion data as a point, turned the entire motion sequence into a high-dimensional space motion track, and used curves to screen out important points [4]. The disadvantage is that it does not accurately reflect real-world differences. Gurbuz et al. proposed a new method with good motion generalization ability, that is, a new method of extracting key frames by calculating the center distance between frames [5], but the accuracy of this algorithm will be harmed by the algorithm's convergence because the threshold is difficult to select. Sun et al. calculated the reconstructed frame using spherical linear interpolation rather than linear interpolation, calculated the ratio error, calculated and subtracted the frame reconstruction error, and obtained the reconstruction error curve using frame subtraction. This method cannot guarantee that the reconstruction

error of key frames obtained at the optimal compression rate is the smallest if the motion expression ability of key frames is not taken into account.

Motion recognition is one of the most widely used research directions in computer vision [6–8], but there has been little research on motion recognition combining video-based motion recognition technology with dance video, and more development based on dance video is needed. This paper summarizes the characteristics of various image behavior datasets based on previous studies on behavior recognition. Ethnic minority dance tasks can be extracted, represented, classified, and identified using this method.

The main innovations of this paper include the following aspects:

(1) For the minority dance video dataset, the research proposes an effective feature extraction method, which extracts the feature of direction histogram by dividing the image into equal parts, and finally creates the direction histogram set

(2) Select the dance motion capture data as the motion test sample, extract the key frames of the motion sequence by the method proposed in this paper, and identify the motion of the test sample by DTW (dynamic time warping).

The core contents of this paper are as follows:

The first section introduces the research background and significance before moving on to the paper's main work. The second section focuses on the technologies involved in human motion recognition. The research's specific methods and implementation are presented in the third section. The fourth section verifies the research model's superiority and feasibility. The fifth section is a synopsis and preview of the entire text.

2. Related Work

2.1. Research Status of Human Motion Recognition. Generally speaking, there are two kinds of input data for human motion recognition and behavior evaluation: video data and 3D bone data. Due to the convenience of obtaining image data, priority is given to research and application, and related work emerges one after another. However, moving image data is prone to occlusion, jitter, angle change, and other problems during shooting, and it is difficult to identify motion. And the advancement of optical inertial motion capture devices allows the 3D bone data of human motion to be captured directly. Specifically, 3D bone data is bone animation data that graphically describes the human body.

Ling et al. used a fixed sliding window and introduced the concept of a fixed time movement action point [9]. To achieve spatial motion alignment, Seulki et al. used the rotational offset between the bodies (extracted from the 3D position of the shoulder and torso) [10]. The Moravec angle operator, which can simultaneously detect the points where the moving object changes greatly in the local space-time dimension [11], has been used to model and refine the 3Dharris operator proposed by Zhang et al. Yan et al. divided the human body on the back into five body parts and used a contrast mining algorithm to detect various postures of the body parts in the spatial domain, which he then collected to create a data dictionary [12]. Al-Qaness et al. depicted motion as a continuous and differentiable function of changing body joint position with time and defined a window around the current time step in which the quadratic Taylor transform can be used to transform it locally [13]. Liu et al. proposed an automatic dynamic weighting method that assigns different importance to different bone joints based on the degree of sports participation, as well as a descriptor based on kinetic energy that provides an example of automatic segmentation and recognition of motion and determines similarity by referring to appropriate standard references [14]. Yin et al. used canny edge detection to extract information about human motion shape and then used edge matching to realize human motion recognition [15]. Millera et al. proposed using a dense optical flow function to describe video content and a motion boundary histogram to describe dense optical flow characteristics, inspired by the dense sampling method of image classification [16]. Tu et al. used atomic motion, object, and posture as explanation functions [17]. They call the co-occurrence relationship "work standard" because it models the co-occurrence statistics among these descriptive features. The weighted combination of these "task-based" subsets can then be expressed as tasks. However, this method is only useful for recognizing action in a few scenes. Because of the inaccurate or incomplete definition of attribute space in natural scenes, recognition accuracy is low [18].

2.2. Development Status of Auxiliary Dance Training. Movement training is an important link in the process of dance education and training. Beginners must go through long-term training and repeated practice before they can understand the essence of dance movements. Traditional mentoring training methods severely limit the effectiveness of trainers to a certain extent. Therefore, using immersive virtual reality technology to develop the virtual system of opera dance education has become a research topic in the field of opera art education.

Jha et al. created a Tai Ji Chuan practice VR interactive system that uses motion capture technology to capture, track, and supervise the dynamic trajectory of trainees, as well as evaluate, track, and imitate the behavior of coaches in real time [19]. At the time, Li et al. created a dance education and training system using advanced mixed reality (MR) and motion capture technology. To achieve [20], the system's functions are divided into teaching, practice, assessment, and other modules. Peng et al. used virtual reality (VR) technology to extract and design elements of local traditional dance content, created VR dance content for target users, and proposed using it in stage performances, as well as researching interactive narrative forms of stage performances [21]. Ji et al. created a virtual system for choreographing bells and dances, using music and dance elements as the basic design units [22]. The system creates a specific action tone, connects each characteristic action unit through data association, and produces the entire dance digitally.

This system is capable of innovating dance movement arrangements as well as dance rhythm innovation. Chaudhary et al. classified ethnic minority dances, used motion capture devices to collect dance movements, used digital 3D software to create ethnic dance role models and costumes, and finally used virtual interactive engine software to realize the dance interaction function [23]. Using virtual reality technology, Chenarlog and others created a virtual dance training system. A human limb sensation acquisition unit, an error correction unit, an image processing unit, and a processed human dance image display unit are the main functional modules of the system [24].

