International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 6657-6668

https://doi.org/10.70517/ijhsa463572

A Study on Enhancing the Creative Effect of Dance Theater Using Artificial Intelligence Image Processing Methods

Qian Lian[1,*]
[1] Department of Dance, Taiyuan Normal University, Taiyuan, Shanxi, 030000, China
Corresponding authors: (e-mail: itsme_lianq@163.com).

Abstract Traditional dance drama creation mainly relies on choreographers' experience, which has problems such as limited innovativeness and low efficiency of movement design. This study explores the use of artificial intelligence image processing methods to enhance the creation of dance drama. A binocular stereo vision depth motion camera is used to capture the dance performer's movements, combined with wavelet transform for multi-scale filtering of dance video images, background subtraction is applied to extract dance movement features, and a dance automatic generation system based on Seq2Seq model is constructed. The system takes audio features as input, and realizes the matching generation of dance and music through the generator, discriminator and self-encoder working together. The experimental results show that the joint angle changes of the dance movements generated by this method are smoother, and the average value of knee pitch angle is 1.315 rad, which is only 1.15% different from the initial set value of 1.30 rad; in the objective evaluation index, the method achieves a beat coverage rate of 26.4%, and a diversity index of 42.1, which is significantly improved compared with the traditional method; in the subjective evaluation, 15 subjects showed a positive response to this method's realism and musical style consistency of the generated dances were highly evaluated by 15 subjects. The study proves that the artificial intelligence image processing method can effectively improve the stability, authenticity and innovation of dance movements in the creation of dance dramas, providing a new technical path for the creation of modern dance dramas.

Index Terms Artificial Intelligence, Image Processing, Dance Theater Creation, Binocular Stereo Vision, Seq2Seq Model, Dance Movement Generation

I. Introduction

Dance is one of the forms of artistic expression and the mother of all arts, and performers can present unique dance movements through body language to show artistic beauty [1], [2]. And dance comes from people's desire for life, and physical performance as a support to show the art. With the rapid development of the economy, the art of dance has also gained a variety of development, and dance drama is one of the most prominent forms of dance, can fully express the artistic ability of dance, and push its value to a higher level, which in turn allows the viewer to find more far-reaching thoughts and feelings, and lays the foundation for the development of the arts [3]-[5]. In the traditional creation of dance drama, mainly relying on the experience of dancers and directors, the choreography is inefficient, and the stage design, due to insufficient funds and space perception, leads to a lower effect of dance drama [6]-[8]. In addition, modern themes are relatively scarce, and they are too straightforward in expressing the content and recounting the events, but too full in the account. As a result, the stage performance presented is more single, and the focus is still on body language, while the overall presentation of the dance drama is not strong, and the dance drama response is not significant [9]-[12].

With the continuous innovation of digital technology, the performing arts are moving towards digitalization, and the art of dance is created with the assistance of technology. In the field of computer vision, artificial intelligence image processing methods represented by generative adversarial network, image transformation, motion capture, etc., can realize stage scene generation through semantic segmentation, style migration through dance kind feature extraction, dance action enhancement through dynamic dance gesture capture, and stage effect improvement through scene reconstruction to build virtual stage [13]-[16].

Dance drama, as a comprehensive art form, requires a high degree of precision and expressiveness in movement design. Traditional dance drama creation mainly relies on the choreographer's personal experience and aesthetic judgment, and the creation process is often time-consuming, inefficient, and limited by the choreographer's knowledge boundaries, facing bottlenecks in innovation. In recent years, the rapid development of artificial intelligence and computer vision technology has provided new possibilities for the creation of dance theater. In particular, the breakthrough of deep learning technology in image recognition, motion capture and sequence generation has laid a technical foundation for the automatic analysis and generation of dance movements. At present, scholars at home and abroad have carried out a series of explorations of applying artificial intelligence technology to dance creation. Some researchers use motion capture technology to analyze the characteristics of dancers' movements and build movement databases; some other scholars try to generate dance sequences based on deep learning models, but most of the research is limited to the recognition or generation of simple movements, which is difficult to meet the comprehensive requirements of dance drama creation on the accuracy of movements, artistry and music matching. In terms of dance video processing, the existing methods mostly use monocular vision system, resulting in insufficient acquisition of spatial information, unable to accurately express the three-dimensional action characteristics of dance; in terms of dance action generation, the existing models mostly use a single neural network structure, lack of in-depth consideration of the artistic characteristics of the dance, and the generated action often lacks stability and realism. In addition, the matching problem between dance and music has not been fully emphasized in previous studies, resulting in the generation of dance that is not coordinated with the rhythm and emotional expression of music. Based on the analysis of existing research, there is an urgent need to establish a set of systematic artificial intelligence image processing methods to realize the accurate capture, feature extraction and intelligent generation of dance movements, and then enhance the creation effect of dance drama. In this study, binocular stereo vision technology is adopted to improve the accuracy of dance action capture, wavelet transform is used for image filtering, combined with background subtraction to realize dance action feature extraction, and a dance automatic generation system is constructed based on the Seq2Seq model. The system realizes the high degree of matching between dance and music through the collaborative work of generator, discriminator and self-encoder. In particular, the discriminator design introduces a dance-music matching evaluation mechanism to ensure that the generated dance movements are not only realistic, but also highly compatible with the music style and rhythm. Through experiments, we verify the advantages of this research method in the quality, stability and music matching of dance movement generation, and provide a new technical path and creative ideas for the creation of modern dance drama.

