www.cad-journal.net
Computer-Aided Design™
AND APPLICATIONS

# Implementation of Computer Aided Dance Teaching Integrating Human Model Reconstruction Technology

Yafang Zhao[1] and Hongpeng Yang[2]

[1]Department of Dance, Zhengzhou Preschool Education College, Zhengzhou 450000, China,
zhaoyafang@zzpec.edu.cn
[2]Academy of Fine Arts, Henan University, Kaifeng, Henan 475001, China,
10170082@vip.henu.edu.cn

Corresponding author: Hongpeng Yang, 10170082@vip.henu.edu.cn

**Abstract.** In the stage of dance teaching and training, advanced science and technology can lay a technical foundation for it, and at the same time give dance art digital characteristics. Dance CAI system meets the requirements of digital dance and provides a convenient visual mode for dance educators and learners in the form of full perspective. In order to achieve better artistic performance, dance computer-aided instruction (CAI) system can be introduced in time. In this article, the reconstruction technology of human motion model based on mixed features is proposed, and the dance motion recognition model is constructed by combining the convolution operation principle. The human body is allowed to use a variety of body language to express itself and communicate with the computer. The early fusion modal information usually has different characteristics, which leads to the inconsistency of the spatial and temporal dimensions of the extracted feature vectors. The method in this article has high accuracy and low error, and can help optimize the dance teaching effect.

**Keywords:** Reconstruction of Human Model; Computer-Aided Instruction; Dance Movement Recognition; Mixed Feature
**DOI:** https://doi.org/10.14733/cadaps.2024.S10.196-210

## 1 INTRODUCTION

In the process of dance movement teaching, advanced science and technology can lay a technical foundation for it, and at the same time give dance art digital characteristics. Motion recognition is defined as being able to automatically identify the content of a specific video and further identify the problem of the motion occurring therein. Adams and Onwadi [1] combined human body model reconstruction technology to implement dance teaching algorithms. Firstly, it is necessary to obtain data on dance movements. This can be obtained through recording or real-time capture devices. The data can be two-dimensional or three-dimensional, including information such as the position, posture, and velocity of various parts of the human body. Create a three-dimensional human body model in a computer, and you can choose different human body model libraries or build your own according to your needs. The human body model should be able to simulate the posture and movements of the human body based on input motion data. Input the obtained dance action data into the human body model, and use algorithms to simulate the corresponding dance actions based on the input data. This requires solving many problems, such as posture matching, motion smoothing, adjusting body proportions and dynamic range, etc. By comparing the human body model with dance movements, it is possible to detect the accuracy of students' dance movements. If the movements are not accurate, the algorithm can provide feedback and suggestions to help students improve their dance movements. Dance CAI system meets the requirements of digital dance and provides a convenient visual mode for dance educators and learners in the form of full perspective. Before the application of deep learning method, the traditional computer vision method of manually extracting features was widely used in dance movement recognition. Angermann et al. [2] used unsupervised single shot depth estimation using perceptual reconstruction for machine vision applications. Unsupervised single shot depth estimation using perceptual reconstruction is an important machine vision and application technology. This technology utilizes deep learning models to restore depth information from a single image, thereby achieving the conversion from a two-dimensional image to a three-dimensional space. Unsupervised single shot depth estimation has a wide range of applications in machine vision and applications. For example, autonomous vehicles need to accurately perceive the depth information of their surrounding environment in order to safely travel and avoid obstacles. Unsupervised single shot depth estimation can provide an efficient and accurate depth perception method for autonomous vehicles. In addition, unsupervised single shot depth estimation can also be applied to fields such as 3D reconstruction, virtual reality, and augmented reality. By restoring depth information from a single image, realistic 3D models can be generated, enabling a more realistic and immersive virtual experience. As a new productive force, AI has brought great changes to the traditional dance industry. Motion recognition based on video is to study how to identify specific dance movements from specified video sequences. On the basis of successful motion capture and feature extraction, motion recognition automatically recognizes dance movements by analyzing the obtained dance movement feature parameters. Guo et al. [3] analyzed videos of three-dimensional human movements through a computer. Generating videos of three-dimensional human movements is a challenging task, but in recent years, significant progress has been made in the field of computer vision. Firstly, it is necessary to obtain data on real human movements. This can be achieved by using motion capture devices such as infrared cameras or accelerometers. These devices can capture human motion information, including joint angles, accelerations, etc. In computer graphics, it is necessary to establish a three-dimensional human body model. This model can be a simple geometric body, such as a sphere, cylinder, etc., or a detailed model, including human muscles, skin, etc.

Liu et al. [4] reconstructed human joint motion using computational fabrics. Computational fabric is an intelligent textile that can perceive and respond to human motion and physiological parameters through sensors and electronic devices. This technology can be used to reconstruct human joint motion, thereby better monitoring and treating human health issues. In the ACM interactive, mobile, wearable, and ubiquitous technology collection, there are many related papers exploring how to use computational fabrics to reconstruct human joint motion. These papers usually introduce the design and implementation methods of computational fabrics, as well as how to use this technology to monitor and improve human movement and health. In order to obtain a realistic and high-precision 3D model of human body, the production process is generally very long, which requires a lot of time and manpower. However, the 3D reconstruction technology, which can obtain the real object model by using simple equipment such as ordinary cameras, can solve this problem to some extent. The application of motion recognition in human interaction is mainly to reduce the contact interaction between the human body and the computer, and realize that the human body naturally carries out various actions, so that the computer can perceive and obtain the state information of the human body through the visual interface. The human body is allowed to use a variety of body language to express itself and communicate with the computer.

