Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 3212308, 11 pages
https://doi.org/10.1155/2022/3212308

Research Article

A Three-Dimensional Animation Character Dance Movement Model Based on the Edge Distance Random Matrix

Yan Jin

Shanghai Normal University, Shanghai 200233, China

Correspondence should be addressed to Yan Jin; jinyan@shnu.edu.cn

Received 11 April 2022; Revised 12 May 2022; Accepted 21 May 2022; Published 31 May 2022

Academic Editor: Ning Cao

In this paper, we use the edge distance random matrix method to analyze and study the dance movements of 3D animation characters and design a 3D animation character dance movement model. Firstly, each dance movement video in the dataset is divided into equal parts, while the segmented videos are subjected to the operation of accumulating edge features separately, and the edge features in all video images within each segment are accumulated into one image, and the directional gradient histogram features are extracted from them. Finally, a set of directional gradient histogram features are used to represent the local appearance and shape of the video dance movement. The reconstructed human movements are obtained mainly by fitting the 3D human coordinates in the image space using the human model and the estimated depth coordinates, which are currently more mature and are chosen from existing techniques. In other combinations, the performance of the method in this paper is better than the recognition results of the benchmark method, especially when the similarities of dance movements in the towel flower combination and the piece flower combination are too high. In response to the problem of the reconstructed action appearing to have ground penetration, slipping, and floating feet, a method is proposed to optimize the fitting human model foot problem according to the foot touching the ground. The experimental results show that the method can replace the traditional motion capture method to a certain extent, simplify the use of motion capture, and reduce the cost of motion capture. In the model deformation stage, to reduce the deformation quality problem during the model motion and improve the efficiency of weight calculation, a model deformation method combining double quaternions and bounded double tuning and weights are given. The static 3D model is tetrahedralization, and then the skeletal control points of the model are set and the weight of each skeletal segment to the model is calculated; next, the 3D model is mixed and bound with the skeletal data using the dual quaternion skinning algorithm; finally, the static 3D model motion is driven by the skeletal data. Experiments demonstrate that the method results in better deformation of the 3D model during rotation.

1. Introduction

In recent years, with the booming development of the field of computer graphics and the further development of computer-related electronic components, the application of 3D animation technology in various fields such as augmented reality, virtual reality, and film animation special effects production is in full swing. In essence, computer 3D animation is a series of continuous images displayed by a computer at a certain rate, giving people a sense of animation [1]. There are multiple similar actions in an action, especially a montage combination, and the same action is divided into different directions. As the three-dimensional human body is complex in its structure and in reproducing the entire process of movement, they have become an important technical difficulty in the production of three-dimensional human animation. However, because the body's gestures can be used to express a relatively rich connotation, the study and production of three-dimensional human animation have great practical significance. The existing 3D human animation methods are very demanding in terms of both the quality of the model itself and the accuracy of the skeletal skin and often require a great deal of human and material effort in the adjustment of the model during the

animation process. With the development of science and technology and the innovation of concepts, the variety of animated short films has gradually increased in terms of form and content, and the number of genres has grown [1].

Physics-based realistic character animation is the process of modeling the physics of a moving human body, giving it properties that it would have in the real world, such as mass, elasticity, and friction. Given a character's state of motion, these properties can be used to calculate the character's state of motion at the next time step to obtain the character's complete motion. The designer needs to optimize the motion model by adjusting the parameters of the physics algorithm to achieve a more realistic result [2]. There are similar dances in the handkerchief flower combination and the piece flower combination. The physics model derived in this way is often referred to as a controller, and each controller corresponds to a specific movement behavior of the character; for example, given an initial state of walking, the walking controller calculates and obtains a model of the character's movement in the walking state [3]. This approach is more realistic than keyframe character animation, but as each controller can only control one specific type of movement for a particular character, it is not universally applicable and may produce incorrect movement data due to state changes during the character's movement. This method is therefore overly dependent on the physical model and is limited and the types of motion that can be achieved are relatively simple [4].

This technique has improved the efficiency of the fields of multimedia and animation technology when setting up the initial motion of a model. The 3D human motion capture technology first requires the acquisition of human motion skeletal data and static model data and then uses the Kinect device to acquire real-time skeletal data and depth images of keyframes while the human body is in motion, combining the static model, skeletal data, depth images, and relevant human deformation algorithms to capture human motion in real time and reconstruct the 3D human dynamic model in motion [5]. Among the four dance combinations in the FolkDance dataset, the similarity between the dance movements in the double flower combination and the inner flower combination is much smaller than that of the handkerchief flower combination and the piece flower combination. With the completion of human motion capture and motion feature extraction, combined with the motion data in the human motion model library, the type of motion performed by the tested person is identified in real time. The 3D human motion capture system is divided into two main parts: the acquisition of human motion data and the recognition of human 3D movements. In terms of accuracy and cost, the Kinect device from Microsoft was chosen for the acquisition of human skeleton data and depth images of keyframes. In terms of motion deformation, the Kinect scanned human model is skeleton-bound to achieve 3D human dynamic model deformation during motion. For motion recognition, an improved dynamic time-adjustment algorithm is used for human motion recognition.

2. Related Work

A channeled attention unit (CAU) module was designed by using STCAN constructed on a dual-stream network, which can effectively extract spatiotemporal information [6]. By using the CAU module, the interdependencies between channels can be modeled and further weight distributions can be generated to selectively enhance the information function. Yang et al. proposed a new action recognition for the characteristics of the independent univariate GM model, which ignores the rotation of the camera, which rarely occurs in action recognition videos and uses the GM to represent the x and y directions, respectively; furthermore, the GM is position-invariant as it is derived from the generic camera motion. Pixels with global motion are subject to the same parametric model and pixels with mixed motion can be considered as outliers, based on which the authors propose an iterative optimized GM estimation scheme that progressively removes outliers and estimates global motion in a course to the fine manner and, finally, LM using a spatiotemporal thresholding-based approach [7]. Through extended contour convolution, SCN can learn from low-level geometric shape boundaries and their temporal dynamics to jointly learn cooccurring features and construct a unified convolutional embedding space, effectively integrating spatial and temporal properties. Geometry-based SCNs significantly improve the recognition of features learned from motion [8].