II. Dance drama action generation based on artificial intelligence image processing

II. A. Typical Dance Visual Image Acquisition

A binocular stereo vision depth motion vision image camera is used to collect video images of dance movements of a typical dance theater performer to record the whole process of the dancer's training, and the accuracy of the collected data is improved by constructing a binocular stereo vision model.

In order to derive the position of the dancer at the spatial point, we set up the camera plane coordinate system formula as follows:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} dx & 0 & -\alpha_0 dx \\ 0 & dy & -\beta_0 dy \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ 1 \end{bmatrix} \tag{1}$$

where, pixel columns are described by β ; physical dimensions on the x and y axes are described by dx and dy , respectively; pixel rows are described by α ; and coordinates of the origin in the image are described by (α_0, β_0) .

The in-space Q -point in camera coordinates are described by $Q_X(X_X, Y_X, Z_X)$, the Q -point world coordinates are described by $Q_W(X_W, Y_W, Z_W)$, and the relationship expression between them is:

$$\begin{bmatrix} X_X \\ Y_X \\ Z_X \end{bmatrix} = \begin{bmatrix} R & T \\ \vec{0} & 1 \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \tag{2}$$

Among them, the translation vector is described by T; the 3×3 rotation matrix is described by R . Based on the stereo imaging theory, the image coordinate system and camera coordinate relationship are derived, as shown in equation (3):

$$Z_X = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \tag{3}$$

where the camera focal length is described by f. Deriving Eq. (1) and Eq. (2) into Eq. (3), the relationship between the world coordinates and pixel coordinates of the Q point can be found with the following expression:

$$Z_X = \begin{bmatrix} c_x & 0 & \alpha_0 & 0 \\ 0 & c_y & \beta_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ \vec{0}' & 1 \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \tag{4}$$

where the internal parameters of the camera are described by c_x, c_y, respectively.

II. B. Methods for automatic tracking of dance movement trajectories

II. B. 1) Dance Video Image Filtering Processing

Wavelet transform [17] is mainly utilized for multi-scale filtering of dance video images by using the translation and expansion functions, and the corresponding frequency bands are different for different values of resolution, and the following wavelet properties will be used to filter the background as well as the high-frequency noise of the dance video images.

Wavelet analysis is used to denoise the dance video image, the dance video image is subjected to wavelet transform, and the transformed wavelet coefficients are thresholded and wavelet inverse transformed to obtain the denoised dance image. In the use of wavelet filtering method for dance video image transformation before the need to prioritize the image for logarithmic transformation, and then the noise is converted, and ultimately obtain the denoised dance video image.

In the process of dance video image denoising, the selection of the threshold value is very important, if the threshold value is determined, it means that the wavelet coefficients smaller than the threshold value are caused by noise. Separately, different thresholds are used on different levels of sub-images, for the target, the background is in the lower frequency band, the wavelet is decomposed, the lowest layer is used as the lower frequency band, and the background around the target is larger than the corresponding band, the wavelet coefficients of the layer are set to 0, and at the same time the noise in the dance video image is effectively filtered.

II. B. 2) Data Capture Based Feature Extraction for Dance Movements

Applying the background subtraction method to detect the target foreground in the dance video, the basic operation steps are as follows: a background model is constructed in priority, and the relevant parameters in the model are learned at the same time, and after learning to obtain the parameters, it is necessary to compare the dance video image corresponding to the current frame with the background model. Set the pixel value in t time period as $x_{(i)}$, then the probability of occurrence of $x_{(i)}$ is:

$$q(x_{(i)}) = \sum_{i=1}^{k} \alpha_{(i,j)} \cdot \beta\left(x_{(i)}, \delta_{(i)}, \phi_{(i)}, \mu_{(i)}\right) \tag{5}$$

where $\alpha_{(i,j)}$ represents the Gaussian distribution weights; $\beta(x_{(i)}, \delta_{(i)}, \phi_{(i)}, \mu_{(i)})$ represents the probability density function with the expression:

$$\beta\left(x_{(i)}, \delta_{(i)}, \phi_{(i)}, \mu_{(i)}\right) = \left(\frac{1}{\sqrt{2\pi}\left|\delta_{(i)}\right|}\right)^{\frac{1}{2}} \cdot k \tag{6}$$

where $\delta_{(i)}$ represents the variance.

The individual pixel values of the first frame image of the dance video are all assigned to the k Gaussian distribution mean, and then the variance is assigned to the larger value to ensure the same value is given.

The model is updated to determine whether the pixels match the established k Gaussian distributions through the set constraints, and at the same time the mean, variance, and weights of the Gaussian distributions are updated in the process:

$$\begin{cases} \delta_{(i)} = (1-\phi)\delta_{(i)} + \phi \\ \phi_{(i)} = (1-\beta)\phi_{(i-1)} + \beta x_{(i)} \\ \mu_{(i)} = (1-\beta)(\phi_{(i-1)} - x_{(i-1)})^2 \end{cases} \tag{7}$$

where $\phi_{(i)}$ represents the mean of the Gaussian distribution and $\mu_{(i)}$ represents the weights of the Gaussian distribution.

When the model finishes training, it is necessary to prioritize according to the Gaussian distribution, which in turn forms the background, as shown in Equation (8):

$$C = \arg\min\left\{\sum_{n}^{i=1}\beta > T\right\} \tag{8}$$

where T represents the threshold value.

