Revija za Kriminalistiko in Kriminologijo
Journal of Criminal Investigation and Criminology
REVIJA ZA KRIMINALISTIKO IN KRIMINOLOGIJO,
ISSN :0034-690X,
STEFANOVA ULICA 2, LJUBLJANA,
SLOVENIA, 1501
Volume 76-Issue 2-(2025)
OPEN ACCESS

Designing AI-Driven Dance Choreography: Motion Capture and Generation Protocols

Yeqin Peng

[1]*The Li Jinhui Music School, Hunan University of Science and Technology, Xiangtan ,411201, China*

[a]*Email: pengyeqin519@163.com*

Abstract: In the era of rapid technological advancement, the emergence of artificial intelligence (AI) technology has reshaped the landscape of dance creation, proposing a new direction for artistic practice through human-machine collaboration. By integrating motion capture technology with generative algorithms and protocols, this study breaks through the limitations imposed by human physiological boundaries, creating a dynamic choreography system that enables virtual-reality interaction. This system extends the possibilities of movement language into other spatial dimensions, After reviewing the current state of AI-driven dance choreography research, this paper delves into two algorithms: a dance notation generation algorithm based on spatial feature fusion and a dance notation generation algorithm based on multi-temporal modeling. Based on the experimental results of these two algorithms, a platform architecture for the automatic generation and display of Latin dance notation is constructed, providing technical support for the documentation and preservation of dance art and culture in the new era.

Keywords: artificial intelligence; motion capture; spatial feature fusion; multi-temporal modeling

1. Introduction

With the steady development of the social economy, collaborative creation models have integrated dance choreography with artificial intelligence. By using motion capture technology to obtain human movement trajectories, artificial intelligence algorithms generate action combinations that comply with human movement norms but exceed physiological limits. These are then selected and optimized by professional dance choreographers, forming a coordinated model that combines human creativity and machine computation [1]. In this process, virtual-reality

integrated dance performances can break through the spatial and temporal constraints of traditional stages, effectively utilizing dynamic tracking technology and holographic projection technology to enable real-time interaction between dancers and virtual characters [2]. Some scholars have found that in real-time interactive scenarios, using wearable technology devices, artificial intelligence can accurately perceive the speed of a dancer's limb movements during improvisational dance, thereby forming a motion trigger feedback loop. This technology does not replace human creativity but transforms traditional dance choreography into dynamic co-creation, allowing technical algorithms and dance art to continuously collide in virtual interaction and spark more artistic inspiration [3]. Currently, several scholars have utilized AI technology to innovate traditional dance, enhancing the efficiency and quality of dance creation. For example, digital models were created based on the grass-weaving movements of intangible cultural heritage inheritors, with fluid dynamics principles serving as the technical foundation, and using technical algorithms to simulate and analyze the amplitude of the dragon's body movements and the dancers' movements multiple times. This ultimately resulted in a design that retains the original "three nods of the dragon's head" ancient ritual while incorporating dance movements compatible with modern ergonomics. This makes it easier for novice dancers to precisely control the 10-meter-long grass dragon and reduces the difficulty of inheriting traditional craftsmanship [4]. This demonstrates that research on AI-driven dance choreography technology has garnered significant attention, and related technical algorithms are a focal point in the practical arts field.

2. Dance Notation Generation Algorithm Based on Spatial Feature Fusion

2.1 Human Skeletal Features

An AI-driven dance choreography system requires motion capture data to accurately record human movement trajectories within a spatial environment [5]. This study adopts Euler angle format to store motion capture data. Since this format cannot directly represent changes in position within space, a conversion operation must be performed on the raw data [6]. The motion data recorded in Euler angle format is transformed into motion data in the world coordinate system, which represents the

position of a moving object relative to the reference frame. Any orientation can be formed using the rotation angles of three basic relative coordinate systems.