The skeletal point information and a 3D point cloud of the human body are obtained by Kinect, and then the template human body is rigidly deformed by segmentation through the skeletal information until it reaches the destination coordinates, and then the deformed model is compared with the point cloud information by ICP algorithm for better alignment, and, finally, the model is flexibly smoothed by TPS algorithm [9]. A 3D segmented human modeling technique based on a single Kinect device is proposed, in which the human body is roughly divided into two parts, further subdivided, and modeled separately, and, finally, all subdivided models are combined [10]. Existing training methods follow a training strategy of starting with a certain value of the learning rate, empirically, and then reducing the learning rate every few training epochs after a certain period. This method allows for the creation of realistic and accurate 3D human models and the rapid construction of many models, which is promising for large-scale crowd 3D animation [11]. An effective feature extraction method is proposed for the dance video dataset, in which the video is equally segmented, followed by an edge feature accumulation operation for each segment, in which all the edge features in each segment are accumulated into one image, and directional gradient histogram features are extracted, and, finally, a set of directional gradient histogram feature vectors are used to characterize the shape of the dance movements of the video [12].

The paper proposes a dance movement recognition method based on the fusion of directional gradient histogram features, optical flow directional histogram features, and audio signature features, and, for the fusion of

heterogeneous features, the paper chooses to use a multicore learning approach to organically fuse the three types of features for dance movement recognition research. Considering that there are few publicly available dances movement datasets, we use a professional motion capture device, Vicon, and invite professional dancers to perform dances according to a design scheme to collect the dance datasets. The process, scheme, and specific content of the dance datasets used in this paper are described in detail, as well as the experimental design, experimental environment, and evaluation criteria. The experimental results are analyzed and compared to verify the effectiveness of the algorithm.

3. Edge Distance Random Matrix Algorithm

This paper introduces the clustering property of community structure; that is, for any social network graph, the community structure is more closely connected inside and more sparsely connected outside, and when a "wanderer" wanders randomly in the network, the chance of returning to the inside of the community will be greater than the chance of returning to the outside of the community. In addition, the sequence of random walks is a Markov chain; that is, the next step that each node should take is not related to the previous step of the node at all but only to the current node's situation; that is, each node may belong to each discovered community, thus enabling overlapping community discovery in random walks to discover community structures. Based on the random wandering process and the edge random wandering process described above, we first calculate the random wandering similarity metric between nodes [13]. Not only is the optical flow direction histogram capable of representing action information but also it is insensitive to scale changes and motion directions, so this feature is used in many research methods of action recognition. If a "wanderer" starts at node in the network graph, the probability of wandering to any node connected to it is calculated as follows:

$$P_{i,j} = \begin{cases} \dfrac{w_{i,j}}{w_i}, \text{others} \\ \\ 1, i, j \in V \end{cases}. \tag{1}$$

In social networks, it is also easy to see that there is a high degree of transferability between friends. In other words, someone's best friend may also be his or her best friend, people with similar interests are usually close friends with each other, or most of one's friends are acquainted with each other. This common type of friend relationship in social networks can be thought of as a ternary friend relationship. This ternary friend relationship is in line with the clustering characteristic of social networks, where communities are more closely connected. In this paper, clustering coefficients are used to quantify the degree of community aggregation in a social network.

The clustering coefficient of a point is a measure of the degree of aggregation in a network from the nodes in the network. In general, the formula for calculating the clustering coefficient for a particular node v in the network is as follows:

$$C(v) = \frac{E\left(e_{i,j}\right)}{T\left(k_i + 1, k_j + 1\right)}. \tag{2}$$

By connecting three nodes, a closed triangle can also be formed. If one edge of the triangle is in a community, likely, the remaining edges are also in that community. However, in actual social networks, the situation of edges in each community that is divided is different. Therefore, when calculating the clustering coefficient of an edge in a network, it is often necessary to add 1 to the numerator to calculate the proportion of the number of triangles containing that edge to the number of all triangles that could contain that edge; for example, the formula for edge i, $e_{i,j}$ in a network is as follows:

$$C(v) = \frac{N\left(e_{i,j}\right) + 1}{\max\left(k_i + 1, k_j - 1\right)}. \tag{3}$$

It tends to have a very strict data treatment, considering that each piece of data can belong to just one category. However, in a real social network, each piece of data may often belong to more than one category, and sometimes even the categories assigned are very ambiguous. Therefore, when classifying data with ambiguity, it is important to consider both the relationship between the data and the strength of the relationship between the data.

$$J_m(U, C) = \sum_{i=1}^{n} \sum_{j=1}^{k} u_{i,j}^{m} \left\| x_i^2 + c_j^2 \right\|^2. \tag{4}$$

Just as for the same sentence, different people speak at different speeds, in action recognition, and although the trajectories of the actions performed by the human body are the same, different people, even for the same action, move at different speeds. Therefore, for two movement sequences with different movement speeds, they should be subjected to movement planning and thus find the most matching movement trajectory for the two movements.

Furthermore, it is possible that the time sequences of the two different movements are only shifted on the time axis, that is, that the time sequences of the two movements being compared are, in the case of reduced shifts, identical. In the above-mentioned cases, the similarity between the two time series cannot be effectively compared simply by calculating the Euclidean distance between the two time series. Instead, the dynamic time regularization algorithm is used when the length of one of the time series is regularized to a certain extent by lengthening the series or shortening it when the length of the test time series differs from that of the two time series in the template, as shown in Figure 1. We cannot rely on this graph file to output the result of action recognition, because the data we need for action classification does not have the connection with the nodes on the original graph data. The distance between the two time series being compared is then calculated, and the similarity between the two is obtained.

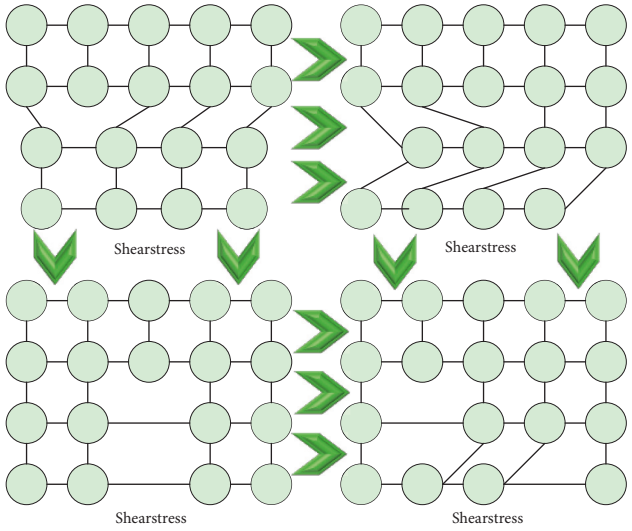

Figure 1: Edge distance random matrix framework.