The early fusion modal information usually has different characteristics, which leads to the inconsistency of the spatial and temporal dimensions of the extracted feature vectors. This is an obstacle to the network fusion of potential multimodal information in low-level feature space. However, the late fusion lacks the correlation exploration of feature levels between modes. Therefore, when the information of different modes is different, it may be too simple to fuse only in the decision-making stage, and the complementary information of each mode can not be fully and effectively utilized. It provides a more convenient visual mode for artistic creators in the form of full perspective, and its optimized operation process provides a more interactive creative platform for directors and actors. In order to achieve better artistic performance, timely introduction of dance CAI system can continuously provide diversified visual modes for dance creators and optimize the interactive communication platform between educators and learners. Aiming at the deficiency of traditional dance teaching mode, this article puts forward the reconstruction technology of human motion model based on mixed features, which realizes the intelligent decomposition and recognition of dance movements and provides technical support for the algorithm realization of dance CAI system.

Ma et al. [5] analyzed an optimal driving mode for electrical impedance tomography scanning suitable for human-computer interaction applications. In order to monitor human physiological parameters in real-time, it is necessary to choose appropriate scanning speed and data acquisition frequency. A slow scanning speed can lead to lag in image reconstruction, while a fast scanning speed may cause an increase in signal noise. Choosing appropriate signal processing and image reconstruction algorithms can optimize the quality and resolution of images. For example, methods based on sparse representation can improve the accuracy of image reconstruction while reducing the dimensionality of measurement data. In summary, the optimal EIT driving mode suitable for human-computer interaction applications should comprehensively consider the above factors to achieve the best image quality and real-time monitoring effect. Motion analysis is to judge the standard of human motion by analyzing various data of human motion, and help users obtain motion information. The reconstruction of human body model is to reconstruct a realistic 3D human body model by using the obtained 3D human body data. When it is needed to evaluate and analyze dance videos, if professionals watch them manually, the efficiency will undoubtedly be very low. At this time, the introduction of human motion recognition technology will greatly reduce the workload of relevant personnel. In this article, the method of dance movement recognition and modeling based on human model reconstruction technology is studied, and the following innovations are made:

(1) In this article, the reconstruction technology of human motion model based on mixed features is proposed, and the dance motion recognition model is constructed by combining convolution operation principle, which provides technical support for the algorithm realization of dance CAI system.

(2) The random projection algorithm is used to reduce the dimension of the feature vector, and the expansion convolution module is introduced to obtain a wider effective receptive field without changing the size of the convolution kernel.

In order to realize the intelligent decomposition and recognition of dance movements and help optimize the effect of dance teaching, this article puts forward the reconstruction technology of human movement model based on mixed features, and combines the convolution operation principle to construct the dance movement recognition model; Experiments verify the motion recognition performance of this method; Finally, the work and contribution of this article are summarized and the possible improvement direction is pointed out.

## 2 RELATED WORK

Computer Aided Design and Manufacturing (CAD/CAM) for head and neck reconstruction is a method of using computer technology for head and neck tumor resection and reconstruction surgery, while traditional surgical planning is carried out through manual measurement and planning by doctors. Padilla et al. [6] compared these two surgical planning methods. When performing head and neck tumor resection surgery, traditional surgical planning requires doctors to manually measure and plan the surgical route and resection range, which has a certain degree of subjectivity and error risk. And CAD/CAM technology can use computers for 3D reconstruction and virtual surgery, accurately calculate tumor location and resection range, and design personalized surgical plans, improving the accuracy and safety of surgery. Shi et al. [7] analyzed an algorithm for reconstructing three-dimensional human motion from monocular videos. This algorithm utilizes video sequences with skeleton consistency to estimate the posture and motion trajectory of the human body through deep learning methods. Specifically, Motionet first preprocesses the input video sequence to extract a consistent human skeleton model. Then, it uses a deep neural network to predict the posture of human joints in each frame. This pose information are used to reconstruct the three-dimensional motion trajectory of the human body. The main advantage of this algorithm is its ability to accurately reconstruct three-dimensional human motion from monocular videos without the need for any special equipment or markers. In addition, Motionet can also handle human bodies with different postures and movements, with good generalization ability. Song et al. [8] conducted through wall human pose reconstruction using ultra-wideband MIMO radar and 3D CNN. The combination of broadband MIMO radar and three-dimensional convolutional neural network (3D CNN) can provide an effective method for reconstructing human posture through walls. Ultra-wideband MIMO radar has the advantages of high resolution, high sensitivity, and strong anti-interference ability, which can penetrate walls and perceive indoor human posture. The signals received by MIMO radar can be converted into images, and then 3D CNN is used for human pose recognition and reconstruction. 3D CNN can extract features from radar images and classify and predict human posture. Compared with traditional 2D image processing methods, 3D CNN can better process the spatial information of radar images, improve the accuracy and robustness of attitude recognition. Therefore, utilizing ultra-wideband MIMO radar and 3D CNN for wall penetrating human pose reconstruction can provide an effective solution for fields such as smart homes and security monitoring. Robust fusion is a technique based on monocular RGBD streams, aimed at reconstructing robust volume performance in human-object interaction. This technology uses a monocular RGBD camera to obtain depth information in the scene, and extracts three-dimensional geometric information in the scene through a series of image processing and computer vision algorithms. On this basis, the technology uses fusion algorithms to fuse multiple depth images to obtain more accurate and robust 3D reconstruction results. In the context of human object interaction, this technology can effectively handle issues such as occlusion, shadows, and changes in lighting, thereby achieving robust volume performance reconstruction in complex scenes. Therefore, robust fusion technology can be applied to fields such as virtual reality, augmented reality, and robot vision, providing new ideas and methods for the development of these fields [9]. Sun et al. [10] analyzed the human motion transfer problem with three-dimensional constraints and detail enhancement. Human motion transfer with 3D constraints and detail enhancement is a complex problem, but it has wide applications in computer vision and human motion capture. IEEE Pattern Recognition is a mathematical method and tool for processing and analyzing patterns, which can be applied to the problem of human motion transfer. In the process of human motion transfer, some mathematical models and methods can be used to achieve the processing of three-dimensional constraints and detail enhancement. For example, methods based on kinematic models can use rigid or elastic models to describe human motion and constrain the range of motion. Machine learning based methods can learn the features of real human motion and then use these features for motion transfer.