In the process of dance action feature extraction, it is necessary to comprehensively consider the differences between the dance video and action video, and at the same time, combine with the characteristics of the dance itself to extract relevant features in the data set to characterize the dance action.

Set the pixel point of the dance video image as $x(i,j)$, and the grayscale value of $h(i,j,t)$ in t time period, after Δt time period, the target image moves to $(i+\Delta i, j+\Delta j, t+\Delta t)$, and the corresponding grayscale value is $h(i+\Delta i, j+\Delta j, t+\Delta t)$, and since the above two points are the same pixel value while the gray scale value remains constant, the following equation is obtained:

$$h(i+\Delta i, j+\Delta j, t+\Delta t) = h(i,j,t) \tag{9}$$

A first-order Taylor series expansion of Eq. (9) yields the following equation:

$$h(i+\Delta i, j+\Delta j, t+\Delta t) = h(i,j,t) + \frac{\partial h}{\partial i}dx + \frac{\partial h}{\partial j}dy + \frac{\partial h}{\partial t}dt \tag{10}$$

After completing the above operations, feature extraction is performed on the data capture based dance movements.

II. B. 3) Automatic tracking of dance movement trajectories

Over the electronic steady image rendering method for steady state tracking as well as surface texture rendering of dance video images, the inverse weighting coefficients of the dance movement trajectories are obtained through the surface texture structure distribution characteristics of the dance video images, and the rendering function of the dance video images is constructed at the same time, as shown in Eq. (11):

$$g(c_i) = \frac{d_i\lambda_i\sum_{n}^{i=1}\frac{\kappa_{(i,j)}u_{ij}}{|u_{ij}|\beta}}{\sum_{m}\frac{u_{ij}}{|u_{ij}|\beta}} \tag{11}$$

where d_i represents the inverse weighting coefficient; λ_i represents the surface texture structure distribution characteristics.

According to the acquired background component, the RGB decomposition of the smooth region of the dance video image can be performed with Eq:

$$\begin{cases} R(c_i) = I_1 + I_2x + I_3y + I_4z + \sum_{m}^{i=0}\lambda_i U(x,y,z) \\ G(c_i) = J_1 + J_2x + J_3y + J_4z + \sum_{m}^{i=0}\beta_i U(x,y,z) \\ B(c_i) = T_1 + T_2x + T_3y + T_4z + \sum_{m}^{i=0}\alpha_i U(x,y,z) \end{cases} \tag{12}$$

where $R(c_i)$, $G(c_i)$ and $B(c_i)$ represent the RGB decomposition results and $U(x,y,z)$ represents the smoothing region of the dance video.

Through the above analysis, the dance action features are configured multimodally, which in turn improves the trajectory tracking ability of the rotational action of the dance video images.

Harris corner detection and template feature matching are effectively combined to automatically track the trajectory of dance action. Priority is given to obtaining a system of linear equations for grid segmentation of dance video images:

$$\begin{pmatrix} a & b & c \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} (I+\pi-B) & \\ \vdots & \\ 0 & x(i,j) \end{pmatrix} \times (i,j,t) \tag{13}$$

where B represents the active contour information of the dance video image.

II. C.Automatic Dance Generation Based on Seq2Seq Modeling

II. C. 1) General Model of Dance Generation

The purpose of this model construction is to generate a model of dance with music by combining music and dance movements based on the principle of deep learning [18], and the specific model structure is shown in Figure 1. Firstly, audio features and dance gesture features are extracted; then the dance gesture is predicted by audio feature input, and the predicted dance gesture is compared by MSE loss function; the audio features are reconstructed by self-coding, and finally, the real dance gesture and the predicted dance gesture are fed into the GAN Loss for training, and finally, the training-matched dance is output.

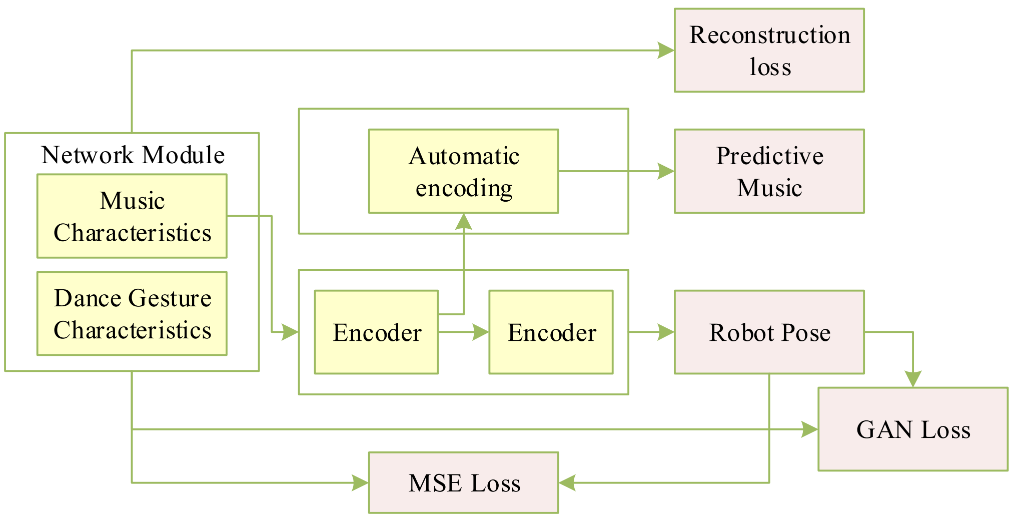

Figure 1: Dance generation model network structure

II. C. 2) Feature extraction

1. Audio Feature Extraction

Audio feature extraction mainly adopts the way of rhythm and tempo feature extraction. According to the characteristics of rhythm, the higher frequency Mel frequency cepstrum is often used to extract the 24-dimensional and 8-dimensional Tempogram features, and it is used as the audio melody; for rhythmic feature extraction, it is necessary to obtain the frequency frame position of the whole piece of music, and the parameter values of the audio frames in the beat, so as to realize the rhythmic classification and extraction according to the values of each specific parameter of the audio.