3. Methodology

3.1. Overview of Human Motion Recognition Methods. At present, the methods of motion recognition are mainly divided into single-layer method and layered method. Single-layer-based methods usually regard motion as a functional category in video and use classifiers to identify motion in video. The hierarchical method identifies high-level operations, mainly by identifying simple or low-level atomic operations in the video. A high-level complex task can be decomposed into a series of subtasks, and subtasks can be decomposed into higher-level tasks.

Motion recognition methods based on single layer can be roughly divided into two categories: spatiotemporal method and sequence model method. The main difference between spatiotemporal model and sequential model method lies in the way of dealing with the time dimension. The method based on spatiotemporal constructs video frames into three-dimensional spatiotemporal volumes on the time axis and extracts features from them, while the method based on sequence model regards human behavior as an ordered observation sequence in the time dimension. Sequence method is usually superior to spatiotemporal method in behavior recognition results, because they consider the sequence relationship of actions.

Hierarchy-based methods usually use a single hierarchy or lower-level subtasks to identify high-level complex tasks. A high-level complex task can be decomposed into a series of subtasks, which can be decomposed into higher-level tasks until they are decomposed into atomic tasks. The advantage of hierarchical behavior recognition method is that it can model the complex structure of human behavior. The hierarchical method is very flexible for simple behaviors and interactions between people. The hierarchical model provides an intuitive and simple interface, which combines prior knowledge and understanding of working structure.

There are two main ways to acquire human posture information, one is to accurately acquire information such as joint coordinates, human bones, motion trajectories, and so on through motion capture equipment, and the other is to acquire various joint positions and bones. This method of extracting specific features from local images for recognition is mainly used for video human motion recognition, and the recognition efficiency is improved by highlighting the motion state of specific joints or specific parts of joints showing specific types of motion.

After estimating the human posture of each frame in the video, the human skeleton information is obtained, the sequence $\{y_1, y_2, \cdots, y_m\}$ is used to represent an action, and each posture y is a set of high-dimensional vectors to form the body joint positions. The posture y_i probability vector in each sequence is expressed as

$$p^{(i)} = \left[p_1^i, p_2^i, \cdots, p_k^i\right]^T. \tag{1}$$

Gesture sequence is represented by matrix $p^{(i)} = [p^{(1)}, p^{(2)}, \cdots, p^{(m)}]$. In the process of recognition, in order to reduce the amount of computation, duplicate information is filtered and key gestures in all gesture sequences are mined.

In order to make the representative actions more robust in terms of inaccurate gestures, soft quantization is used to allocate each action symbol. More precisely, K symbols in the dictionary are used to represent a gesture, and each symbol is associated with probability P_i to measure the distance between gestures.

$$P_i = \frac{e^{-\mathrm{dist}(s_i, y)}}{\sum_{j=1}^{K} e^{-\mathrm{dist}\left(s_j, y\right)}}. \tag{2}$$

$\mathrm{dist}(s_i, y)$ represents the distance between the measuring point y and the convex hull formed by the active sample. More intuitively, a small distance produces a large probability, whereas a large distance will produce a small probability.

The application range of motion recognition algorithms is also different due to the different algorithm structure and feature description used, and there is no perfect universal algorithm that can be applied to all classification problems; so, the effectiveness of human motion recognition should be relative. It is very important to choose the appropriate algorithm according to the application range.

3.2. Dance Feature Extraction. Feature extraction refers to extracting feature information from the motion dataset to describe the target behavior in the image, which is an essential feature in motion recognition research. The extracted features appear to play a significant role in the accuracy and robustness of behavior recognition methods. The light direction histogram function is used to describe dance movement motion information. Furthermore, the study of dance movement recognition should take into account the impact of music on dance, as dancers dance to music, and the style of music is related to the type of dance.

Many frames of data are nonuniversal and uncertain when using motion capture devices to capture motion data information, and some frames of data may appear in the motion when incomplete or redundant. These frames are extremely useful and beneficial for motion detection. It also increases the amount of calculations required. As a result, motion capture data optimization and dimension reduction are beneficial in laying a better foundation for motion recognition, while motion capture data key frame extraction can achieve better data dimension reduction and feature extraction. This paper adjusts the extraction order of

each feature, improves the complementarity between features, and obtains feature descriptors with more accurate information based on the use of multiple features to form feature vectors. The optical flow features of each region are extracted, and then 3D SIFT features are extracted, with the constraints of optical flow parameters added to the SIFT features during the interest point selection process.

In dance movements, most movements require two arms and two legs to jointly determine a dance movement, but there is no fixed correspondence between arms and legs. In the process of extracting optical flow features, the background information of images with little change or little change according to the optical flow value can be filtered out; so, the areas where information is finally extracted are the upper body, the lower body, and the whole body. And the human body joint coordinates obtained from the motion capture device divide the human body region as shown in Figure 1.

In this paper, the displacement field formed by the superposition of optical flows between consecutive frame pairs is taken as the initial feature of optical flows. Optical flow uses the transformation method to calculate the displacement vector field d_t of the pixel (x, y) between two consecutive frames at time t and $t + 1$ and finds the function u, v that minimizes the energy function. Optical flow calculation consists of data items and smoothness items to optimize the global energy function. Mathematically, it is expressed as

$$E_{\text{Global}} = E_{\text{Data}} + \lambda E_{\text{Smooth}} \tag{3}$$

E_{Data} is a data item, which measures the consistency between the optical flow and the input image, E_{Smooth} is a smooth term, λ represents the flow field that tends to change smoothly, and E_{Global} means optimizing global energy.