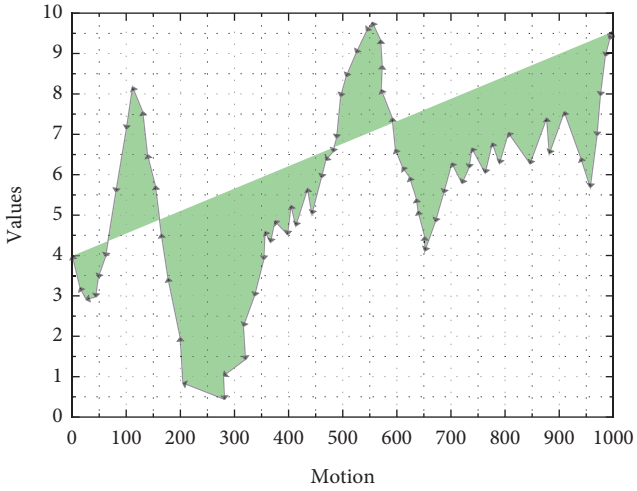

Figure 2: Motion velocity frame diagram.

The main idea of the DTW algorithm is to find a matching path with the minimum distance for two time series of different lengths using a dynamic programming method, which is a point-to-point mapping relationship between the two series. If the two sequences are of different lengths, the DTW algorithm can make this difference in the time axis disappear, so that the distortion between the two can reach a minimum value. The way to eliminate the difference in the time axis is to stretch or shorten the time axis of one of the sequences and then make it reach a maximum overlap with the time axis of the other sequence.

$$\left|X_t - \mu_{i,t}\right| \geq 2.5\sigma_{i,t}^2. \tag{5}$$

In action recognition, the time series is the representation of the data, and comparing the similarity of two actions is, in time series terms, comparing the similarity of two sequences. In a time series, the two time series that need to be compared for similarity may not be of equal length, which in the field of action recognition manifests itself in the different speeds of actions of different people. Because of the considerable randomness of the action signal, even actions performed by the same person at different moments in time may not have the same length of time. In this case, the traditional Euclidean distance cannot effectively find the distance or similarity between two time sequences [14]. That is, in most cases, the two sequences have very similar shapes, but these shapes are not aligned in the time series. In this case, if you want to compare the similarity of the two time series, the easiest way to align them is to do a linear scaling, as shown in Figure 2. The short sequence is scaled linearly to the same length as the long sequence and then compared for similarity. However, such a calculation does not consider the fact that the duration of the segments in the motion clip may vary in length from one situation to another, so recognition is unlikely to be optimal, and therefore dynamic time regularization is more often used.

When the training data is linearly divisible, the hard interval needs to be maximized. The hard interval is relative to the training dataset or to the feature space, meaning that no classification errors can occur, and the model learned from the hard interval can be referred to as a hard interval support vector machine. When the training data is approximately linearly divisible, it is necessary to introduce slack variables to maximize the soft interval, which is relative to the hard interval and allows for a certain amount of error to exist. For the error measure, a parameter needs to be defined which measures how much error the data made rather than how many errors the data made. The human body is divided into posture parameters and shape parameters, where the posture parameter is the rotation angle of the joint, and the shape parameter is the height, fatness, thinness, and so forth, which makes it very convenient to reconstruct a body close to a real person from information such as 3D human key points. The model learned from the soft interval can be called a soft interval support vector machine.

In the skinning algorithm of the model, the calculation of the transformation matrix is the calculation of the interpolation [15]. During the movement of a model of the human body, the interpolation of the bones where the movement occurs should be calculated based on the time difference between the two keyframes of the movement of the bones. Quaternions, double quaternions, and matrices can all represent rotations and can be transformed into each other. When the model is interpolated with quaternions, the interpolated quaternions are first converted into a transformation matrix. The absolute transformation matrix of each bone is then calculated and the new position of each vertex in the model is calculated.

4. 3D Animation Character Dance Movement Model Design

The convolution principle of this module is divided into three steps: regular convolution, Ghost generation, and

feature map concatenation. Firstly, the standard convolution is used to obtain the intrinsic feature map, and then each channel layer of the intrinsic feature map is linearly transformed to produce a Ghost feature map with the same number of channels as the intrinsic feature map, and, finally, the Ghost feature map and the intrinsic feature map are stitched together [16]. The model learned by soft margin can be called a soft margin support vector machine. The linear transformation used in the experiments is achieved by deep convolution. The basic principle of the Ghost Module is like that of deep separable convolution, but the difference lies in the fact that point-by-point convolution is performed first, followed by deep convolution, and the number of deep convolutions is increased. Compared to the standard convolution operation, the computational effort of the Ghost Module is significantly reduced. On the other hand, how this convolution method is computed is understood as an augmentation of the feature map.

The human contour is mainly used in early methods of human posture estimation, where the model contains the approximate width and contour of the main joints and torso, and the body parts are represented by rectangular boxes or contour edges. The model is trained from 3D scans of thousands of people and obtains statistics on human pose and shape. The greatest advantage of the model is its compact and intuitive parametric treatment of the human form into pose and shape parameters, where the pose parameter is the angle of rotation of the joints and the shape parameter is an indication of body shape such as height and fatness. This makes it easy to reconstruct a near-real-life form from information such as 3D human key points.

The introduction of parameters related to the virtual human muscle fibers, combined with the multilayered model structure and biological anatomical fiber vectors, better simulates the effect of anatomical fiber vectors with local directional stretching characteristics on the secondary movements of the virtual human model, achieving a realistic animation effect on movements such as biceps flexion, as shown in Figure 3 to demonstrate whether the effect of biceps contraction based on anatomical fiber vectors under simulation is compared as shown in Figure 3.