2. Action feature extraction

For the extraction of dance movement features, OpenPose is used to extract the human body movement features to obtain the 18 key points of the human body.

Through the extraction of audio features and action features, the following sets of feature data are obtained:

(1) Audio features: $M_i = \langle m_i^1, m_i^2, \cdots, m_i^{32} \rangle$, i.e. the feature dimension of each audio is 32.

(2) Beat feature: $B_i = \langle b_i^1, b_i^2, b_i^3 \rangle$, i.e., the size of the beat dimension of a video frame section is 3.

(3) Pose features: $p_i = \langle P_i^1, P_i^2, \cdots, P_i^{36} \rangle$, i.e., the size of the pose dimension in a video frame is 36.

(4) Audio data: $M = \{M_1, M_2, \cdots M_n\}$, which indicates that there are a total of n frames of audio data, and each frame is represented as M_n.

(5) Dance gesture data: $p = \{p_1, p_2, \cdots, p_n\}$, denoting a total of n frames of dance gesture, with each frame denoted as P_n.

II. C. 3) Dance generation model module design

1. Generator design

In order to make the dance generation better, the Seq2Seq model [19] is chosen to design the generator, which includes: encoder and decoder. However, the Seq2Seq model can not represent all the semantic information, will be based on the model, adding the attention mechanism, so as to become a context vector that can be summarized in different moments of different semantic information. The generator adopts the tacotron2 model, which is very suitable for solving the task of sequence generation.

2. Discriminator Design

In order to make the dance gesture and music matching more accurate, the matching result is evaluated by adding a discriminator. If the match is made, the dance gesture is continued to be generated, and vice versa, a new dance gesture is generated. The specific discriminator is mainly carried out through the loss function of equation (14).

$$\begin{aligned} & \mathrm{L}_{GAN}(G,D)=\mathrm{E}_{(P,M)}\left[\log D((P,M))\right] \\ & +\frac{1}{2}\mathrm{E}_{M}\left[\log\left(1-D\left(G(M),M\right)\right)+\log\left(1-\left((W,M)\right)\right)\right] \end{aligned} \tag{14}$$

In Eq. (14), G denotes the generator; M denotes the music; $G(M)$ denotes the generation of dance gestures; $D\left(G(M),M\right)$ denotes the size of the probability that the dance matches with the music, and if the value is close to 1, it indicates that the two match more. If the value is in the range of 0, it indicates that the probability of matching dance and music is small.

After generating the dance gesture, the discriminator is used to judge the match between the dance and the audio, so as to confirm whether the audio is consistent with the real gesture. If the pose vector is denoted as $p=\{P_1,P_2,P_2,\cdots,P_n\}$, the difference between the front and back frames is denoted as $R=\{P_1-P_2,P_2-P_3,\cdots,P_{n-1}-P_n\}$, and the audio features are represented as $M=\{M_1,M_2,\cdots M_n\}$. All of them are inputted into the discriminator, and then calculated by equation (14).

3. Self-encoder module design

The role of the self-encoder is to reconstruct the audio features. If M_i denotes audio features and B_i denotes beat features, the match between music and dance can be improved by inputting the two features into the encoder and then encoding and decoding them. Specific reconstruction:

$$f_i=Encoder\left(Concat\left(M_i,B_i\right)\right) \tag{15}$$

$$M_i^{\sim}=Decoder\left(f_i\right) \tag{16}$$

where $M_i^{\sim}$ denotes the reconstructed audio and f denotes the extracted low-dimensional audio features. $Concat$ denotes the splicing processing of the extracted parameters, and $Encoder$ and $Decoder$ denote the neural networks to be learned.

4. Realistic Choreography

Relying solely on the generator to transform audio features into dance movements may prevent a single dance gesture from being well presented. To solve this problem, an improved Pix2Pix algorithm is used to transform a single dance, resulting in a robotic dance action.

III. Dance drama creation generation and effect analysis

III. A. Experimental preparation

Conduct comparison experiments, remembering the method of this paper as experimental group A, and the two traditional methods as experimental group B and experimental group C. Compare the three groups of experiments to see whether the trajectory of the bipedal joint angle change is smooth or not when the robot generates the dance movement automatically.

Select a dance robot as the experimental object, use matlab to write a program to read the music signal, the music signal after sub-frame processing is shown in Figure 2. It can be obtained that the frequency range of the base tone of the music used in the experiment is 105HZ~260HZ, the frequency range of the music signal is 21HZ~21kHZ, the peak size range is 2.7ms~10.1ms, and the maximum absolute value error of the control matlab is 0.8. The sampling frequency range is set to be 12kHZ, 26kHZ, and 49kHZ, and the window length range is taken to be 35ms~55ms. 55ms.