Vo et al. [11] analyzed the spatiotemporal beam adjustment of dynamic 3D human body reconstruction in the field. Dynamic 3D human body reconstruction in the field is a challenging task as it involves precise capture and reconstruction of human motion in 3D space. Spacetime beam adjustment is a technology used to optimize sensor data collection and processing, which can be used to improve the accuracy and efficiency of dynamic 3D human body reconstruction in the field. IEEE Pattern Analysis and Machine Intelligence is an international academic conference that involves the application of machine learning and artificial intelligence in various fields. At this conference, scholars and experts will discuss the latest research achievements, technologies, and trends, including spatiotemporal beam adjustment techniques for dynamic 3D human body reconstruction in the field. Therefore, applying spatiotemporal beam adjustment technology to dynamic 3D human body reconstruction in the field and analyzing its performance at IEEE mode analysis and machine intelligence conferences can provide valuable references for further development in this field. Restoring old photos through deep potential spatial translation is a method of image restoration using deep learning techniques. In this method, Wan et al. [12] used a neural network structure called autoencoder, which has two parts: encoder and decoder. Firstly, the encoder converts the input image into its potential representation, and the decoder converts the potential representation back to the original image. By training an autoencoder to encode and decode the complete image, only the damaged image is processed using the encoder to recover the image. This technology has published multiple related papers in the IEEE Journal of Pattern Analysis and Machine Intelligence. These papers describe how to use deep learning techniques to solve various image processing problems, including image restoration, super-resolution, image classification, and object detection. These technologies have been widely applied in various practical applications, such as medical image processing, remote sensing image analysis, digital image restoration, etc. Wang and Liu [13] analyzed the impact of game teaching on primary school students' dance learning. The impact of game teaching on primary school students' dance learning is profound. Firstly, it can stimulate the learning interest of primary school students and enhance their initiative and enthusiasm. Through games, students can learn dance in a relaxed and enjoyable atmosphere, thereby reducing their sense of strangeness and fear towards dance. Secondly, game teaching helps primary school students better understand and master dance movements. As dance is a dynamic art form, learning dance requires continuous practice and repetition. Through game teaching, students can unconsciously master dance movements during play, thereby improving their dance skills. Finally, game teaching can also cultivate the innovative ability and independent choreography ability of primary school students. By involving students in games, teachers can discover their individual characteristics and innovative potential, thereby guiding students to independently choreograph. This not only cultivates students' creativity, but also enables them to better understand the connotation and significance of dance.

Unstructured fusion is a real-time 4D geometric and texture reconstruction method that uses commercial RGBD cameras to obtain geometric shape and texture information of object surfaces. This method is different from traditional structured methods because it does not require any preset markers or structured scenes and can be applied to various objects and environments. Xu et al. [14] conducted real-time 4d geometric and texture reconstruction using commercial rgbd cameras. The basic principle of unstructured fusion is to obtain point cloud data on the surface of an object using an RGBD camera, and then use deep learning or other algorithms to estimate the surface normal vector and texture coordinates of each point. This information can be used to generate 4D geometric models of objects, including their shape and texture. The advantage of unstructured fusion method is that it can be applied to various different scenes and objects without requiring any preset markers or structured scenes. In addition, due to its use of commercial RGBD cameras, it can be easily integrated into existing systems. Human dynamics in monocular videos with dynamic camera motion is a complex research field that involves multiple disciplines such as computer vision, computer graphics, and human dynamics. This research aims to extract and understand human motion information from monocular videos, in order to better understand aspects such as human behavior, exercise, and health. Yu et al. [15] analyzed human dynamics in monocular videos with dynamic camera motion. In ACM Transactions on Graphics, there are many related papers exploring how to extract and model human dynamics from monocular videos. These papers typically introduce methods for tracking and modeling human motion using computer vision and graphics techniques, as well as how to use this information to understand and simulate human dynamics. Zhai [16] conducted feature expression and attribute mining for dance action recognition. Dance action recognition is a challenging task as it involves classifying and analyzing dynamic and complex actions in video or image sequences. The methods based on feature representation and attribute mining have extensive applications in this field, but they do face some complexity. Identifying dance movements requires effective feature extraction and expression. Features can include the geometric shape, posture, motion trajectory of the human body, as well as the spatial and temporal structure of dance movements. Extracting these features and accurately expressing them requires a deep understanding and understanding of dance. In addition, feature extraction and expression may require a lot of manual work, which can be very time-consuming and difficult to expand. There may be significant inter class variations between different dance movements, for example, the styles and techniques of various dances can greatly change the expression of the same movements. At the same time, even within the same dance, there may be significant intra class variations in the same movement, for example, due to differences in the performer's personal style, skill level, or emotional expression. Neural animation mesh is a technology used for modeling and rendering human behavior, based on a combination of neural networks and graphics. Through neural animation grids, realistic human behavior animations can be generated, including walking, running, jumping, and more. ACM graphics is an international academic conference in the field of computer graphics, covering various aspects of computer graphics, including human behavior modeling and rendering. At this conference, scholars and experts will showcase their latest research achievements and technologies, including neural animation grids. Therefore, Zhao et al. [17] used neural animation grids to model and render human behavior, and presented relevant results at ACM graphic conferences. This can provide new ideas and methods for the development of computer graphics. Zhao and Ye [18] analyzed the spatial reconstruction of audiovisual media based on artificial intelligence and virtual reality. The spatial reconstruction of audiovisual media based on artificial intelligence and virtual reality is an emerging research field that combines various technologies such as artificial intelligence, virtual reality, and audiovisual media, aiming to create a brand-new media experience. In this experience, viewers can enter an audio-visual media space constructed by artificial intelligence through virtual reality technology. This space can be dynamically adjusted according to the audience's behavior and emotions, forming a personalized experience. For example, viewers can explore this space and interact with virtual environments through head-worn displays and gesture recognition devices. Zheng et al. [19] used sparse inertial sensors (such as accelerometers and gyroscopes) to collect human motion data and used deep learning algorithms to reconstruct human posture. During the training process, selecting appropriate training data and optimizing the placement of sensors are crucial for improving model performance and generalization ability. In order for the model to adapt to various human movements and postures, it is necessary to select training data with sufficient diversity. This includes data on different genders, ages, body types, exercise types, and postures. Before training, appropriate preprocessing of the data, such as normalization, noise removal, interpolation, etc., can improve the accuracy and robustness of the model. The arrangement of sparse inertial sensors in different parts of the human body can affect the accuracy and reliability of data collection. Through experiments and simulations, the optimal sensor placement can be found to improve data quality and reconstruction accuracy.