In a file of type x, a sparse matrix is stored, with each row of the matrix representing a node number and each column representing a different feature. Take, for example, the Cora dataset originally used by FastGCN. The Cora dataset has been a popular dataset for graph deep learning in recent years and consists of machine learning papers. After removing words that have no real meaning and words that occur too infrequently, 1433 unique words are left. A 1433-dimensional feature was constructed in the original x-class file based on the 1433 unique words contained in the Cora dataset. 0 and 1 describe whether each word is present in the paper.

For error measurement, a parameter needs to be defined. This parameter measures how much error the data has made rather than how many errors the data has made. The start and endpoints of the same action for different subjects were then interconnected, thus completing the construction of a meaningful graph data file. Although this graph data file can be used for FastGCN training, due to the nature of the FastGCN network, it cannot be relied upon to output action recognition results, as the data we need to classify for action does not have a connection to the nodes on the original graph data.

Motion features in a video describing human action are often essential features in action-based recognition studies, and the optical flow method is often used to extract these features in some of the current action recognition studies. The concept of optical flow has been introduced a long time ago and can be thought of as the instantaneous velocity generated when pixels on the moving surface of a target change in a video, as shown in Figure 4.

Its main principle is to determine the change in pixel position based on the change in greyscale of all pixels in the video image and the correlation between pixels, so optical flow is generally generated by the motion of the target in the video, the motion of the camera, or the motion of both together. In this paper, we use an optical flow-based approach to extract motion features from dance videos to characterize the motion information of dance movements [17]. That is, most of the time, the two sequences have very similar shapes, but the shapes are not aligned over the time series. The advantage of using the optical flow direction histogram is that as optical flow is sensitive to noise, scale variation, and direction of motion, the optical flow direction histogram is not only able to represent motion information but also insensitive to scale variation and direction of motion and therefore has been used in many research methods for motion recognition.

Therefore, the core problem of multicore learning is to learn both the optimal kernel function parameters and the corresponding weights of the kernel function in the process of optimization. Methods based on multicore learning usually obtain better results than methods based on single kernel learning, but the problem facing multicore learning is that its time complexity and spatial complexity are too large.

5. Analysis of Validation Experiments

However, existing methods often use a lower resolution due to concerns about computational overload. Therefore, a layer of upsampling was removed and the pooling layer was removed to ensure a high-resolution feature representation while keeping the network lightweight [18]. Based on this design, the size of the feature maps in the latter two stages is kept consistent with that of the backbone network, and there are only two upsampling layers.

In the inference phase, most of the existing methods for obtaining the coordinates of the heat map and converting them to their original resolution use the Argmax function, but Argmax results in discrete integers only, which inherently imposes a limit on the accuracy of the coordinates. However, it is observed that the true value of the heat map of the training data is obtained at each coordinate point using 3D Gaussian centrality, as shown in Figure 5.

The traditional Euclidean distance cannot effectively find the distance or similarity between two time series. This problem can be avoided by the iterative training method,

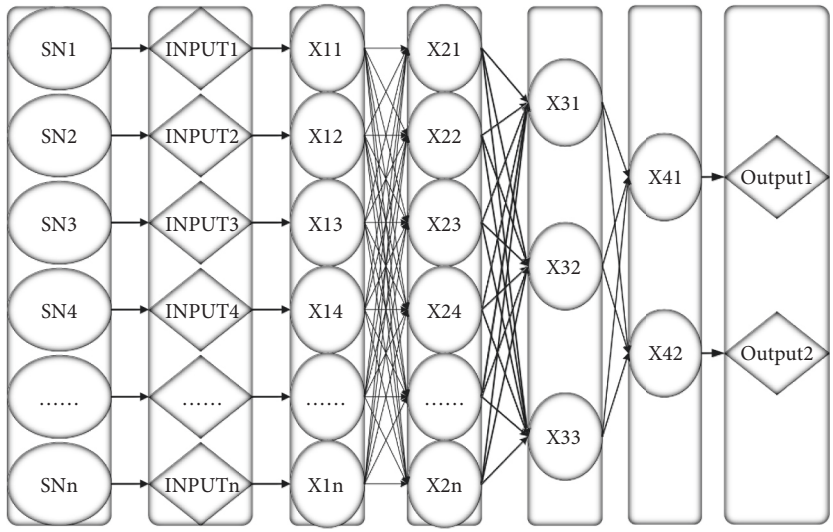

Figure 3: Sample propagation diagram.

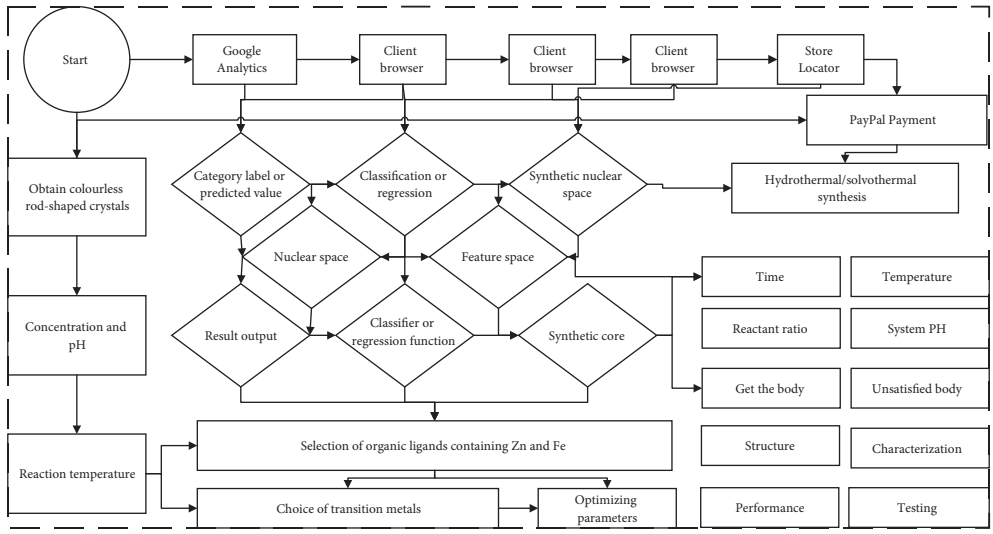

Figure 4: Schematic diagram of a linear combination of multi-core learning kernel functions.