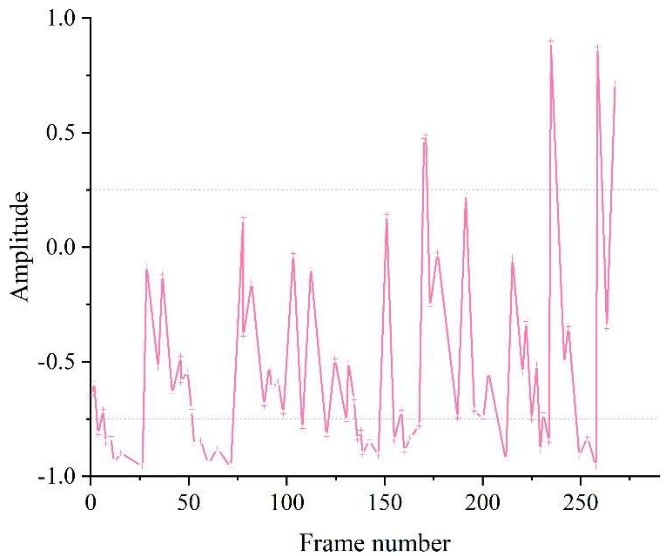

Figure 2: The music signal after the frequency processing

Calculate the feature parameters of the music signal shown in the above figure, take 3100 training set samples randomly from the initial data set, choreograph the dance movements according to the feature parameters, calculate the dance movement parameters, and choreograph the bipedal dance movements for the music. Finally set the corresponding bipedal dance movement of the music as cross walk, and the bipedal movement pattern is shown in Fig. 3.

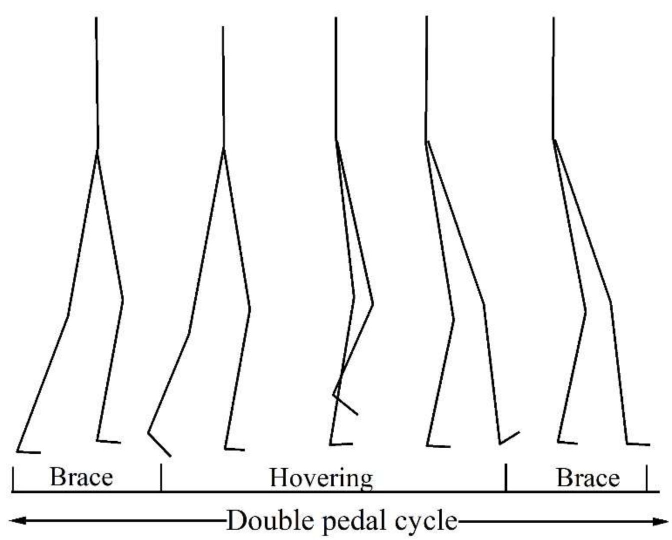

Figure 3: Set the dance movement

Three groups of experiments use the same group of dance movement library, the test can be obtained three groups of experimental bipedal dance movement cycle time is about 2.9s, the dance movement are more stable. The third dance movement, trunk squatting and knee forward flexion, was selected to measure the joint angle under this movement.

III. B. Experimental results

Firstly, the knee pitch angle was measured, and the initial value of this angle setting was 1.30rad, and the measurement results of the three groups of experiments are shown in Table 1. In the table, the variation of knee pitch angle in experimental group A is not large, the extreme values are 1.30rad and 1.32rad, and the average value of pitch angle is 1.315rad, while the maximum values of group B and group C are 1.36rad and 1.37rad, and the minimum values are 1.30rad and 1.22rad, which is a large difference from the initial value.

Table 1: The results of the first set of experiments

The number of cycles of double pedal	A group of knee knees	B group of knee knees	C group of knee knees
1	1.32	1.36	1.37
2	1.31	1.35	1.29
3	1.32	1.35	1.28
4	1.32	1.36	1.35
5	1.31	1.35	1.34
6	1.32	1.36	1.35
7	1.30	1.29	1.37
8	1.31	1.30	1.34
9	1.32	1.36	1.32
10	1.32	1.35	1.22

The hip joint pitch angle was measured, and the initial value of this angle setting was 0.32 rad. The measurement results of the three groups of experiments are shown in Table 2. The extreme values of experimental group A were 0.28rad and 0.34rad, and the average value of pitch angle was 0.323rad, while the maximum values of group B and group C were 0.36rad and 0.38rad, and the minimum values were 0.26rad and 0.25rad, respectively, and the experimental group A was closer to the initial setting angle.

Table 2: The results of the second group of experiments

The number of cycles of double pedal	A group of hip sides	B group of hip sides	C group of hip sides
1	0.33	0.36	0.33
2	0.33	0.34	0.25
3	0.28	0.33	0.36
4	0.33	0.36	0.37
5	0.33	0.32	0.38
6	0.29	0.26	0.29
7	0.33	0.27	0.33
8	0.34	0.27	0.35
9	0.34	0.36	0.36
10	0.33	0.26	0.26

Finally, the pitch angle of the bare joint was measured, and the initial value of this angle setting was -0.23 rad. The measurement results of the three groups of experiments are shown in Table 3. The extreme values of experimental group A are -0.24rad and -0.25rad respectively, and the average value of pitch angle is -0.243rad, while the maximum values of group B and group C are -0.20rad and -0.19rad respectively, and the minimum values are -0.27rad and -0.28rad respectively, compared with experimental groups B and C, the bare joint pitch angle of experimental group A basically did not change, and was very close to the initial value is very close. In summary, this method utilizes artificial intelligence technology to iteratively process the collected parameters, simplifies the process of parameter identification and calculation, and makes the joint angles of the bipedal dance movements more closely match the initial angles, and the changes in the angle trajectories are very smooth, so the stability of the robot dance movements is better than the traditional method.