Let the pixel point $p(x, y)$ have a gray value of $i(x, y, t)$ at time t. After time Δt, the point moves to $(x + \Delta x, y + \Delta y, t + \Delta t)$, and the gray value becomes $I(x + \Delta x, y + \Delta y, t + \Delta t)$. Since these two points are the same pixel point, the gray value remains unchanged, which can be expressed as follows:

$$I(x + \Delta x, y + \Delta y, t + \Delta t) = I(x, y, t). \tag{4}$$

Optical flow is sensitive to noise, scale change, and motion direction. The histogram of optical flow direction can not only represent motion information but also is insensitive to scale change and motion direction; so, it is used in many research methods of motion recognition. In this way, the optical flow histograms of all cells in the formed block are connected in series to form the optical flow histogram feature vector of the block.

We use $2L-$norm normal form to normalize the histogram. The specific form of $2L-$norm is shown in formula (5):

$$2L-\text{norm}, v \longleftarrow \frac{v}{\sqrt{\|v\|_2^2 + \varepsilon^2}}. \tag{5}$$

Finally, the feature vectors of the optical flow histogram of all blocks are concatenated to form the HOF feature of the image. You can use the following formula to calculate specific dimensions:

$$V = \text{binNum} \times \text{cellNum} \times \text{blockNum}, \tag{6}$$

where binNum represents the number of directional columns, cellNum represents the number of cell grids in each block, and blockNum represents the number of blocks in the image.

3D_SIFT is the three-dimensional information of SIFT features, which adds time information, and supplements the dynamic information of human body while maintaining static information. By encoding the local information of time and space of dynamic images, the robustness to direction and noise is improved. Therefore, the 3D_SIFT function is mainly used to select a point of interest from the image, then calculate the gradient direction and size of the whole neighborhood of the point of interest, and finally construct a 3D_SIFT descriptor by sub-histogram coding.

Scale space scales the original image according to certain rules and finally forms a pyramid-shaped multiscale spatial representation sequence. The spatial scale diagram is shown in Figure 2.

After the scale space is constructed, the local extremum points are selected as the key points. The key points are usually located at the edges and corners of the objects in the image, that is, the rapidly changing points. Direction determines the direction and gradient of each key point. The 2D gradient amplitude and direction of each pixel are specifically defined as

$$\begin{aligned} m_2 D(x, y) &= \sqrt{L_x^2 + L_y^2}, \\ \theta(x, y) &= \tan^{-1}\left(\frac{L_y}{L_x}\right), \end{aligned} \tag{7}$$

where x, y is the coordinate of the pixel in the image, and L_x, L_y is obtained by finite difference approximate calculation.

Then, use L_x, L_y and L_t to calculate the gradient size and direction of 3D:

$$\begin{aligned} m_3 D(x, y, t) &= \sqrt{L_x^2 + L_y^2 + L_t^2}, \\ \theta D(x, y, t) &= \tan^{-1}\frac{L_y}{L_x}, \\ \varphi(x, y, t) &= \tan^{-1}\frac{L_t}{\sqrt{L_x^2 + L_y^2}}. \end{aligned} \tag{8}$$

Because $\sqrt{L_x^2 + L_y^2}$ is positive, there is always $\varphi(-\pi/2, (\pi/2))$, and each corner is represented by a unique (θ, φ)