## 3 METHODOLOGY

After the human body is detected and tracked, the next step of motion recognition is needed. In human movement, due to the influence of external environment, soft clothes and other factors, contour information cannot well reflect the action details of human movement in many cases. In human motion recognition, bone data has attracted more and more attention because of its robustness and compactness. In traditional methods, the human structure is usually modeled by manually constructing or learning the characteristics of human joint points. However, these methods have the problem of too complicated design process, and often ignore the internal relationship between joint points, which cannot achieve satisfactory results.

In dance movement teaching, most movements are realized by the performer's two arms and two legs. On the pixel block of the image, the pixel value in the pixel block is operated with the corresponding convolution kernel to obtain the output value of the corresponding block; Then select new pixel blocks step by step, move the convolution kernel, and operate in turn to get the output after convolution of the whole image. In the task of dance movement recognition, data modes are generally divided into three categories: video data, depth image and bone movement sequence. According to the different data modes of the recognition task, different algorithms or models are designed to complete the recognition task. The coordinate division of human joints is shown in Figure 1.

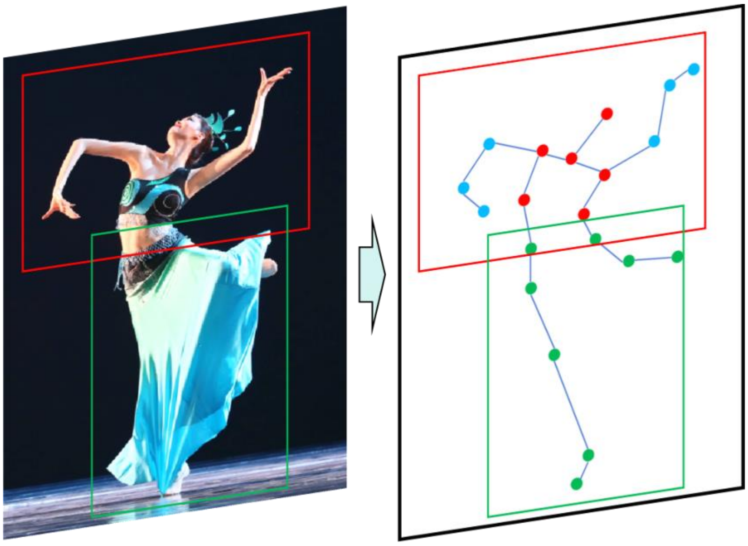

**Figure 1**: Division of human joint points.

One of the most natural ways of communication between people is action. Different actions have different meanings, so we can get more information by analyzing human actions. If we can track the human body accurately, we can observe and learn some behaviors of the human body more conveniently. A dance usually contains a variety of basic dance movements, and the existing action recognition methods generally recognize a single action, so it is needed to segment the dance video sequence before recognition, so that each video segment only contains one action. Video frame sequence segmentation stems from the fact that action video contains many types of actions. The basic idea is to extract key frames by various methods, and then realize video frame segmentation according to key frames to obtain the main action sequence or key sequence composed of key frames in the video.

In the stage of video motion recognition research, we usually face a lot of calculations, and these calculations are not only the complexity of the algorithm itself, but also affected by a large number of invalid actions in the video. Invalid action means that the stage of human motion recognition will increase the amount of calculation, but it cannot improve the accuracy of motion recognition. Therefore, it is needed to segment the video frame sequence to obtain the key action sequence, so as to remove the complicated calculation caused by invalid actions.

In order to calculate the motion posture data of each bone and joint in the human motion model, it is needed to transform the posture data collected by sensors, so establishing a reasonable coordinate system can quickly realize the posture transformation. The mutual transformation of the same point $P_A(X,Y,Z)$ in the established three kinds of coordinate systems can be calculated by the translation $R$ and rotation transformation $T$ between coordinate axes:

$$P_A^{'} = RP_A + T \tag{1}$$

The rotation amount usually assumes that the initial angle of each axis is 0° or 180°, and the translation amount is the 3D length parameter of the model bone structure.

Human body joint point recognition connects human bodies through joint points in Kinect, so Kinect judges the coordinate position of human bodies by analyzing the front and side joint points. The density estimator of the following body components is defined as:

$$f_c(\hat{x}) \propto \sum_{i=1}^{N} w_{ic} \exp\left( -\left\| \frac{\hat{x}-\hat{x}_i}{b_c} \right\|^2 \right) \tag{2}$$

Where $\hat{x}$ is that coordinate of 3D space, $N$ is the numb of pixels, $w_{ic}$ is the pixel weight, $\hat{x}_i$ represent the projection of the pixel $x_i$ into the world space, and $b_c$ represents the width of each component.

In the actual reasoning process, only the initial human model needs to be established according to these shape parameters, and then the driving can be completed through the attitude parameters. The actual driving equation can be determined by forward dynamics:

$$\vec{S}_i = \vec{R}_{parent(i)} * \left( l_i \vec{O}_i \right) + \vec{S}_{parent(i)} \tag{3}$$

Where $\vec{S}_i, \vec{R}_{parent(i)}$ is the position of the $i$-th key point and its parent node, and $\vec{O}_i$ is the relative direction vector from the parent node to the current $i$-th key point (a unit vector representing the direction), then the relationship between rotation and extension can be transferred by using the bone length $l_i$ and the rotation matrix $\vec{R}_i$ in the shape parameters.