In an inertial coordinate system, a rotation operation must first be performed. Assuming the rotation matrix is R, the initial offset positions of all non-root nodes in the BVH file are (x_0, y_0, x_0), and the position of the parent node of this node is (x_p, y_p, x_p). After one rotation, the position of node L relative to its parent node P is (x_1, y_1, x_1), so the new node L is located at (x_1, y_1, x_1). This can be calculated using the following formula:

$$(x_1,y_1,z_1)=R\square(x_0,y_0,z_0)$$

In practical research, it was found that BVH files are stored in ZXY rotation order. Therefore, the rotation matrix can be calculated using the following formula:

$$R=R_Z R_X R_Y$$

In the above formula, $R_Z$is obtained by rotating around the Z-axis, $R_X$is obtained by rotating around the X-axis, and $R_Y$is obtained by rotating around the Y-axis. The following formula can be used:

$$R_Z(-r)=\begin{bmatrix}\cos(-r) & \sin(-r) & 0\\ -\sin(r) & \cos(-r) & 0\\ 0 & 0 & 1\end{bmatrix}=\begin{bmatrix}\cos(r) & -\sin(r) & 0\\ \sin(r) & \cos(r) & 0\\ 0 & 0 & 1\end{bmatrix}$$

$$R_X(-p)=\begin{bmatrix}1 & 0 & 0\\ 0 & \cos(-p) & \sin(-p)\\ 0 & -\sin(-p) & \cos(-p)\end{bmatrix}=\begin{bmatrix}1 & 0 & 0\\ 0 & \cos(p) & -\sin(p)\\ 0 & \sin(p) & \cos(p)\end{bmatrix}$$

$$R_Y(-y)=\begin{bmatrix}\cos(-y) & 0 & -\sin(-y)\\ 0 & 1 & 0\\ \sin(-y) & 0 & \cos(-y)\end{bmatrix}=\begin{bmatrix}\cos(y) & 0 & \sin(y)\\ 0 & 1 & 0\\ -\sin(y) & 0 & \cos(y)\end{bmatrix}$$

Based on the hierarchical analysis of the human joint node tree structure, the coordinates of the current joint nodes are not only influenced by the rotation of their corresponding parent nodes but also by the collective rotation of all other parent nodes [7].

Assuming that the rotation matrices of all nodes in L are R_1, R_2,..., R_m, the formula for calculating the displacement of a node is as follows:

$$(x_1,y_1,z_1)=(R_m\square(R_{m-1}(\square(R_1\square(x_0,y_0,z_0)))))$$

Combining the above analysis, the offset of any node relative to the nodes in this

coordinate system can be obtained, thereby accurately determining the position coordinates of each node.

Assuming that the offset quantities of the human joint nodes and their preceding parent nodes are L_0, $L_{(1),}$…, $L_{(r),}$where $L_r$represents the node, then the offset quantities of these nodes relative to the node are O_0, $O_{1,}$…, O_{r-1}.

In the world coordinate system, assuming that node P_{root}represents the coordinates of all nodes in the human skeleton, the node coordinates in this coordinate system are as follows:

$$P=P_{root}+O_{r-1}+\square+O_1+O_0$$

After the above calculations and analysis, the conversion of motion capture data stored in Euler angle form to spatial human motion data represented by three-dimensional coordinates in the world coordinate system can be expressed as follows:

$$(x_1,y_1,z_1)=(R_m\square(R_{m-1}(\square(R_1\square(x_0,y_0,z_0)))))$$

2.2 Action Recognition

Regarding the recognition and generation of dance patterns, the focus is on studying changes in limb movements. Therefore, this study employs the Long Short-Term Memory (LSTM) network algorithm, which has unique control gates capable of effectively capturing contextual relationships in continuous data. Based on this, limb movement recognition is performed, demonstrating superior performance in handling dynamic continuous data and high efficiency in dance pattern generation [8]. Recurrent Neural Networks (RNNs) are neural networks well-suited for processing continuous data. However, ordinary RNNs struggle to capture relevant information when the distance between contextual relationships is either very long or very short. To address this issue, researchers proposed the Long Short-Term Memory (LSTM) network algorithm to resolve problems such as gradient explosion or vanishing gradients [9]. The actual internal structure is shown in Figure 1 below:

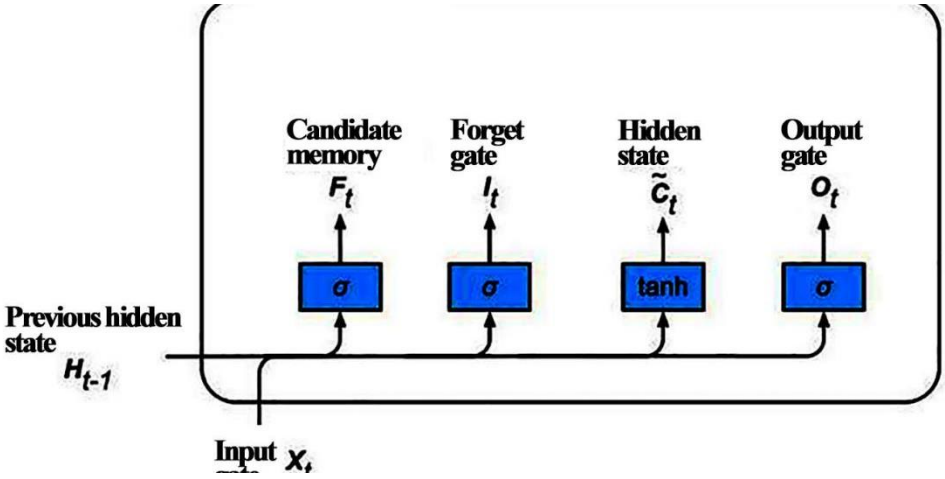

Figure 1 Structure diagram of the LSTM algorithm

At each time step ^{t}, an LSTM unit has a memory cell with the current state c^{t}. Three sigmoid gates—the input gate i^{t}, the forget gate f^{t}, and the output gate o^{t}—control the reading or modification of information in the core cell c^{t}[10]. Generally speaking, the information update process of a Long Short-Term Memory Network Algorithm unit at time t is as follows:

$$i^{t}=\sigma(W_{xi}x^{t}+W_{hi}h^{t-1}+W_{ci}c^{t-1}+b_{i})f^{t}=\sigma\left(W_{xf}x^{t}+W_{hf}h^{t-1}+W_{cf}c^{t-1}+b_{f}\right)o^{t}$$

$$=\sigma(W_{xo}x^{t}+W_{ho}h^{t-1}+W_{co}c^{t-1}+b_{o})c^{t}$$

$$=f^{t}c^{t-1}+i^{t}\tanh(W_{xc}x^{t}+W_{hc}h^{t-1}+b_{c})h^{t}=o^{t}\tanh(c^{t})$$

In the above formula, σrepresents the sigmoid function, x^{t}represents the input vector at time t, h^{t}represents the output of the value information used at time t and previous times, all W matrices represent the connection weights between two nodes, and b_{i}, b_{f}, b_{o}, and b_{c}refer to the bias vectors.

2.3 Experimental Results Comparison

This paper investigates the application of the long short-term memory network algorithm structure to form a human value action recognition method [11]. The neural network is used to learn input information, select and retain valuable content, and to avoid overfitting, a dropout layer is added after the first network layer. The last layer is a fully connected layer, and the activation function uses the widely used SoftMax function to predict the probability values of each category required for human action recognition. The actual experimental structure design is shown in Table 1:

Table 1 Architectural Design Results of the Long Short-Term Memory Network Algorithm

Layer (type)	Output Shape	Param #
lstm_1 (LSTM)	(None, 40, 256)	275456
lstm_2 (LSTM)	(None, 128)	197120
dense_1 (Dense)	(None, 25)	3225

In experimental studies, the accuracy of human action recognition is obtained by dividing the total number of correctly recognized human lower limb actions by the number of human lower limb actions recognized by the group. By altering the input feature types of the action recognition network, the impact of different feature representations and feature combinations on the final results is compared and analyzed [12]. Among these, the input features are set in two forms: one uses only Joint features, and the other uses only Line features. Different feature combinations combine these two types of features. The action recognition network is constructed using a long-short term memory network. By utilizing two feature fusion methods to construct a spatial feature fusion action recognition method, dynamic temporal and spatial information can be carried, directly presenting posture and direction changes during movement, and showcasing the topological structure of the human body and the correlations between bones [13]. The specific experimental results are shown in Table 2.3:

Table 2 Lower-limb (left leg)