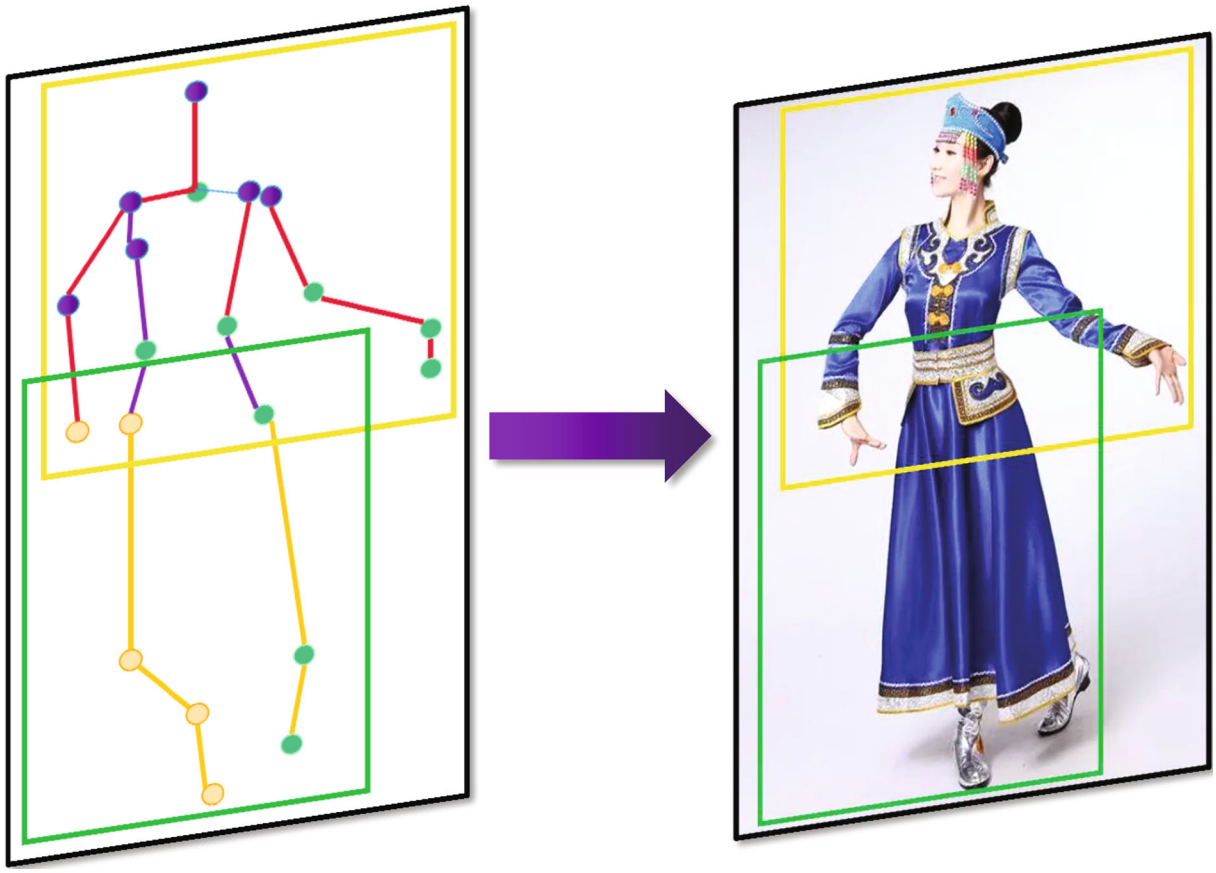

Figure 1: Dividing human body region map according to joints.

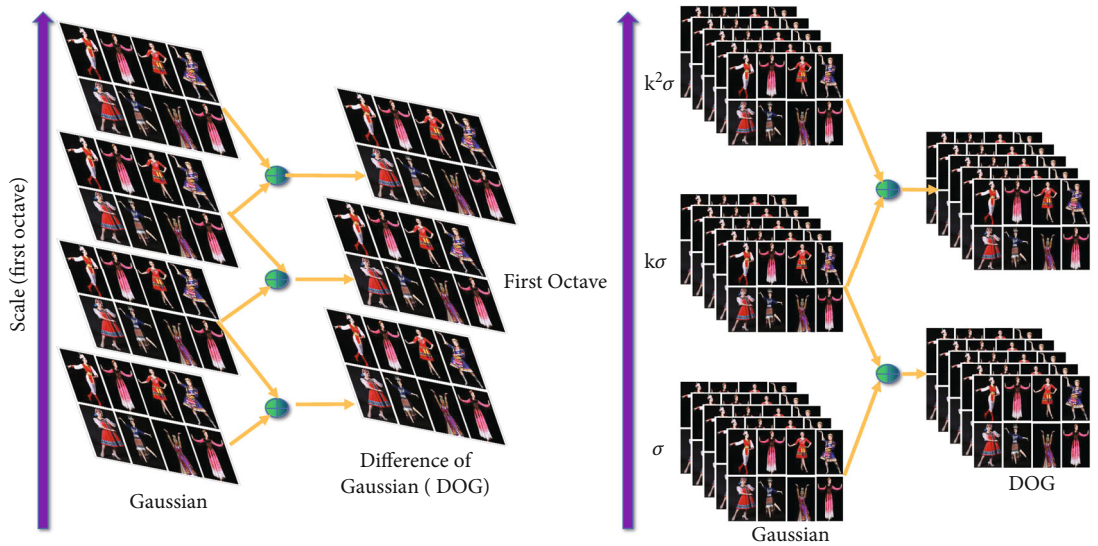

Figure 2: Spatial scale diagram.

pair; so, the gradient direction of each pixel in 3D is represented by two values.

3.3. Motion Recognition of Capture Motion Data. Regardless of the motion capture method used, the majority of the motion data are stored frame by frame in the form of a motion sequence in the process of human motion recognition. This motion sequence can be thought of as a time sequence, and the problem of identifying a motion sequence can be thought of as a problem of time sequence