III. C. Subjective evaluation of dance movement generation

In order to prove the usability and reasonableness of the proposed method, dance movement generation experiments are carried out on the experimental platform, and the results of the dance generation experiments are firstly compared with the results of some 2D dance movement generation tasks to show that the method of this paper is capable of generating dance movements with high quality. Then the experimental results of this paper's method and other methods are compared to prove the effectiveness and robustness of this paper's method.

Five objective evaluation metrics: FID, Beat Coverage, Hit Rate, Diversity and Multimodality are all used in the experiments.Firstly, the comparison experiments between the two-person dance movement generation and some two-dimensional single dance movement generation tasks are conducted, which are used to illustrate the feasibility of this paper's method. The experimental results are shown in Table 4. From the metrics in the table, it can be seen that this paper's method is able to obtain better metrics on the experimental platform, and compared with the other best methods, the beat coverage is improved by 1, and the diversity metrics are improved by 3.5. The beat hit rate is lower than the best method by 12.2, and the multimodality metrics are lower than the best method by 7.95. Only the FID metrics are more disparate from the results of the 2D dance generation task, which are higher than the best method by 16.3, which is due to the fact that there are more frames with missing keypoints due to occlusion in the two-person dance video, which are made up by the linear interpolation algorithm, and there is a gap with the real samples.

Table 3: Experimental results

The number of cycles of double pedal	A group of knee knees	B group of knee knees	C group of knee knees
1	-0.24	-0.27	-0.28
2	-0.25	-0.25	-0.27
3	-0.24	-0.27	-0.26
4	-0.24	-0.26	-0.19
5	-0.24	-0.20	-0.26
6	-0.25	-0.21	-0.21
7	-0.24	-0.26	-0.27
8	-0.24	-0.20	-0.26
9	-0.25	-0.27	-0.28
10	-0.24	-0.24	-0.26

From the objective metrics, our method is able to generate high quality dance movements on the experimental platform, which illustrates the usability and feasibility of the proposed method.

Table 4: Comparison of results of dance action

Methods	FID	Beat Coverage(%)	Hit Rate(%)	Diversity	Multimodality
LSTM	75.4	1.7	0.6	26.4	-
Dancing2Music	21.6	17.5	67.3	32.5	20.5
Huang, et al	7.8	22.6	71.6	38.1	16.4
Ours	24.1	23.6	59.4	41.6	12.55

In order to investigate the performance of the proposed method and other dance movement generation methods in the dance generation task, the comparison experiments of different methods on the experimental platform are conducted, and the experimental results are shown in Table 5. The experimental results show that the method in this paper outperforms the current mainstream 2D dance movement generation methods in four indexes, including FID, beat coverage and hit rate, and diversity, and only slightly lags behind in the multimodality index, indicating that the method in this paper outperforms the traditional methods.

Table 5: Experimental results of different methods

Methods	FID	Beat Coverage(%)	Hit Rate(%)	Diversity	Multimodality
LSTM	161.5	0.5	0.9	9.3	3.8
Dancing2Music	51.4	17.6	50.3	22.4	10.1
Huang, et al	30.5	24.5	54.6	30.4	14.5
Ours	22.6	26.4	60.1	42.1	12.7

In order to further evaluate the quality of the dance movements generated by the method of this paper on the experimental platform, a subjective evaluation experiment was designed. Fifteen general college students were invited as subjects and asked to evaluate the dance movements after a brief training. Two dance movements generated by the two methods were randomly given during the evaluation, and the subjects were asked to choose the dance that they thought was more realistic without considering the music, and at the same time to choose the one that was more consistent with the style of the music with considering the music. In the experiment we limited the amount of time subjects had to make each choice, and we also ensured that each two-by-two pairing of dance combinations was evaluated by everyone, and that a sufficient number of evaluations were made to minimize the interference caused by subjective factors. Figures 4 and 5 show the results of the subjective evaluation experiments, pink: the method in this paper, red: LSTM, green: Dancing2 Music, gray: Huang, etal, yellow: Real Dances.

The results show that compared with other dance movement generation methods, this paper's method is superior in terms of realism and consistency with the music style of the generated dances.The LSTM method accumulates a large amount of errors when generating longer dance sequences, which leads to its low evaluation in terms of realism. The method in this paper effectively improves the realism of the generated dances and also obtains better user ratings. In terms of consistency with the music style, the use of advanced music features allows for the generation of dances that are more consistent with the music style.