Video frames are clustered based on the distance between features, and each cluster in the clustering result represents an action. Because this method does not need to consider whether there is obvious interval or transition state between actions, it is also possible to extract features for clustering in general videos, so it can meet most application scenarios. Early video key frame extraction is mainly used in video compression to facilitate video transmission and storage, while the purpose of video key frame extraction in video motion recognition is defined as expressing the richest motion information with minimum data, reducing the calculation amount in recognition and removing the influence of repetitive motion and inconspicuous motion information on recognition.

Deep learning method is an indirect segmentation method, which requires a lot of data for training to achieve good results. However, in the actual segmentation scene, there may not be enough data for training, and at the same time, it is needed to label the action type of each video frame, which is extremely workload and very difficult to implement. The proposed method is a direct segmentation method, which can achieve excellent segmentation accuracy without complicated labeling and training, and can be well applied to actual segmentation requirements. The operation principle of human motion model is shown in Figure 2.

All kinds of movements in the dance video are composed of simple human movements that are constantly changing. Although in most cases, it is impossible to accurately identify what kind of dance movements are based on a single human movement, there are many movements with high similarity in the human movements that make up a complex movement. In dance movements, there are movements performed by different actors at different speeds, and even the same actor cannot guarantee that the length of the time series formed between the same movements is exactly the same. Therefore, this article will use dynamic time warping to calculate the similarity between each action sequence after segmentation in a video frame sequence. If the trends of two action sequences are the same or similar, the regular distance is small, thus judging that the two action sequences belong to the same or similar actions.

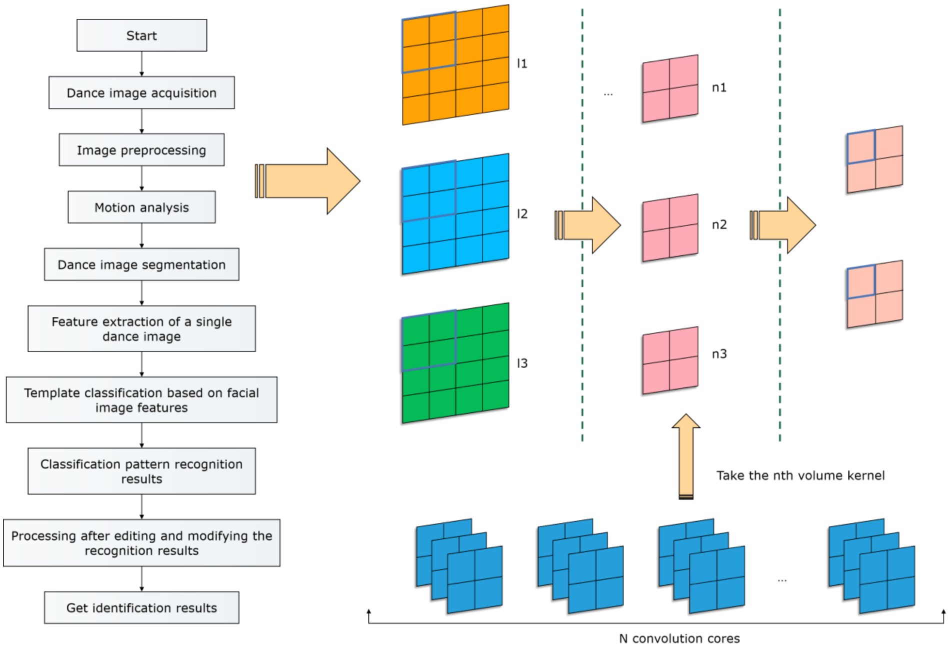
Start
Dance image acquisition
Image preprocessing
Motion analysis
Dance image segmentation
Feature extraction of a single dance image
Template classification based on facial image features
Classification pattern recognition results
Processing after editing and modifying the recognition results
Get identification results
I1
I2
I3
...
n1
n2
n3
Take the nth volume kernel
...
N convolution cores

**Figure 2**: Operation principle of human motion model.

In the skeleton topology diagram, the strength of the connection is represented by the thickness of the line segment between nodes, and the dotted line represents the connection other than the physical connection. In the process of drawing, for the newly-built connections between nodes, several connections with the highest connection strength are selected to add, while for the existing connections, the thickness of the line segment is modified according to the change of connection strength. In this article, the original image is scaled according to a certain law, and a multi-scale space is obtained, as shown in Figure 3.

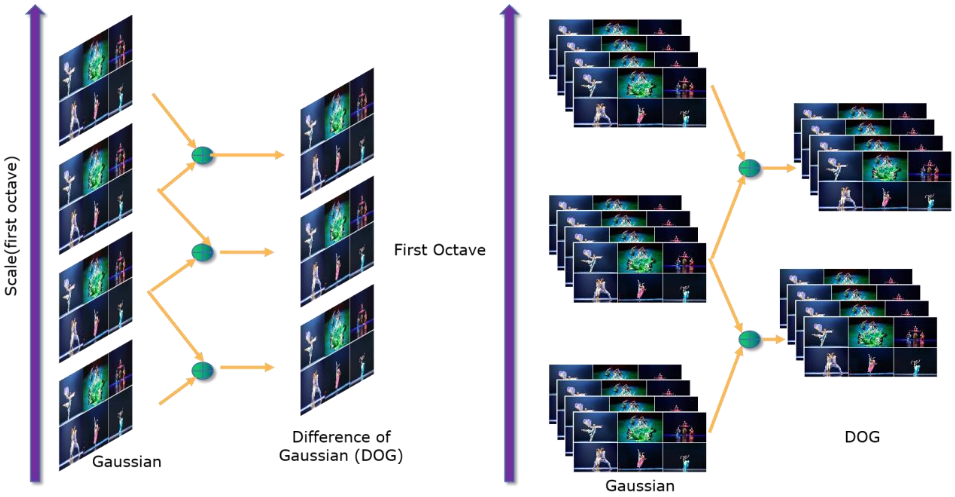

**Figure 3**: Schematic diagram of scale space.

With the introduction of nonlocal and self-attention blocks, due to their ability to build channels and spatial attention modules, nonlocal mechanisms can be used to capture space-time and channel information. Due to the variety and complexity of dance movements, the whole torso of some dancers does not move, only the position of joint points of limbs changes. Some dance movements, such as running and jumping, will cause great displacement of the human torso, and more often, the limbs are doing various movements while the human torso moves. Due to the variety and complexity of dance movements, the whole torso of some dancers does not move, only the position of joint points of limbs changes. Some dance movements, such as running and jumping, will cause great displacement of the human torso, and more often, the limbs are doing various movements while the human torso moves. It is obviously not accurate to represent the movement of the whole human body only by the displacement of one or several joints of the human body. Noise usually obeys Gaussian distribution, so the influence of noise on differential operation can be overcome by setting threshold.