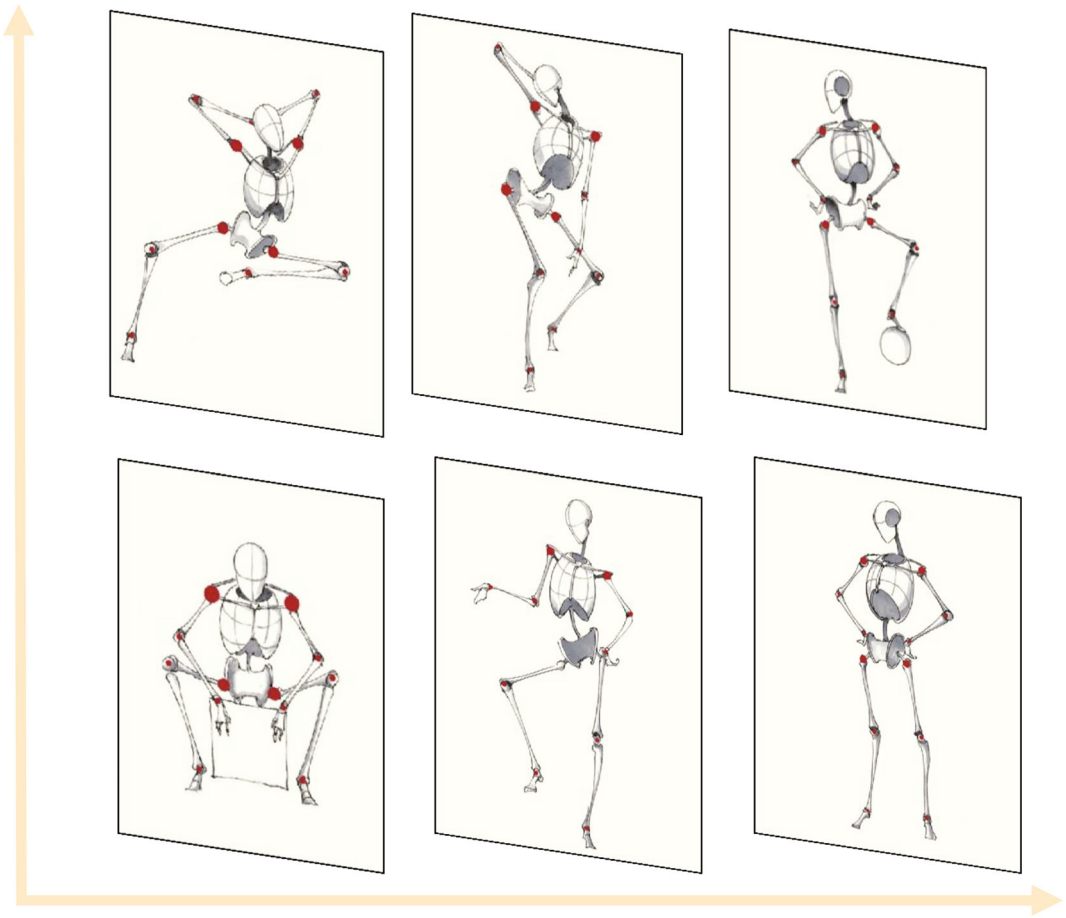

Figure 3: Three-dimensional example of human motion skeleton sequence.

Table 1: Comparison of recognition effects of features described.

Database		Upper body	Lower body	Whole body	
JHMDB	RGB	53.6	51.6	53.3	61.5
	Flow	60.1	60.8	56.8	69.4
	RGB + flow	66.2	61.7	66.9	74.1
MPII cooking	RGB	30.3	16.9	27.4	44.5
	Flow	48.9	0	57.1	58.6
	RGB + flow	50.4	0	55.4	92.7

identification. That is to say, two action sequences with similar action logic may or may not have similar captured values, making it difficult to compare and match human action sequences. In order to avoid errors caused by different numerical and logical similarity of behavior sequences, the DTW (dynamic time warping) algorithm is chosen as the logical similarity measurement method of human behavior sequence matching algorithm in this paper.

Dynamic time warping can be classified as an optimization problem. The one-to-one mapping relationship between two sequences is described according to the time warping function that meets the requirements, and then the time corresponding to the minimum cumulative distance between two sequences is found. A great advantage of regular function is that it can solve the time alignment problem of different action sequences. Get the coordinate information of human bones through Kinect, as shown in Figure 3.

Each action consists of a certain number of skeleton sequences. When building an action template, the coordinates of 20 skeleton points per frame must be stored in the template file. A template consists of the same actions performed by different people according to the same tasks.

In the process of setting templates, due to the differences in height and body shape of each person, it is necessary to

standardize the templates and make standard comparison, as shown in formula (9).

$$P_c(x,y) = \left(\frac{sl_x + sr_x}{2}, \frac{sl_y + sr_y}{2}\right), \tag{9}$$

where $P_c(x,y)$ is the standard center point of the obtained bone points, sl_x, sr_x is the x coordinate of the bone points on the left shoulder and the right shoulder, respectively, and sl_y, sr_y is the y coordinate.

Then, by taking the difference between the 20 bone points and the centroid, the human bone coordinate points after the centroid normalization are obtained. Because each person's shoulder width is different, and the shoulder width can form a relatively standard proportional relationship with the height, fat, and thinness of the human body, it can be calculated that the distance between the two shoulders of human skeleton coordinates can be standardized. Here, the distance between the two shoulders is obtained by formula (10).

$$D_s = \sqrt{(sl_x - sr_x)^2 + (sl_y - sr_y)^2}, D_s = \sqrt{(sl_x - sr_x)^2 + (sl_y - sr_y)^2} \tag{10}$$

where D_s is the European distance between two shoulders. The x, y of the 20 bone points with standardized center is divided by D_s, so that the coordinates of the bones to be matched with standardized scale can be obtained.

Select a time series as a reference, for each element in the time series, find another time series and its elements, then calculate the mean value, and finally get the mean value series. The average sequence length obtained by using the reference average value matches the reference sequence length.

The most direct idea of time series averaging is to find the corresponding relationship between each element by DTW algorithm and average each element to get the average sequence.

$\bar{A}$ is defined as the average sequence of multiple time series, and then

$$\bar{A} = \stackrel{\Delta}{=} \arg\min_{A \in X} d_D(A,S), \forall_X \in X^L, \forall_L \in [1,+\infty), \tag{11}$$

where X^L is an arbitrary time series of length L. And the average distance $d_D(A,S)$ is the average of DTW similar distances between sequence A and each time series in S.

$$d_D(A,S) = \frac{1}{N}\sum_{j=1}^{N} d_{\mathrm{DTW}}(A,S_j). \tag{12}$$

According to the above definition, an optimal time series that minimizes the sum of similarity distances can be obtained. However, you cannot subjectively limit the length of the average time series. The traditional method is to define the average order between each series in advance and then use iterative algorithm to average two time series at a time.

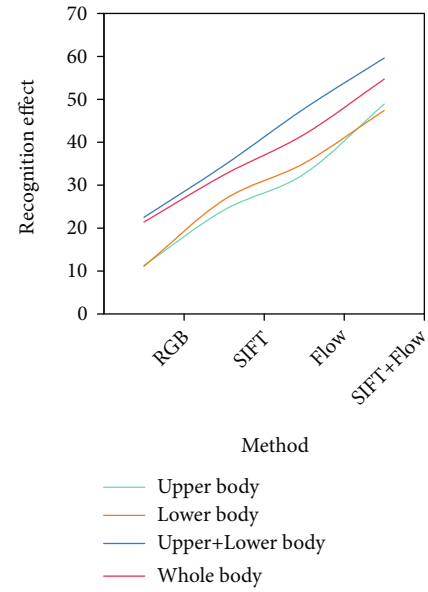

Figure 4: Different human body regions extract different feature recognition effects.