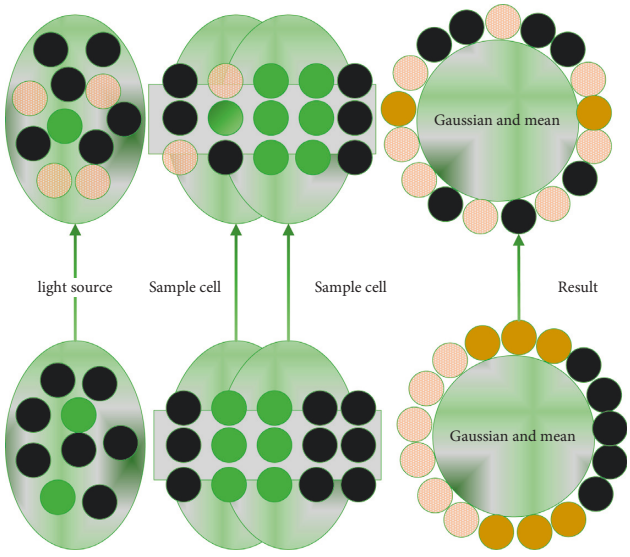

Figure 5: 3D Gaussian centrality.

which provides an alternative method of obtaining pre-trained models and can save a significant amount of time. The theory of universal approximation states that the more implicit units a network contain, the better it is at approximating arbitrary functions, that is, the better the network can achieve a global optimum given enough parameters. However, in practice, the network tends to fall into local optima during training, which is more likely for small networks with weak generalization ability, and it is more difficult to jump out of local optima when the learning rate is small. Therefore, it is necessary to increase the learning rate again during the training of the small network to jump out of the local optimum, as shown in Figure 6.

The iterative training method focuses on the learning rate during training. Existing training methods follow a training strategy that starts with a learning rate of a specific value, which is empirical, and then decreases the learning rate every few training grounds after a certain period [19]. Iterative training differs by changing the iteration period of

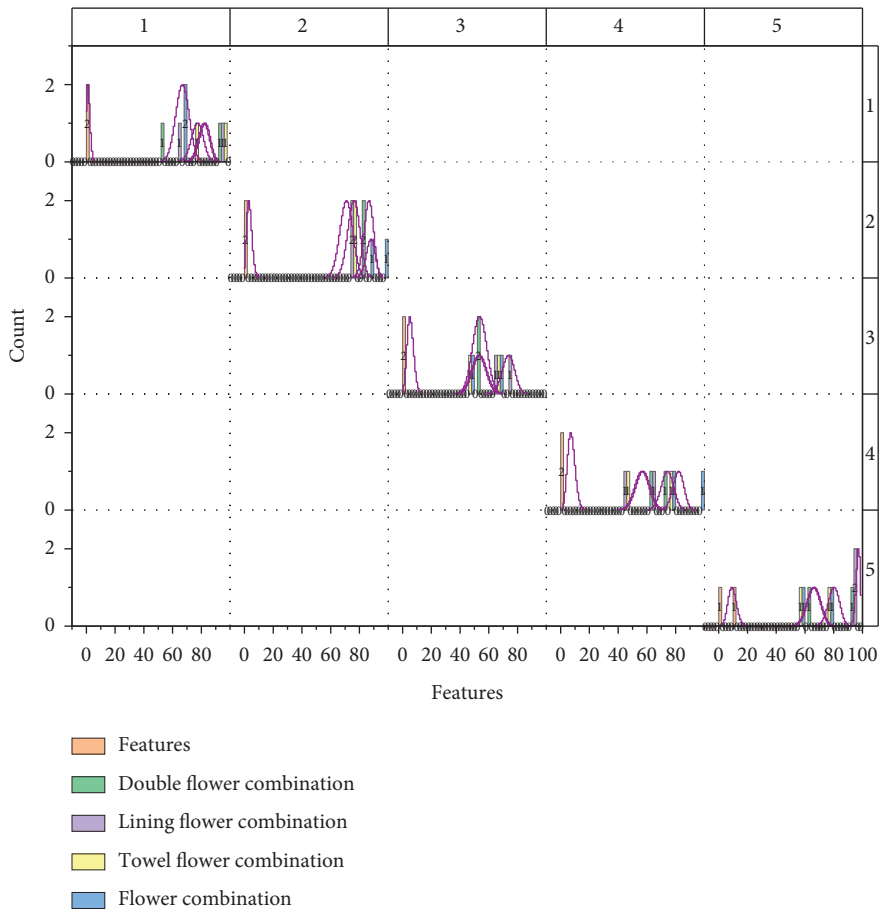

Figure 6: Experimental results of the two HOG extractions on the dataset.

the learning rate, and the training method can be divided into multiple phases: the first phase of the training method is identical to the conventional training method; subsequent training methods follow the principle of initializing the network parameters using the optimal training model from the previous phase, resetting the learning rate, retraining from the specified period, and repeating the phrase several times.

6. Analysis of Results

6.1. Algorithm Performance Analysis. The HOG features are used to characterize the local appearance and shape of the movements, and when the similarity between movements in a dance combination is too high, it will make recognition more difficult and affect the recognition accuracy. From the table, we can see that the recognition rate of the HOG feature is 42.8% and 40% for the double-flower combination and the piece-flower combination, respectively, and is higher than that of 33.3% and 29.2% for the hand towel and piece-flower combinations. Among the four dance combinations in the FolkDance dataset, the similarity between the dance movements in the heeled double flower combination and the rip flower combination is much smaller than that in the hand towel flower combination and the Katana combination, but there are similar dance movements in both the hand towel flower combination and the Katana combination, especially in the Katana combination where there are multiple similar movements and the same movement is divided into different directions, which also makes it more difficult to identify the dance movements. In this group, the recognition rate of HOG features was the lowest among the four groups at 29.2%.

In this paper, the HOF features used to characterize the movement information of the dance movements were 38.1% and 33.3% in the combination of the double flower in the following step and the flower in the piece, respectively, which were not very different from the 37.5% and 33.3% in the combination of the hand towel and the flower in the piece. It is only related to the current node situation; that is, each node may belong to each discovered community, so that overlapping community discovery can be achieved in the random walk to discover the community structure. In addition, the results of the first two groups also show that when the similarity between the movements in the dance combinations is low, the performance of the HOG feature in this paper is better than the HOF feature for complex dance movement recognition, as shown in Figure 7.