In video action recognition, invalid actions include not only transitional actions, but also human actions with too high similarity in a video. In order to reduce the complexity of calculation, it is only needed to analyze the human action sequence in a class of effective actions when recognizing actions. After selecting the best segment from the similar action sequence segments judged by the similarity of action sequences, other segments will be used as new data to enrich the database in the training process, and they will be filtered in the testing process, and only the best action sequence will be used for action recognition. The 2D gradient amplitude and direction of each pixel are specifically defined as:

$$m_2 D(x,y)=\sqrt{L_x^2+L_y^2} \tag{4}$$

$$\theta(x,y)=\tan^{-1}\left(\frac{L_y}{L_x}\right) \tag{5}$$

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

$$m_3 D(x,y,t)=\sqrt{L_x^2+L_y^2+L_t^2} \tag{6}$$

$$\theta D(x,y,t)=\tan^{-1}\frac{L_y}{L_x} \tag{7}$$

$$\phi(x,y,t)=\tan^{-1}\frac{L_t}{\sqrt{L_x^2+L_y^2}} \tag{8}$$

Because $\sqrt{L_x^2+L_y^2}$ is positive, there is always $\phi\left(-\frac{\pi}{2},\frac{\pi}{2}\right)$, and each corner is represented by a unique $(\theta,\phi)$ pair.

It is obviously not accurate to represent the movement of the whole human body only by the displacement of one or several joints of the human body. Noise usually obeys Gaussian distribution, so the influence of noise on differential operation can be overcome by setting threshold. According to the analysis of the characteristics of human sparse point cloud, outliers in human sparse point cloud are judged by the average distance, rather than being removed directly by median filtering. If we judge by distance, we can use statistical filtering algorithm to judge by calculating the average distance between each data point and its neighboring points. If the result conforms to Gaussian distribution, its shape is determined by mean and standard deviation. If the average distance exceeds the specified range, the point is judged as a noise point and needs to be deleted.

## 4 RESULT ANALYSIS AND DISCUSSION

Dance CAI system provides users with a sound dance action model and environment simulation model base. The system broadens the imagination of dance choreographers and endows dance creation and performance with agility. With the help of simulation plan, the perspectives of each session and seat are set respectively, so as to maximize the artistic effect and spatial effect of dance. For the modeling of bone sequence in time series, time chart is composed of continuous frames connected by corresponding joint points in time dimension. Because the time window is predefined, there is no flexibility in identifying different actions. Therefore, a time convolution module including long, medium and short scales is proposed to obtain the time characteristics in different time domain receptive fields. In the experiment, the error test of motion recognition between the improved algorithm and the traditional algorithm is shown in Figure 4. The comparison of the accuracy of motion recognition of the algorithm is shown in Figure 5.

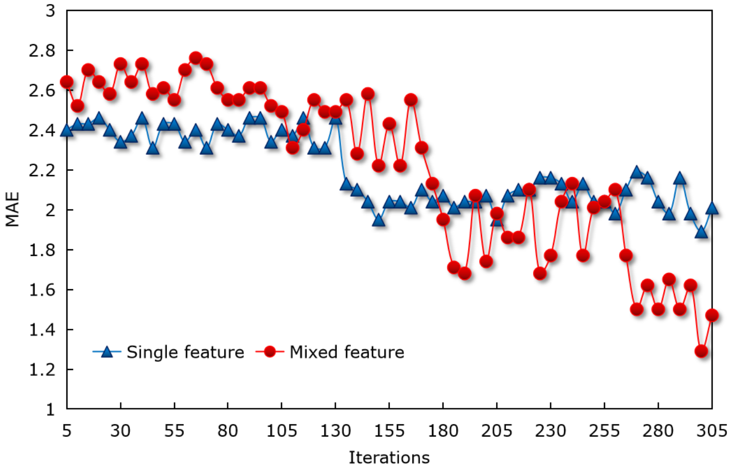

**Figure 4**: Algorithm error test.

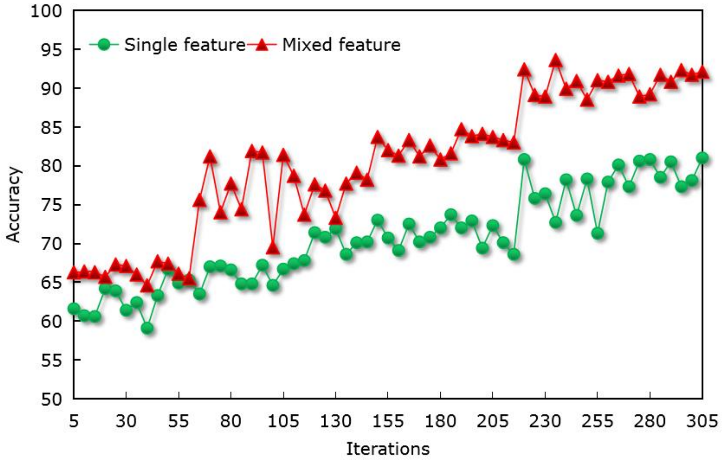

**Figure 5**: Algorithm accuracy test.

The accuracy of this method is higher (more than 95%) and the error is lower (about 15% lower than the traditional method). In time series, it is very important to deal with the time dimension. An action may contain multiple stages, and different sub-stages have different effects on the final recognition results. Based on this, this article introduces the time attention module, and solves the problem of which frames should be focused on in time series by giving corresponding weights to different frames, so that the network can better extract the time series characteristics. In the prototype system, it is needed to calculate the system load in combination with the running state of the main hardware resources in the system. Figure 6 shows the results of the system load test.