4. Experiment and Results

The databases used in this paper include minority dance video (MDV_data) collected from motion capture devices, and the popular JHMDB and MPII Cooking databases. In this paper, the image dance motion recognition process is to first determine the upper body, lower body, whole body image areas, and main motion sequences according to human posture and then extract 3D_SIFT features and optical flow. Finally, the function descriptor is used as the input of DTW to realize dance action recognition. The recognition effects of the described features are shown in Table 1.

In the experiment, the color features expressed by JHMDB and MPII datasets have obvious recognition effect, but in the process of dance movement recognition, complex background pixels affect the feature acquisition effect, thus reducing the recognition efficiency when 3D_SIFT is used. It means that the static information improves the performance and gets the effect of motion recognition in the above database. At the same time, the recognition of human body regions separated by human body posture will also lead to different recognition rates, and when using the whole body region, it is higher than other parts and has the highest proportion in the whole image.

Different feature recognition effects are extracted from different human body regions, as shown in Figure 4.

The recognition rate of motion after feature extraction using the upper or lower body is relatively close, but the recognition rate of the lower body is slightly higher, as can be seen from the recognition rates obtained by various methods. The rate at which a combination of human body parts is recognized. The reason for this is that the dance times of upper and lower body movements is similar when choosing a movement category, and the lower limbs are the primary movements. Furthermore, dance movements

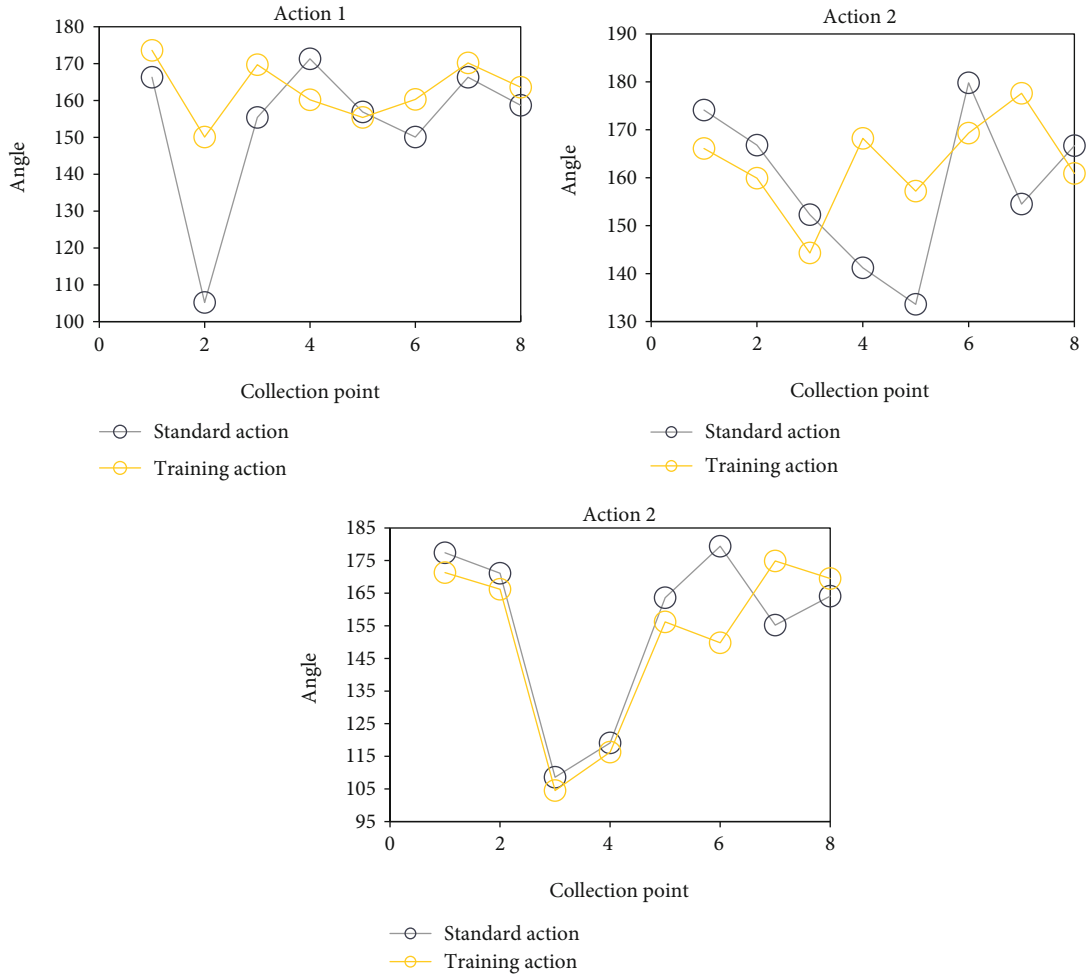

Figure 5: Comparison between each key joint angle and standard action curve.

Table 2: Comparison of experimental results of two HOG extraction methods on MDV_data dataset.

Method	Experimental result
This paper extracts HOG	32.3%
Traditional HOG extraction	22.5%

are frequently named separately from upper and lower limb movements, and separate training can improve recognition accuracy. Compute the angles formed by the joints in the decomposed movements using the movement information from the dance instructor's three hand movements, as shown in Figure 5. There are some differences between trainer movements and standard movements, as can be seen, and trainers of various levels can conduct targeted training according to their own level.

In Figure 5, it can be seen that the right wrist-right elbow-right shoulder angle is too large, the waist-right knee-right ankle angle is too large, the back leans forward slightly, and the right leg bends. At the same time, the angle between the neck-right shoulder-right elbow is too large, and the angle between the left wrist-left elbow-left shoulder is too small; so, the bending of the left elbow and left wrist should be reduced in the next training.

In this paper, the recognition results of HOG features extracted from the images generated by the cumulative edge feature algorithm are compared with those extracted from the original dance action images in two datasets. Table 2 compares the recognition results of HOG feature extraction used in this paper with that of existing MDV_data dataset.