Method	**Features**	**Dataset 1 Accuracy (%)**	**Dataset 2 Accuracy (%)**
LSTM	Joint	58.5	60.8
LSTM	Line	93.75	94.6
LSTM	Joint+Line	96.41	96.76

Table 3 Lower-limb (right leg)

Method	**Features**	**Dataset 1 Accuracy (%)**	**Dataset 2 Accuracy (%)**

LSTM	Joint	59.4	60.1
LSTM	Line	93.2	94.1
LSTM	Joint+Line	95.97	96.78

Based on the analysis in the table above, it can be concluded that the feature fusion recognition method achieves higher actual accuracy, with a performance improvement of 2.5%, making it suitable for the feature representation part of dance choreography generation research.

Compared with existing methods, the recognition accuracy of the proposed method has increased by 2%, with specific results shown in Table 4:

Method	**Dataset 1 (%) — Left**	**Dataset 1 (%) — Right**	**Dataset 2 (%) — Left**	**Dataset 2 (%) — Right**
Template+DTW	80.36	77.17	62.56	59.38
HMM	90.05	90.82	90.31	89.25
RNN	89.90	91.13	90.46	92.11
GRU	92.86	94.55	91.81	92.00
LSTM	94.23	94.62	93.43	95.02
Spatial-Feature-Fusion	96.41	95.97	96.76	96.78

Table 5Experimental results compared with existing methods

Method	**Dataset 1 (%) — Left Leg**	**Dataset 1 (%) — Right Leg**	**Dataset 2 (%) — Left Leg**	**Dataset 2 (%) — Right Leg**
Template+DTW	80.36	77.17	62.56	59.38
HMM	90.05	90.82	90.31	89.25
RNN	89.90	91.13	90.46	92.11
GRU	92.86	94.55	91.81	92.00
LSTM	94.23	94.62	93.43	95.02
Spatial-Feature-Fusion	96.41	95.97	96.76	96.78

3. Dance Score Generation Algorithm Based on Multi-Temporal Modeling

3.1 Bidirectional Gated Recurrent Unit

The gated recurrent unit (GRU) is a variant of recurrent neural networks that first combines the hidden layer state passed from the previous neural node with the input data of the current node. Using the Sigmoid activation function, the data is mapped to the range [0,1], thereby obtaining the gate states of the reset gate and update gate [14].

The specific formula is as follows:

$$r=\sigma(W_{h^{t-1}}^{r}X^{t})z=\sigma(W_{h^{t-1}}^{r}X^{t})$$

In the above formula, the gate r controls the reset of data, and the gate z controls the update of data. Based on this, the reset gate is used to obtain the reset data, as shown in the following formula:

$$h^{t-1'}=h^{t-1}Or$$

The hidden layer state $h^{t-1'}$ of the previous node after reset is then combined with the current input data X^t using a tanh activation function to map the output data of the hidden state h' to the range [-1,1], thereby obtaining the current state of the hidden layer. The specific formula is as follows:

$$h^{'}=\tanh(W_{h^{t-1'}}^{X^{t}})$$

After the recurrent gating unit resets the hidden state, it enters a new phase, where the network selectively remembers and forgets some information. The updated gating z can be obtained using the following formula:

$$h^{t}=zOh^{'}+(1-z)Oh^{t-1}$$

In the above formula, `$(1-z)Oh^{t-1}$` represents the original hidden state that is selectively forgotten, and `$z\square h^{'}$` represents the h' that includes the existing node information after selective memory b.

3.2 Generation Algorithm

Based on the characteristics of dance movements and motion capture data in choreography, there is a strong correlation between consecutive action sequences and different data frames. Therefore, the experimental study employs a multi-temporal modeling approach to generate a dance notation algorithm [15]. During algorithm execution, motion capture data is obtained, and a deep neural network that fuses joint features and line space features is automatically generated to form the network framework. The input refers to the number of frames of motion capture data after feature extraction, where half represents the joint feature sequence and the other half represents the line feature sequence [16]. The relationship between the network input and

hidden layers is analyzed using the following formula:

$$h_t^1 = W_{X_1 h} X_t^1 + W_{hh}^1 h_{t-1}^1 + b_h h_t^2 = W_{X_1 h} X_t^2 + W_{hh}^2 h_{t-1}^2 + b_h$$