The difference in recognition rate between the audio signature features in this paper is not significant in the four groups, which also shows that the audio features have

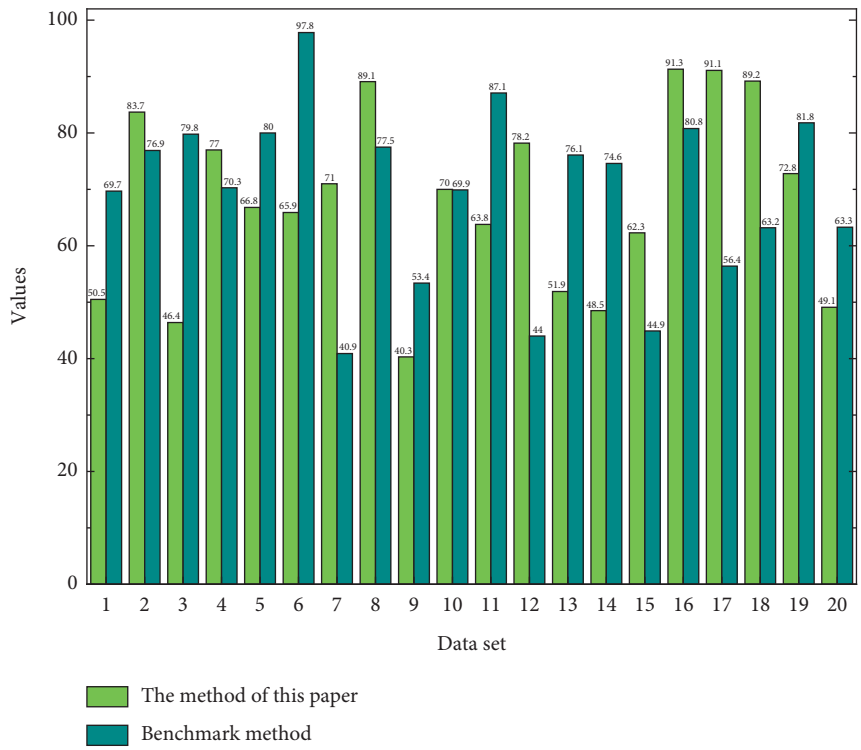

Figure 7: Comparison of the experimental results between this method and the benchmark method in the four groups.

maintained a good recognition rate and are not affected by factors such as the complexity of the movements. Compared to a single feature, the recognition rate for all groups of dance movements has improved significantly. Despite the influence of similar dance movements, the recognition rates of the multifeature fusion method are 50% and 45.8% for the hand towel and flower combinations, respectively, which are higher than the recognition rates of all single features in both combinations. The results for the first two groups are 52.4% and 53.3%, respectively, which are lower but more robust than the performance of single features in the same group. This suggests that by, fusing audio features in dance movement recognition and learning the appropriate weights through the training process, the impact of complex and similar movements can be reduced, thus maintaining a certain level of accuracy.

In the four dance combinations, the overall recognition rate of this method is higher than that of the benchmark method. Only in the case of the Satsuki combination is the recognition rate of 53.3% lower than the 60% of the benchmark method. In the other combinations, especially in the hand towel and piece flower combinations where the dance movements are too similar, the method performs better than the benchmark method. This also indicates that the baseline method based on the fusion of trajectory features cannot accurately characterize the dance movements when the dance movements are too complex and when there are similar movements and self-obscuration, as shown in Figure 8.

It can be easily seen that the modularity value of the EDME algorithm in this paper does not change much with the increase in the network size, and the experimental results are relatively stable. Compared with the LFM algorithm, which is more effective in the overlapping community in recent years, the difference is not too big, and, even in some networks, the difference is greatly improved. For example, the improvement was 16.38% in the relatively large Word Association network. The largest improvement is achieved when compared to the LC algorithm, which only considers adjacent edges in the network. This shows that the algorithm works relatively well in general-purpose networks, which indicates that the algorithm has good generality. Use Kinect equipment to obtain real-time skeleton data and depth images of key frames, combine static models, skeleton data, depth images, and related human deformation algorithms to capture human motion in real time, and reconstruct the 3D human dynamic model during motion.

The experiments were designed to test the AP scores after training with and without the module, without the use of the iterative training method and the postprocessing function parameterized by Soft-Argmax. The experimental analysis shows that the module has a significant effect on improving the average accuracy by about 2 points, indicating that the module can improve the prediction accuracy and that the increase in the number of model parameters and floating-point operations is within an acceptable range.

6.2. Analysis of Experimental Results. Due to the limited length of the collected video and the number of actors, the model would not generalize well to real scenes if trained using raw pixel information. Therefore, a more abstract feature was chosen, a two-dimensional sequence of key point coordinates containing action information. To detect the

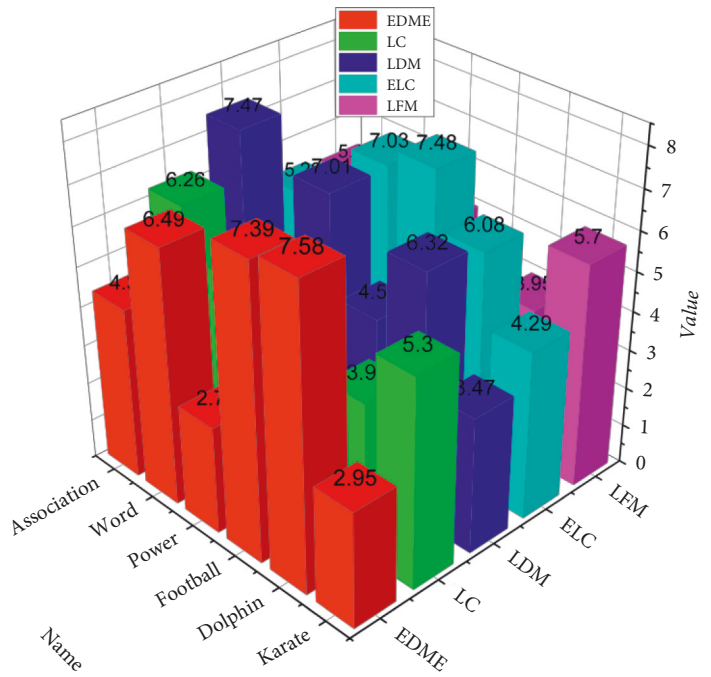

Figure 8: Comparison of experimental results of modularity of classical real network experimental data.