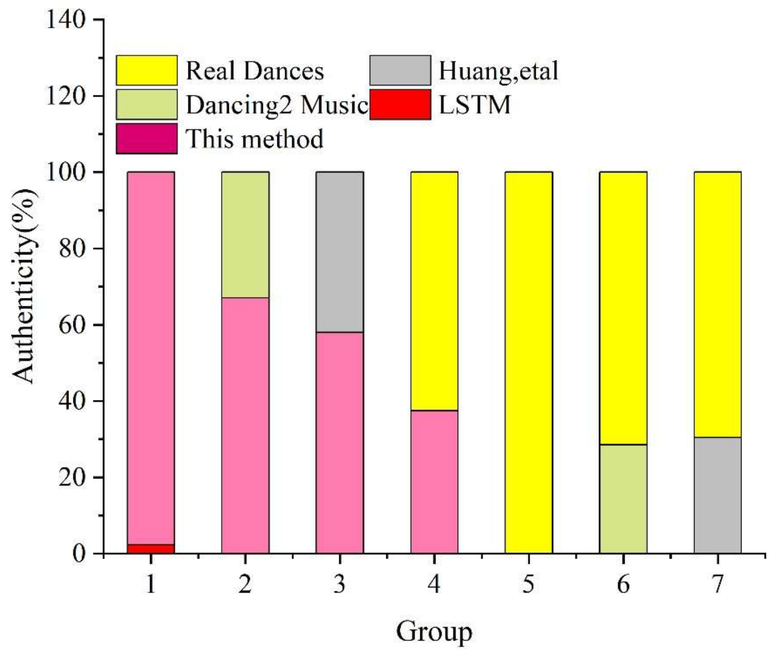

Figure 4: Subjective evaluation of dance

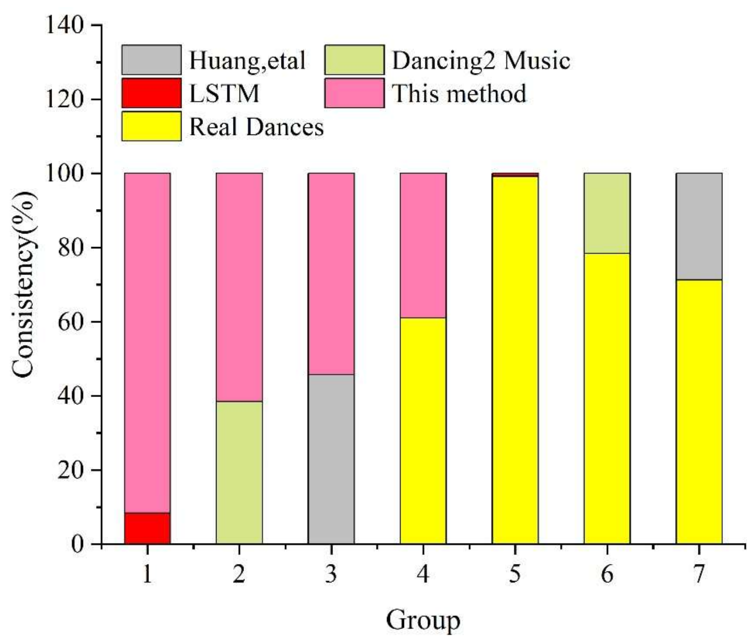

Figure 5: The consistent subjective evaluation results of dance and music style

III. D. Example analysis and experimental results

III. D. 1) Application cases

A team of professional contemporary dancers partnered with a technology company to create and evaluate new dance movements using generative adversarial networks and motion capture technology. The dance team had several meetings with the technology partner early in the project to clarify the goal of using AI technology: to generate innovative dance movements and evaluate their performance quality in real time. Once the goal was set, both parties identified the technical resources and types of data needed.

The team collected data on the choreography dancers performing standard dance movements through motion capture technology. This data included information such as the coordinates of the dancers' key body points, movement speed and rhythm. The collected data was preprocessed, including denoising, normalization, and time-series analysis, so that it could be used to train AI models. The technical team trained the generative adversarial network using the preprocessed data. In this process, the generator tries to create new dance moves while the discriminator evaluates the naturalness and innovativeness of these moves. After several rounds of iterations, the AI model was able to generate dance moves that met the artistic requirements of dance and were highly innovative.

The generated dance movements were delivered to the choreographer's interface and the choreography team choreographed the dance based on the movement suggestions provided by the AI model. During this process, the team also utilizes AI for real-time performance quality assessment. The AI system analyzes the dancers' performance data and provides feedback on technical execution and expressiveness, helping the dancers to adjust their movements for optimal performance. After the newly choreographed dances were performed in-house, the audience provided feedback via an electronic questionnaire. The results showed a significant improvement in the creativity of the new movements and the overall expressiveness of the dance. The dance team used the audience feedback and AI evaluation results to further optimize the choreography.

III. D. 2) Comparative analysis of experiments

The same dance team was selected for the experiment to create two choreographic passages with similar themes using the AI technique and traditional methods in this paper. Each dance passage was performed by the same group of dancers to ensure consistency of performers' skills.The AI choreographed passages utilized previously described generative dance movements and were assessed for movement quality through a long and short-term memory network. Traditional choreography relies solely on the experience and creativity of the choreographer for movement design.

Performances of both choreographic approaches were videotaped, and the videos were subsequently scored by independent professional judges and audience members. Judging criteria included innovativeness, technical execution, artistic expression and audience response. In addition, feedback from dancers' experiences of the choreographic process was collected to assess the actionability of the two methods in dance practice and dancer satisfaction.

AI choreography significantly outperformed traditional methods in terms of innovativeness scores. Due to the AI technology's ability to analyze and synthesize large datasets, the generated dance movements showed novel elements and combinations, whereas the traditional method was limited by the choreographer's personal experience and known skills. Both methods scored similarly on technical execution, indicating that the movements generated by AI technology are technically feasible and can be effectively executed by the dancers. The audience and judges gave high marks to the artistic expression of the AI dance segments, especially in terms of matching the movements to the music and conveying the emotion of the overall performance.