In MDV_data dataset, the color of moving objects is similar to the background, and the results of this paper are 32.3% higher than the existing 22.5% HOG recognition results. By accumulating edge features from the generated image and extracting HOG features, it can be seen that the effect of the above situation is not as good as directly extracting HOG features from dance images.

Figure 6 shows the experimental results of the three functions and algorithms in this paper on four groups of dance combinations in JHMDB dataset.

It can be seen that the recognition rate of dance movements of individual features in each group is still relatively low. HOG function is used to express the local appearance and form of movements. If the similarity of movements in

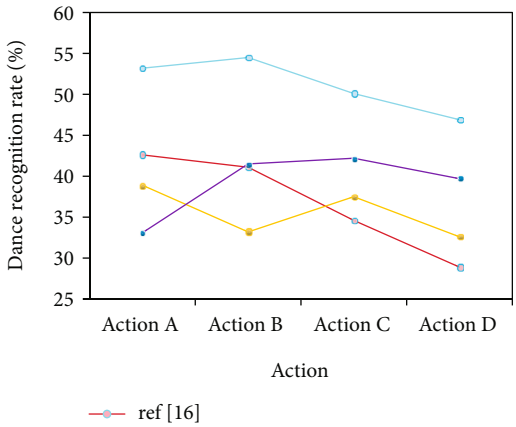

Figure 6: Single feature on JHMDB dataset and comparison of experimental results of this method.

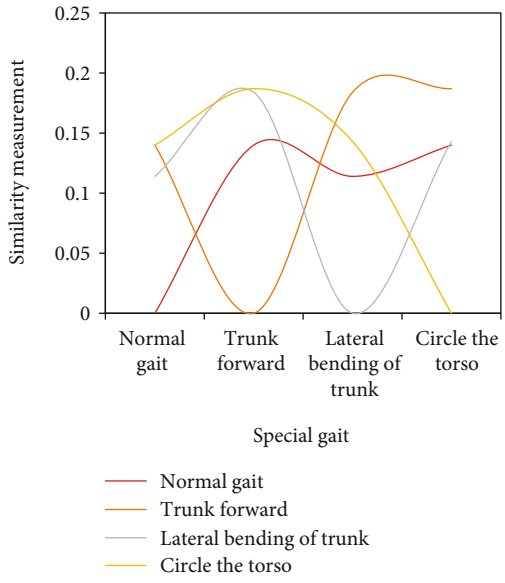

Figure 7: Measurement results of similarity of special gait of the human body.

dance combination is too high, the recognition difficulty will increase, and the recognition accuracy will be affected.

Among the four dance combinations in JHMDB dataset, the similarity of dance actions of action A and action B is much smaller than that of action C and action D. Action C and action D have similar dance steps, especially action D. Many similar movements and the same movements are divided into different directions, which also increase the difficulty of identifying dance movements.

In human daily life, gait is a regular and periodic movement, which achieves normal movement through the cooperation of nervous system, bones, and muscles. Generally speaking, normal gait refers to the gait of healthy adults walking in the most natural and comfortable posture, which is characterized by stable body, sufficient stride, and minimum energy consumption. Relevant indicators and recognition results are shown in Figure 7.

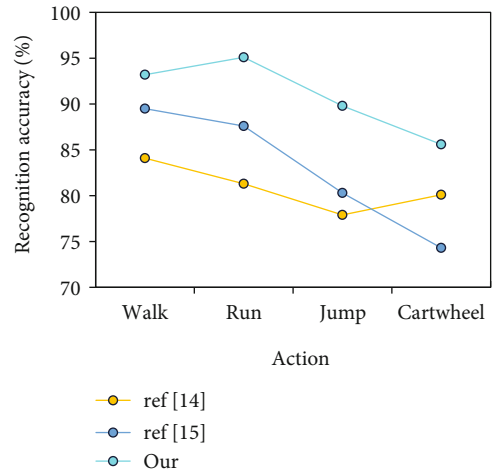

Figure 8: Identification accuracy of three methods in four action tests.

Through template similarity matching, the similarity between various gait movements is calculated, and the posture of the trunk during walking is classified in order to help doctors observe and diagnose.

The method proposed in this paper is used to extract the key frames of the motion sequence, and the DTW algorithm is used to identify the motion of the test samples. The accuracy of action recognition is introduced as the evaluation standard of action recognition, four basic actions such as walking, running, jumping, and cartwheel are recognized, respectively, and the two methods in references [14, 15] are selected for comparison. In this paper, the accuracy of recognition is used as the standard to measure the performance of the algorithm. Figure 8 shows the recognition accuracy of three algorithms of four basic motion recognitions.

As shown in Figure 8, this method can effectively identify four basic actions, and the recognition accuracy is higher than the other two methods.

5. Conclusions

To sum up, as an important part of Chinese traditional culture, the development of minority dance is often influenced by social culture, environment, customs and religious beliefs, etc. Dance can also fully reflect the cultural background of a country. A method of analyzing the performance characteristics of minority dance performances based on human motion recognition is proposed. 3D_SIFT, which adds time in three dimensions by fusing dynamic features based on posture, uses optical flow features to express changes in time and space according to actors' actions. The changing trend of main joints in continuous frames of human posture can help to identify the same movement under different orientations and body postures, thus determining the upper and lower body regions of the human body in the image. The DTW algorithm is used to realize motion recognition, and the algorithm is verified by simulation experiments. At the same time, this paper only uses basic minority dance training

movements as experimental movements, analyzes the stage performance characteristics of modern dance, classical dance and other dance styles, and tells the future research focus.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The author does not have any possible conflicts of interest.