touchdown state of the foot key points, a fully convolutional structured network with residual connections is used, taking as input a 3D coordinate sequence of human key points and transforming the input by temporal convolution, as shown in Figure 9. The convolutional model can process data in both batch and temporal parallelism, but the RNN cannot be parallelized in temporal order; in the convolutional model, the length of the gradient path between input and output is fixed, independent of the sequence length, which avoids the gradient vanishing and gradient explosion problems that affect RNN models, and the convolutional structure is also able to accurately control the temporal length, which was experimentally found to be useful for modeling the temporal nature of the foot touchdown estimation task This was found to be useful for temporal modeling of the foot touch estimation task.

To model foot touch discrimination, details of the construction of the dataset and the design of the network structure are presented in detail; that is, a large amount of data on actions such as walking and descending stairs is collected and labeled, and the network structure is chosen in such a way that the current foot key point state can be judged using multiple frames of consecutive actions as input, fully because the actions consist of consecutive frames. For example, given the initial state of walking, the walking controller will calculate and obtain the character. Finally, the effectiveness of the foot touchdown discriminator on a self-built dataset is quantitatively evaluated, and the effect of the foot touchdown status on consecutive frames is visualized. The experiments show that the proposed foot touch discriminator can, to a certain extent, replace physical sensors in capturing touchdown conditions.

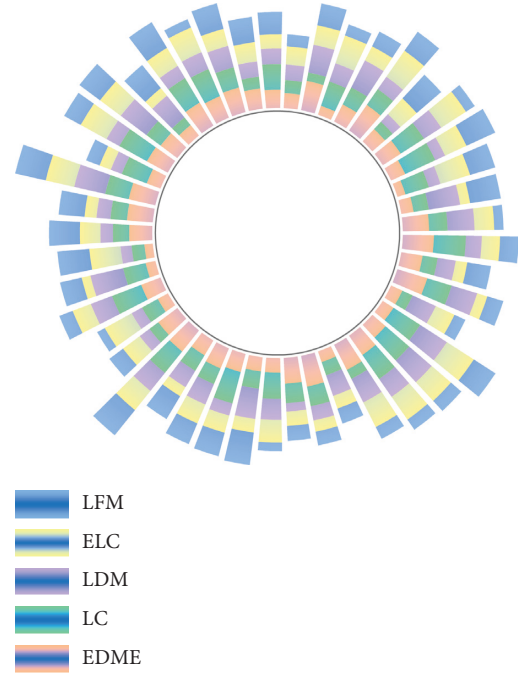

Figure 9: Comparison of action recognition rates.

To analyze the effectiveness of using the foot touch judgment for solving slip and foot-through-ground problems, the ground is input for reference to show the difference between the two. Kinematic metrics, primarily three kinematic physical metrics, were metrically evaluated for both methods, including floating, ground penetration, and skip a step, and these metrics were evaluated for accuracy of foot

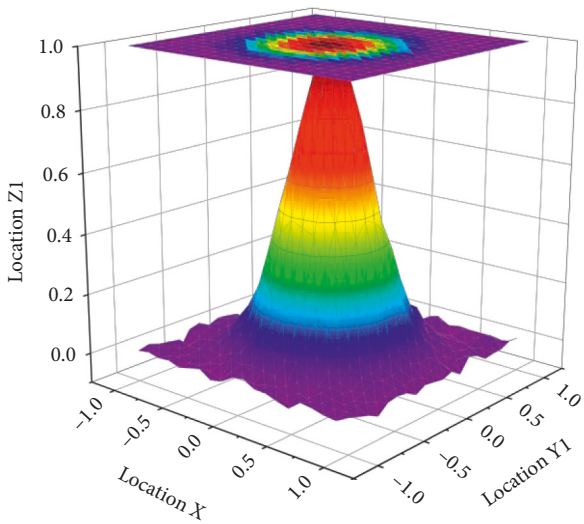

Figure 10: Dance movement display results.

contact optimization. Specifically, given a foot contact ground label, the floating, slipping, and ground penetration states of the foot joints are calculated. Floating is when the foot is in contact but more than three centimeters off the ground; ground penetration is when the foot passes three centimeters into the ground in all cases, and slip is when the foot moves more than two centimeters in the contact state. The lower the metric result, the better.

The motion capture technology allows the data of the nodes to be obtained. The paper initially chose vertex animation, which yielded unsatisfactory results after which the paper chose a method like skeletal animation to obtain an image of the terminated action; the main idea is as follows. The physical model obtained in this way is usually called a controller. Each controller corresponds to a specific motion behavior of the character. The way to obtain the image of the next frame is to find the world coordinate system of its root node at the next frame and the position of other nodes relative to the root node, the root node is calculated as a vector difference, and the relative position of each node is stored as a constant in the action library. In this way, the world coordinate system of the positions of all the nodes of the body can be obtained and an image of the body at the time of the frame can be drawn. With this skeletal animation-like algorithm, it is possible to avoid distortions in the human animation caused by computational errors. It is possible to ensure that errors are fixed in time while ensuring that each frame is a normal image of the human body in action, as shown in Figure 10.

As shown in Figure 10, from left to right, the Laban dance score is generated based on rule, dynamic planning, and extreme learning, respectively. Because they are recognized on the premise of the same cut movements, they differ only in the recognition symbols. The rule-based approach is easier to understand and requires only conditional judgments in recognition, so the computational effort is almost nonexistent. In the process of animation production, a lot of workforce and material resources are often required to adjust the model. However, there are also significant disadvantages, as the human body has many complex movements and many classes, with each rule corresponding to one movement, so many rules need to be developed, which can lead to conflicts between the rules.

In addition, the rules focus on the posture and do not describe the process of the movement in detail, but the analysis of the movement of the center of gravity of the support bar needs to focus on the process, so the rule-based approach is not effective in identifying the movement of the support bar. However, the dynamic time regularization and limited learning-based methods provide a better description of the process and are more effective than the rule-based ones in identifying the support bar. However, the dynamic programming-based approach has some gaps in accuracy compared to the extreme learning and is computationally intensive in terms of interframe distance and pathfinding, so this paper chose to adopt extreme learning as the action recognition algorithm nested in the Laban dance score automatic generation platform.