The backpropagation process is as follows:

$$h_t^{1'} = W_{X_1 h} \quad {}' X_t^1 + W_{hh}^{1'} h_{t+1}^1 + b_h' h_t^{2'} = W_{X_2 h} \quad {}' X_t^2 + W_{hh}^{2'} h_{t+1}^2 + b_h'$$

3.3 Experimental Results Comparison

In the experiment, the same data set was selected for training and analysis. The initial learning rate was set to 0.0, and the number of network neurons was selected as 128, 256, and 512 for different group studies. To prevent overfitting, the Dropout value was adjusted to 0.25, and the recurrent Dropout value was adjusted to 0.20. First, the two types of features were input into the neural network model for training. After setting the network layers, the number of neurons, and various parameters to ensure the neural network was in optimal condition, the experimental results were evaluated and analyzed. The specific experimental results are shown in Table 6 below:

Table 6 Accuracy (%) on Dataset 1

Method	**Features**	**Epochs**	**Left Leg**	**Right Leg**
Bi-GRU	Joint	70	58.5	59.4
Bi-GRU	Line	20	93.7	93.2
Bi-GRU	Spatial Feature Fusion	35	97.3	97.0
Method	**Features**	**Epochs**	**Left Leg**	**Right Leg**
Bi-GRU	Joint	70	60.8	60.1
Bi-GRU	Line	20	94.6	94.1
Bi-GRU	Spatial Feature Fusion	35	97.0	97.4

Based on the analysis in Table 6, it can be observed that after integrating spatial features into the network input, the network converged gradually under the condition that the iteration parameters reached 35, and the network performance reached its optimal state, with an average recognition accuracy of 97%. This indicates that the multi-temporal dance score generation algorithm proposed in this study not only converges quickly but also achieves higher network recognition accuracy, which is of great significance for the research on automatic dance score generation in the new era.

4. Application of the Automatic Latin Dance Score Generation

Demonstration Platform

Based on the motion capture and generation protocol algorithm proposed in the previous section, this study, after mastering dance movement theory, motion segmentation, and related elements, designed a specific architecture for a digital platform for the generation and display of dynamic art,The overall architecture design is divided into two parts: the generation module and the multi-dimensional display module [17]. The former can take three-dimensional human motion capture data as input content to generate dance notation corresponding to the movements, while the latter integrates dance notation, multi-angle videos, capture data, and animations into a unified format, providing dynamic art with diverse recording methods from multiple perspectives. This offers an engaging platform for learning and appreciating dynamic art in the new era, making it more convenient for people to learn and inherit [18]. Although current research on automatic generation is still in its early stages, and there are few explorations of the integration of dance and computer technology, the overall architectural design research in this paper has achieved significant results, clearly demonstrating the feasibility and effectiveness of AI-driven dance choreography technology [19]. Future platform designs should regard three-dimensional human motion capture data as the foundational basis, organically integrating dance movement theory with computer technology architecture to truly achieve the goal of automatic generation centered on motion capture data and generation protocol algorithms, thereby making outstanding contributions to the digital recording, preservation, and transmission of dynamic art in the new era [20].

Conclusion

In summary, as an important basis for technological reform and innovation in various fields in the new era, the integration and exploration of artificial intelligence in the field of dance art have become increasingly in-depth, with an increasing number of technical algorithms being mastered, truly achieving the organic unity of the natural flow of dance art and safe control. The dance notation generation algorithm based on spatial feature fusion and the dance notation generation algorithm based on multi-temporal modeling proposed in this study have been proven to be safe and effective

through experimental results. Based on this, a Latin dance notation automatic generation and display platform has been constructed, intuitively demonstrating the unique performance of the applied technical algorithms. This provides new insights for cross-disciplinary technological innovation in the new era.

Funding

Project "Interdisciplinary Research on Construction of Graduate Training System for Master of Arts in Music Education" supported by the 2024 Research Project on Graduate Teaching Reform in Hunan Province, Project Number: 2024LXBZZ082.