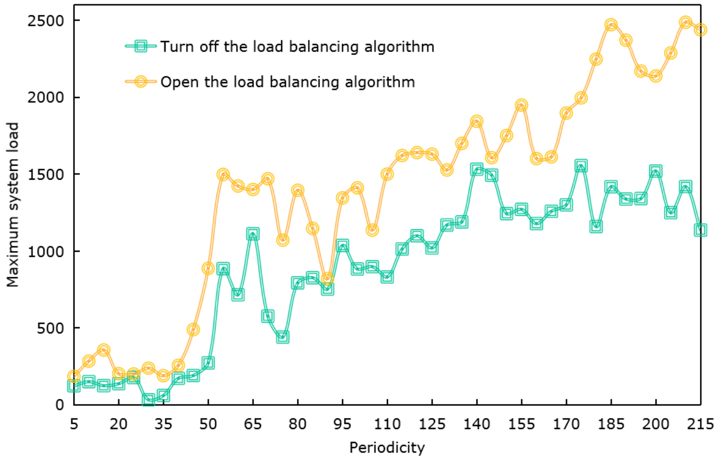

**Figure 6**: Load situation of the system.

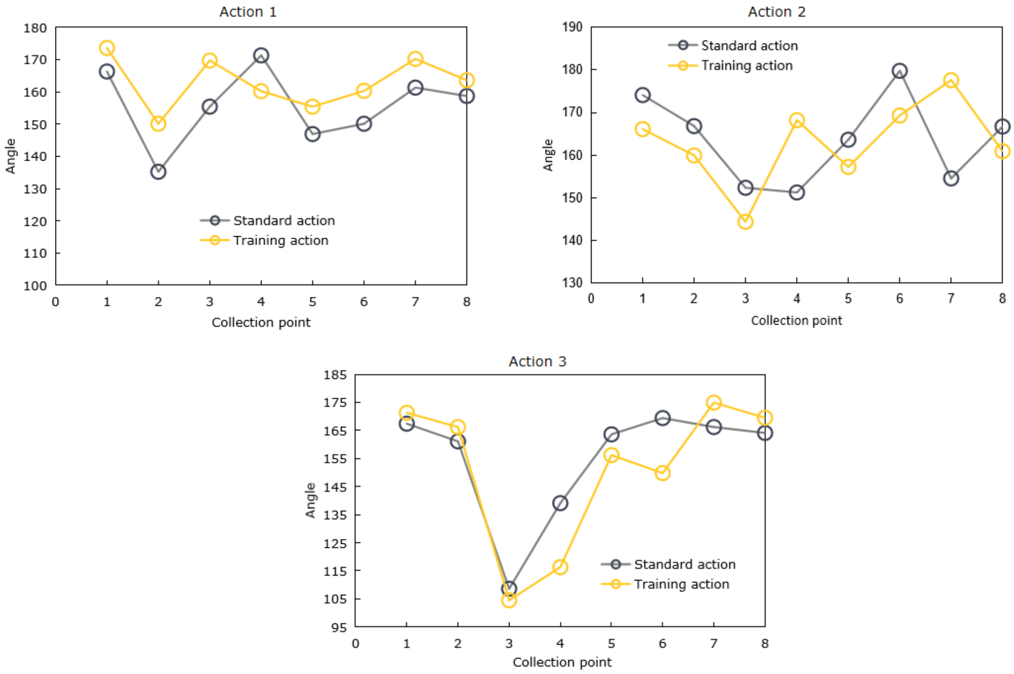

**Figure 7**: Comparison of key joint angles with standard action curves.

After the load balancing mechanism is turned off, the system load will increase rapidly with the user's visit. After adopting the load balancing scheduling algorithm, although the load of the server will increase due to the increase of the number of visiting customers, it is generally lower than that of turning off the load balancing algorithm.

By collecting the data of the dancer's three hand movements, and calculating the included angle of each joint in the decomposition process, as shown in Figure 7. As can be seen from Figure 7, the included angle between the right wrist, the right elbow and the right shoulder is large, and the included angle between the waist, the right knee and the right ankle is also large. In the next exercise, your right arm should be straighter, and your body should be slightly tilted forward and your right leg should be slightly bent.

Dance CAI system broadens the imagination of choreographers and gives dance teaching and performance flexibility. The choreographers can design the simulation plan according to the times and the audience's perspective to ensure that the presented artistic image can meet the expectations. The system can also be used to simulate the stage lighting, scenery and effects, efficiently complete the design of the lighting needed for a specific scene, adjust the proportion of stage space and ensure the artistic effect.

## 5 CONCLUSIONS

Applying CAI function to the practical workflow of dance teaching is an inevitable trend of dance teaching gradually moving towards modern design mode. Through the rapid identification of dance movements, related music and dance fragments can be quickly obtained. In this article, the application of motion recognition technology in dance CAI system is studied, and the dance motion recognition algorithm based on human model reconstruction technology is realized. In the actual teaching work, the morphological and molecular changes of each limb joint of the designated dancing performance object when making a certain dance action should be taken as the key reference object for setting the parameters of the design scheme. Only the setting of various parameters and indicators in the virtual design environment can better follow the characteristics of the actual dancer's skeleton, and the dance action teaching scheme has its inherent feasibility in the practical performance. The simulation shows that the accuracy of this method is bound to exceed 95%, and the error is reduced by about 15% compared with the traditional method. For the 3D reconstruction method of human model designed in this article, although the key steps in 3D reconstruction have been improved, the accuracy can be further improved, and the reconstruction time and efficiency can also be improved.