7. Conclusion

This paper analyzes the current situation of the task of generating 3D human animations and presents a method for generating 3D human animations based on action recognition and a priori models, which can switch the most similar models according to different actions and can effectively alleviate the model distortion. A 3D human animation generation technique based on action recognition and a priori models is proposed. The a priori model library is constructed by simply adding skeletal point information to the model and roughly binding the skeletal skin and then using a quadratic-based rotational deformation method; the model with the closest movement characteristics is selected from the model library based on the results of movement recognition and deformed on its basis. Therefore, it is of great practical significance to study and make 3D human animation. Existing 3D human animation generation methods have high requirements on both the quality of the model itself and the accuracy of skeletal skinning. This method effectively improves the local distortion of the model deformation. For the study and analysis of the characteristics of dance movements at the same time, this paper proposes an effective method of extracting features by equal segmentation of the dance movement videos. The segmented videos are then subjected to an edge feature accumulation operation separately, where all the edge features in the video images within each segment are accumulated into one image, directional gradient histogram features are extracted, and, finally, a set of directional gradient histogram features vectors are used to characterize the appearance and shape of the dance movements in the video. Our dataset is therefore recorded in fixed scenes and with a single-person dancing, with a total of 84 dance movement videos from three people and four groups, as well as other single-person, multicategory datasets for other aspects of dance research.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by Shanghai Normal University.

References

[1] Y.-S. Liu, Z.-Z. Qiu, X.-C. Zhan, H.-N. Liu, and H.-N. Gong, "Study of statistical damage constitutive model of layered composite rock under triaxial compression," *Applied Mathematics and Nonlinear Sciences*, vol. 6, no. 2, pp. 299–308, 2021.

[2] Y. Q. Zhu, Y. M. Cai, and F. Zhang, "Motion capture data denoising based on LSTNet autoencoder," *Journal of Internet Technology*, vol. 23, no. 1, pp. 11–20, 2022.

[3] T. L. Gomes, R. Martins, J. Ferreira, R. Azevedo, G. Torres, and E. R. Nascimento, "A shape-aware retargeting approach to transfer human motion and appearance in monocular videos," *International Journal of Computer Vision*, vol. 129, no. 7, pp. 2057–2075, 2021.

[4] M. Cancan, M. Imran, S. Akhter, M. K. Siddiqui, and M. F. Hanif, "Computing forgotten topological index of extremal cactus chains," *Applied Mathematics and Nonlinear Sciences*, vol. 6, no. 1, pp. 439–446, 2021.

[5] A. Switonski, H. Josinski, and K. Wojciechowski, "Dynamic time warping in classification and selection of motion capture data," *Multidimensional Systems and Signal Processing*, vol. 30, no. 3, pp. 1437–1468, 2019.

[6] L. Chuan-Xi, H. Jun, L. Wei-Xiong, H. Zhi-Yong, S. Qi-Bin, and M. James, "Study on water damage mechanism of asphalt pavement based on industrial CT technology," *Applied Mathematics and Nonlinear Sciences*, vol. 6, no. 1, pp. 171–180, 2021.

[7] J. Yang, X. Guo, K. Li, M. Wang, Y. K. Lai, and F. Wu, "Spatio-temporal reconstruction for 3D motion recovery," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 6, pp. 1583–1596, 2019.

[8] P. B. Zhang and Y. S. Hung, "Articulated deformable structure approach to human motion segmentation and shape recovery from an image sequence," *IET Computer Vision*, vol. 13, no. 3, pp. 267–276, 2019.

[9] L. Hailong, "Role of artificial intelligence algorithm for taekwondo teaching effect evaluation model," *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 2, pp. 3239–3250, 2021.

[10] K. Yin, H. Huang, E. S. L. Ho et al., "A sampling approach to generating closely interacting 3d pose-pairs from 2d annotations," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 6, pp. 2217–2227, 2018.

[11] E. Marchi, B. Schuller, A. Baird et al., "The ASC-inclusion perceptual serious gaming platform for autistic children," *IEEE Transactions on Games*, vol. 11, no. 4, pp. 328–339, 2018.

[12] D. Liu, C. Li, and P. Zhong, "Walking model and planning algorithm of the over-obstacle pipe climbing robot," *Applied Mathematics and Nonlinear Sciences*, vol. 6, no. 1, pp. 243–262, 2021.

[13] F. De Chaumont, E. Ey, N. Torquet et al., "Real-time analysis of the behaviour of groups of mice via a depth-sensing camera and machine learning," *Nature biomedical engineering*, vol. 3, no. 11, pp. 930–942, 2019.

[14] L. Cen, D. Ruta, L. M. M. S. Al Qassem, and J. Ng, "Augmented immersive reality (AIR) for improved learning performance: a quantitative evaluation," *IEEE Transactions on Learning Technologies*, vol. 13, no. 2, pp. 283–296, 2019.

[15] P. Paliyawan, W. Choensawat, and R. Thawonmas, "Mossar: motion segmentation by using splitting and remerging strategies," *Multimedia Tools and Applications*, vol. 77, no. 21, pp. 27761–27788, 2018.

[16] G. Saul and C. Ells, "Shadows illuminated. Understanding German expressionist cinema through the lens of contemporary filmmaking practices," *Acta Universitatis Sapientiae, Film and Media Studies*, vol. 16, no. 1, pp. 103–126, 2019.

[17] W. A. Lyons, E. C. Bruning, T. A. Warner et al., "Megaflashes: just how long can a lightning discharge get?" *Bulletin of the American Meteorological Society*, vol. 101, no. 1, pp. E73–E86, 2020.

[18] J. Robb, "Art (Pre)History: rnvcnbae," *Journal of Archaeological Method and Theory*, vol. 27, no. 3, pp. 454–480, 2020.

[19] Y. Qi, G. Jiang, M. Yu, Y. Zhang, and Y. S. Ho, "Viewport perception based blind stereoscopic omnidirectional image quality assessment," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 10, pp. 3926–3941, 2020.