Project "Innovative Practice Research on 3D Holographic Immersive Teaching in Appreciation of Chinese and Foreign Dance Works" supported by the 2024 Key Project of Undergraduate Teaching Reform Research in Hunan Province, Project Number: 202401000878.

References

[1] Yuanfan Deng. Reflections on AI Choreography and the Subjectivity of Dance Creation [J]. Chinese Dance Studies, 2024(2):114-124.

[2] Junjie Wang. Dance Challenges Technology, Technology Inspires Dance: A Discussion of Merce Cunningham's Computer-Based Choreography Experiments [J]. Dance, 2023(1):96-100.

[3] Yinan Zhang. An Analysis of the Application of Artificial Intelligence Technology in Dance Choreography [J]. Shangwu, 2023(8):76-81.

[4] Tingting Liang. Research on the Application of Virtual Reality Technology in Dance Teaching Practice [J]. Shangwu, 2023(9):96-98.

[5] Zhen Wu. Research on the Protection and Development of Intangible Cultural Heritage Dance Using Motion Capture and Virtual Reality Technology: A Case Study of Tibetan Dance [J]. Dance, 2023(4):90-96.

[6] Chenyuan Ma. Replacing the "Physical Body" with "Digital" in Dance: An Analysis of Holographic Dance in "Dance Monster" [J]. Journal of the Beijing Dance Academy, 2024(1):37-44.

[7] Xiaojian Chen. Design of a Demonstration System Based on Motion Capture Technology and Unity 3D [J]. Automation and Instrumentation, 2023(12):144-147.

[8] Yikun Wang, Xinnan Xu. On the "Mind-Body Realm" in Environmental Dance

Education [J]. Art Education, 2025(3):180-183.

[9] Jin Li. Innovative Development of Dance Creation in the Digital Age [J]. Dance, 2023(3):98-100.

[10] Junda Huang, Yan Yu, Li Shang. A Comparative Study of Reverse Movements in Waltz Dance Among Athletes of Different Skill Levels [J]. Boxing and Martial Arts, 2025(1):88-90.

[11] Siqi Yin, Lin Zhu. Application of Human Body Capture Technology in College Sports Dance Training [J]. 2024(1):132-134.

[12] Jizhou Duan, Tianyu Jiang, Qiang Zhang, et al. Research on Body Coordination Training Based on Musical Rhythm Combined with 3D Visual Evaluation Methods [J]. Mechanical Manufacturing and Automation, 2025, 54(1):197-201.

[13] Gang Zheng, Lijuan Sun, Yan Wei. Research on the Digital Preservation and Inheritance of Gannan Tibetan Dance [J]. Art Market, 2025(2):67-69.

[14] Xing Tao, Fanshan Yu, Yuejie Song, et al. A method for human upper limb motion capture without alignment based on inertial sensors [J]. Flight Control and Detection, 2024, 7(2):28-35.

[15] Zihao Song, Yaxian Xin, Biao Wang, et al. A Method for Segmenting Human Motion Segments Based on Spatio-Temporal Features of Motion Data [J]. Artificial Intelligence and Robotics Research, 2025, 14(1):138-153.

[16] Na Su, Chengcheng Liu. The Disembodied Body and the Reconstructed Belief: The Dual Characteristics of Digital Virtual Character Performance in Metaverse-Themed Films [J]. Film Review, 2023(10):13-17.

[17] Bo Wu. Research on 3D Skeletal Behavior Recognition Based on Motion Flow Self-Attention [J]. Innovation and Application of Science and Technology, 2024, 14(25):72-75.

[18] Jubo Ma, Zhouyi Chen, Jinjian Wu. Surface defect detection of aluminum-based discs using dynamic visual sensors [J]. Acta Automatica Sinica, 2024, 51:1-13.

[19] Zhize Wu, Sheng Chen, Ming Tan, et al. Skeleton Behavior Recognition Based on Cross-Channel Feature Enhanced Graph Convolutional Networks [J]. Pattern Recognition and Artificial Intelligence, 2024, 37(8):703-714.

[20] Yaodong Gu, Shun Wang, Yining Xu. Progress in Biomechanical Research on Competitive Swimming During the Paris Olympic Cycle [J]. Journal of Medical Biomechanics, 2024, 39(4):576-585.