## 6 ACKNOWLEDGEMENT

This work was supported by 2022 Special Project for Cultural Research of Henan Xing Cultural Project: Research on the Integration and Communication of Tea culture in Henan (No. 2022XWH166).

*Yafang Zhao*, https://orcid.org/0009-0005-6283-9009
*Hongpeng Yang*, https://orcid.org/0009-0002-5238-3421

## REFERENCES

[1] Adams, S.-O.; Onwadi, R.-U.: An empirical comparison of computer-assisted instruction and field trip instructional methods on teaching of basic science and technology curriculum in Nigeria, International Journal of Social Sciences and Educational Studies, 7(4), 2020, 22-35. https://doi.org/10.23918/ijsses.v7i4p22

[2] Angermann, C.; Schwab, M.; Haltmeier, M.; Laubichler, C.; Jónsson, S.: Unsupervised single-shot depth estimation using perceptual reconstruction, Machine Vision and Applications, 34(5), 2023, 82. https://doi.org/10.1007/s00138-023-01410-5

[3] Guo, C.; Zuo, X.; Wang, S.; Liu, X.; Zou, S.; Gong, M.; Cheng, L.: Action2video: generating videos of human 3d actions, International Journal of Computer Vision, 130(2), 2022, 285-315. https://doi.org/10.48550/arXiv.2111.06925

[4] Liu, R.; Shao, Q.; Wang, S.; Ru, C.; Balkcom, D.; Zhou, X.: Reconstructing human joint motion with computational fabrics, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 3(1), 2019, 1-26. https://doi.org/10.1145/3314406

[5] Ma, G.; Hao, Z.; Wu, X.; Wang, X.: An optimal electrical impedance tomography drive pattern for human-computer interaction applications, IEEE Transactions on Biomedical Circuits and Systems, 14(3), 2020, 402-411. https://doi.org/10.1109/TBCAS.2020.2967785

[6] Padilla, P.-L.; Mericli, A.-F.; Largo, R.-D.; Garvey, P. B.: Computer-aided design and manufacturing versus conventional surgical planning for head and neck reconstruction: a systematic review and meta-analysis, Plastic and Reconstructive Surgery, 148(1), 2021, 183-192. https://doi.org/10.1097/PRS.0000000000008085

[7] Shi, M.; Aberman, K.; Aristidou, A.; Komura, T.; Lischinski, D.; Cohen-Or, D.; Chen, B.: Motionet: 3d human motion reconstruction from monocular video with skeleton consistency, ACM Transactions on Graphics (TOG), 40(1), 2020, 1-15. https://doi.org/10.1145/3407659

[8] Song, Y.; Jin, T.; Dai, Y.; Song, Y.; Zhou, X.: Through-wall human pose reconstruction via UWB MIMO radar and 3D CNN, Remote Sensing, 13(2), 2021, 241. https://doi.org/10.3390/rs13020241

[9] Su, Z.; Xu, L.; Zhong, D.; Li, Z.; Deng, F.; Quan, S.; Fang, L.: Robustfusion: Robust volumetric performance reconstruction under human-object interactions from monocular rgbd stream, IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(5), 2022, 6196-6213. https://doi.org/10.1109/TPAMI.2022.3215746

[10] Sun, Y.-T.; Fu, Q.-C.; Jiang, Y.-R.; Liu, Z.; Lai, Y.-K.; Fu, H.; Gao, L.: Human motion transfer with 3d constraints and detail enhancement, IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(4), 2022, 4682-4693. https://doi.org/10.1109/TPAMI.2022.3201904

[11] Vo, M.; Sheikh, Y.; Narasimhan, S.-G.: Spatiotemporal bundle adjustment for dynamic 3d human reconstruction in the wild, IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(2), 2020, 1066-1080. https://doi.org/10.1109/TPAMI.2020.3012429

[12] Wan, Z.; Zhang, B.; Chen, D.; Zhang, P.; Wen, F.; Liao, J.: Old photo restoration via deep latent space translation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(2), 2022, 2071-2087. https://doi.org/10.1109/TPAMI.2022.3163183

[13] Wang, Y.; Liu, Q.: Effects of game-based teaching on primary students' dance learning: the application of the personal active choreographer, International Journal of Game-Based Learning, 10(1), 2020, 19-36. https://doi.org/10.4018/IJGBL.2020010102

[14] Xu, L.; Su, Z.; Han, L.; Yu, T.; Liu, Y.; Fang, L.: Unstructuredfusion: real-time 4D geometry and texture reconstruction using commercial rgbd cameras, IEEE transactions on pattern analysis and machine intelligence, 42(10), 2019, 2508-2522. https://doi.org/10.1109/TPAMI.2019.2915229

[15] Yu, R.; Park, H.; Lee, J.: Human dynamics from monocular video with dynamic camera movements, ACM Transactions on Graphics (TOG), 40(6), 2021, 1-14. https://doi.org/10.1145/3478513.3480504

[16] Zhai, X.: Dance movement recognition based on feature expression and attribute mining, Complexity, 2021(21), 2021, 1-12. https://doi.org/10.1155/2021/9935900

[17] Zhao, F.; Jiang, Y.; Yao, K.; Zhang, J.; Wang, L.; Dai, H.; Yu, J.: Human performance modeling and rendering via neural animated mesh, ACM Transactions on Graphics (TOG), 41(6), 2022, 1-17. https://doi.org/10.1145/3550454.3555451

[18] Zhao, X.; Ye, S.: Space reconstruction of audiovisual media based on artificial intelligence and virtual reality, Journal of Intelligent & Fuzzy Systems, 40(4), 2021, 7285-7296. https://doi.org/10.3233/JIFS-189554

[19] Zheng, Z.; Ma, H.; Yan, W.; Liu, H.; Yang, Z.: Training data selection and optimal sensor placement for deep-learning-based sparse inertial sensor human posture reconstruction, Entropy, 23(5), 2021, 588. https://doi.org/10.3390/e23050588