IV. Conclusion

Artificial intelligence image processing technology provides a revolutionary way for the creation of dance drama. Experimental validation shows that the binocular stereo vision technology can capture the dancer's spatial position information with high precision, the combination of wavelet transform and background subtraction method effectively extracts the dance movement features, and the dance generation system based on the Seq2Seq model successfully realizes the coordinated matching of the dance and music. Comparison experiments show that the joint angles of the dance movements generated by this method change smoothly, and the average value of hip pitch angle 0.323rad differs from the initial set value of 0.32rad by only 0.94%, which is much better than the traditional method. Among the objective evaluation indexes, the beat coverage rate of the method reaches 26.4%, which is 25.9 percentage points higher than that of the traditional LSTM method; the diversity index reaches 42.1, which is 87.9% higher than that of the Dancing2Music method; and the beat hit rate is 60.1%, which is 59.2 percentage points higher than that of the LSTM method. The subjective evaluation results were even more significant, with 15 subjects evaluating the realism and musical style consistency of the dances generated by the method significantly higher than other methods. The application case analysis confirms that the method can effectively improve the movement innovativeness and artistic expression in the creation of actual dance dramas. Overall, the artificial intelligence image processing method breaks through the limitations of traditional dance theater creation and expands the space of dance art expression through data-driven dance movement generation, opening up a new path for the integration of dance theater art and technology.

References

[1] Peng, Z. (2023). The Current Situation and Development of Dance Creation in the New Media Era. Art and Performance Letters, 2023, 4 (5): 32, 38.

[2] Rae, J. (2018). Drawing the Language of Dance. PAJ: A Journal of Performance and Art, 40(2), 42-45.

[3] Zhou, Y. (2022, November). Analysis on the Application of Technical Skills in the Dunhuang Dance Drama “Silk Road”. In 8th International Conference on Arts, Design and Contemporary Education (ICADCE 2022) (pp. 49-52). Athena Publishing.

[4] Haowei, Z. H. A. N. G. (2024). A Study on the Metaphorical Features of Chinese and Korean National Ballets in the 20th Century: Taking the National Ballets ‘Raise the Red Lantern’and ‘Shen Qing’as Examples. Journal of Sinological Studies (JSS), 1(1), 1-7.

[5] Ren, B. (2025). Commercial Strategy Analysis of Chinese Dance Drama--Taking The Works of Zhou Liya And Han Zhen as An Example. Advances in Education, Humanities and Social Science Research, 13(1), 135-135.

[6] Carey, K., Moran, A., & Rooney, B. (2019). Learning choreography: An investigation of motor imagery, attentional effort, and expertise in modern dance. Frontiers in Psychology, 10, 422.

[7] Zhang, M., & You, C. (2024, July). Exploring the Fusion of Eastern and Western Cultures in Dance Drama Creation—A Case Study of the Children's Ballet Drama “Grassland Hero Sisters”. In 2024 5th International Conference on Mental Health, Education and Human Development (MHEHD 2024) (pp. 427-432). Atlantis Press.

[8] Lian, Q. (2025). Optimization of Image Recognition Technology for Dance Theatre Creation and Evaluation of Its Effectiveness. J. COMBIN. MATH. COMBIN. COMPUT, 127, 1653-1665.

[9] Pogrebnyak, M., Nemkovich, O., & Hromchenko, V. (2024). Author’s dance theatre of the XX–the beg. of the XXI centuries:‘The idea of freedom in the theory of drama’and its influence on shape. Amazonia Investiga, 13(80), 176-188.

[10] Elliott, F. A. (2023). Dancing and the Stance: Mapping a Creative Practice in African Dance-Drama. Dance Research, 41(2), 194-206.

[11] Padmanabha, K. V., & Kumar, S. (2020). Health communication through folk media: a study based on Yakshagana–A South Indian dance drama on Covid-19. Mass Communicator: International Journal of Communication Studies, 14(3), 20-26.

[12] Jin, X. (2024). Piercing the Veil of God, Revealing Human Nature—Analyzing the Dance Drama “The Orphan of Zhao” from the Perspective of “People”. Open Access Library Journal, 11(12), 1-8.

[13] Tang, H., & Sebe, N. (2021). Total generate: Cycle in cycle generative adversarial networks for generating human faces, hands, bodies, and natural scenes. IEEE Transactions on Multimedia, 24, 2963-2974.

[14] Yin, W., Yin, H., Baraka, K., Kragic, D., & Björkman, M. (2023). Multimodal dance style transfer. Machine Vision and Applications, 34(4), 48.

[15] Pang, Y., & Niu, Y. (2023). Dance video motion recognition based on computer vision and image processing. Applied Artificial Intelligence, 37(1), 2226962.

[16] Ying, F., & Bo, Z. (2021). Building virtual scene construction and environmental impact analysis based on image processing. Scientific programming, 2021(1), 9979862.

[17] Morteza Saadatmorad,Ramazan Ali Jafari Talookolaei,Samir Khatir,Nicholas Fantuzzi & Thanh Cuong Le. (2025). Frugal wavelet transform for damage detection of laminated composite beams. Computers and Structures,314,107765-107765.

[18] Shiva Paudel,Tami Brown Brandl,Gary Rohrer & Sudhendu Raj Sharma. (2025). Deep learning algorithms to identify individual finishing pigs using 3D data. Biosystems Engineering,255,104143-104143.

[19] Xiao Yaoqiang,Li Yi,Yuan Jin,Guo Songrui,Xiao Yi & Li Zhiyong. (2021). History-based attention in Seq2Seq model for multi-label text classification. Knowledge-Based Systems,224,