References

[1] P. Wang, H. Liu, L. Wang, and R. X. Gao, "Deep learning-based human motion recognition for predictive context-aware human-robot collaboration," *CIRP Annals - Manufacturing Technology*, vol. 67, no. 1, pp. 17–20, 2018.

[2] J. Yu, J. Sun, S. Liu, and S. Luo, "Multi-activity 3D human motion recognition and tracking in composite motion model with synthesized transition bridges," *Multimedia Tools and Applications*, vol. 77, no. 10, pp. 12023–12055, 2018.

[3] Y. Wang, W. Wang, S. Tian, M. Li, P. Li, and X. Chen, "Human motion recognition based on electrostatic signals," *Jiqiren/Robot*, vol. 40, no. 4, pp. 423–430, 2018.

[4] W. Liu and S. Li, "Human motion target recognition using convolutional neural network and global constraint block matching," *IEEE Access*, vol. 8, no. 99, pp. 69378–69388, 2020.

[5] S. Z. Gurbuz and M. G. Amin, "Radar-based human-motion recognition with deep learning: promising applications for indoor monitoring," *IEEE Signal Processing Magazine*, vol. 36, no. 4, pp. 16–28, 2019.

[6] W. Cai, B. Zhai, Y. Liu, R. Liu, and X. Ning, "Quadratic polynomial guided fuzzy C-means and dual attention mechanism for medical image segmentation," *Displays*, vol. 70, p. 102106, 2021.

[7] J. Zhou, X. Wei, J. Shi, W. Chu, and W. Zhang, "Underwater image enhancement method with light scattering characteristics," *Computers and Electrical Engineering*, vol. 100, p. 107898, 2022.

[8] W. Cai, M. Gao, R. Liu, and J. Mao, "MIFAD-net: multi-layer interactive feature fusion network with angular distance loss for face emotion recognition," *Frontiers in Psychology*, vol. 12, 2021.

[9] K. Ling, H. Dai, Y. Liu, A. X. Liu, W. Wang, and Q. Gu, "UltraGesture: fine-grained gesture sensing and recognition," *IEEE Transactions on Mobile Computing*, vol. 22, no. 99, p. 1, 2020.

[10] S. Kyeong, W. Shin, M. Yang, U. Heo, J. R. Feng, and J. Kim, "Recognition of walking environments and gait period by surface electromyography," *Frontiers of Information Technology & Electronic Engineering*, vol. 20, no. 3, pp. 342–352, 2019.

[11] Z. Zhang, Y. A. Yingchun, W. U. Zhaohui, and M. A. Qian, "A posture recognition system for rat cyborg automated navigation," *Chinese Journal of Electronics*, vol. 27, no. 4, pp. 687–693, 2018.

[12] H. Yan, Y. Zhang, Y. Wang, and K. Xu, "WiAct: a passive WiFi-based human activity recognition system," *IEEE Sensors Journal*, vol. 20, no. 1, pp. 296–305, 2020.

[13] M. A. Al-Qaness, M. Abd Elaziz, S. Kim et al., "Channel state information from pure communication to sense and track human motion: a survey," *Sensors*, vol. 19, no. 15, p. 3329, 2019.

[14] X. Liu and G. Zhao, "3D skeletal gesture recognition via discriminative coding on time-warping invariant Riemannian trajectories," *IEEE Transactions on Multimedia*, vol. 23, no. 99, pp. 1841–1854, 2021.

[15] Y. Zheng, R. Yin, Y. Zhao et al., "Conductive MXene/cotton fabric based pressure sensor with both high sensitivity and wide sensing range for human motion detection and E-skin," *Chemical Engineering Journal*, vol. 420, article 127720, no. 22, 2021.

[16] L. Millera, H. C. Agnewb, and K. S. Pilzb, "Behavioural evidence for distinct mechanisms related to global and biological motion perception," *Vision Research*, vol. 142, no. 142, pp. 58–64, 2018.

[17] Z. Tu, H. Li, D. Zhang, J. Dauwels, B. Li, and J. Yuan, "Action-Stage Emphasized Spatiotemporal VLAD for Video Action Recognition," *IEEE Transactions on Image Processing*, vol. 28, no. 6, pp. 2799–2812, 2019.

[18] T. Plotz and Y. Guan, "Deep learning for human activity recognition in mobile computing," *Computer*, vol. 51, no. 5, pp. 50–59, 2018.

[19] D. Hachuel, A. Jha, K. Staller, C. D. Velez, and A. Martinez, "Mo2049 - Augmenting Gastrointestinal Health: A Deep Learning Approach to Human Stool Recognition and Characterization in Macroscopic Images," *Gastroenterology*, vol. 156, no. 6, p. 937, 2019.

[20] R. Li, Z. Liu, and J. Tan, "Human motion segmentation using collaborative representations of 3D skeletal sequences," *IET Computer Vision*, vol. 12, no. 4, pp. 434–442, 2018.

[21] Y. Peng, H. Lee, T. Shu, and H. Lu, "Exploring biological motion perception in two-stream convolutional neural networks," *Vision Research*, vol. 178, pp. 28–40, 2021.

[22] X. Ji, Q. Zhao, J. Cheng, and C. Ma, "Exploiting spatio-temporal representation for 3D human action recognition from depth map sequences," *Knowledge-Based Systems*, vol. 227, article 107040, no. 4, 2021.

[23] S. Chaudhary and S. Murala, "Depth-based end-to-end deep network for human action recognition," *IET Computer Vision*, vol. 13, no. 1, pp. 15–22, 2019.

[24] V. A. ChenarlogH and F. Razzazi, "Multi-stream 3D CNN structure for human action recognition trained by limited data," *IET Computer Vision*, vol. 13, no. 3, pp. 338–344, 2019.