International Journal for Housing Science and Its Applications
Publish August 4, 2025. Volume 46, Issue 3 Pages 3102-3113

https://doi.org/10.70517/ijhsa463260

Analyzing Body Language and Emotional Expression Mechanisms in Dance through Computer Vision

Guojing Tan[1,*] and Jianan Wang[1]
[1] School of Performing Arts, Sichuan University of Media and Communications, Chengdu, Sichuan, 610000, China
Corresponding authors: (e-mail: 13982280543@163.com).

Abstract In this study, a hybrid CNN-BLSTM model integrating biomechanical feature extraction and graph theory is proposed as the core of computer vision technology for the recognition of emotion dynamics in dance body language. Through the Euler angle matrix transformation and de-rotation and de-translation process, biomechanical features such as joint position, bone angle, and human body orientation are quantified, and a force effect parameter system including lightness and smoothness is constructed. The synergistic mechanism between movement learning and emotional expression is explained from the perspective of cognitive psychology by combining movement concept and schema theory. The experiments are based on DanceDB and FolkDance datasets, and the CNN-BLSTM model with deep and shallow feature fusion is used for validation. The results show that the proposed model achieves an average recognition accuracy of 43.48% and 52.37% on DanceDB and FolkDance datasets, respectively, which is an improvement of 7.46%-12.20% compared with single-feature methods such as directional gradient and optical flow direction. In key frame extraction, the multimodal feature fusion strategy reduces the compression rate to 2.96% and improves the accuracy and F1 score to 95.54% and 91.87%, respectively. It is shown that the model significantly enhances the emotion resolution of complex dance movements through joint modeling of spatio-temporal features.

Index Terms computer vision analysis, dance movement, body language, CNN-BLSTM, emotion recognition

I. Introduction

Dance art is a unique art form, which can not only show a variety of moving dance posture, but also express the emotion of the characters through body movements, and convey the meaning that the audience can not imagine [1]-[3]. Dance, as a silent art, uses human body as a medium to convey emotions with the body, so body language and emotional expression are two indispensable parts in the art of dance [4], [5].

Dance body language is more demanding in terms of aesthetic sense compared with the language art of the general public, and it needs to use carefully designed movements, modeling, and scheduling to delicately describe the character image characteristics, emotional changes, and storyline development [6]-[9]. For dancers, the most important way to show their physical qualities and inner emotions is body language, if the expression of body language is not in place, it will affect the perception of the whole work, and can not fully express the connotation that the creator wants to show [10]-[13]. Only when the dance body language is fully compatible with the character image, character psychological activities, storyline, to present a vivid character as well as a vivid story [14]-[16]. One of the most important parts of the art of dance is emotional expression, which is to show the dancer's feelings and inner world through body language [17], [18]. The ultimate goal of emotional expression is to show profound human concern and cultural value. Emotion is one of the most crucial elements in dance performance, which can flow naturally from the dancer's heart or be expressed through body movements [19]-[21]. Emotions can be expressed in different emotional states such as passionate, pain and sorrow, joy and delight [22]. When dancers are able to express these rich emotional states through body movements, the audience will more easily enter into the emotional realm of the dance performance [23]-[25]. Another important aspect of emotional expression is to express the theme, and expressing the theme through dance can make the audience understand and think more deeply about the theme while expressing emotions [26], [27].

Taking computer vision technology as the core, this study proposes a systematic methodological framework, aiming at realizing the accurate recognition of dance emotion through the quantification of movement features, the elaboration of movement schema theory and the construction of deep learning model. Firstly, from the biomechanical point of view, based on the Euler angle principle and motion capture data, biomechanical features such as joint position, bone angle, and human body orientation are quantified, and spatial deviation is eliminated through the process of de-rotation and de-panning. We construct a complete feature system containing force effect parameters such as lightness, smoothness, etc., which provides a data basis for emotion analysis. We also combine the movement concept and schema theory to explore the empowering effect of movement concept group and schema theory on dance aesthetic education from the perspective of cognitive psychology, emphasizing the importance of concept-driven processing in movement learning. The synergistic relationship between movement memory and creative expression is also explained through the schema formation mechanism. On this basis, a hybrid model integrating convolutional neural network CNN and bi-directional long and short-term memory network BLSTM is proposed. CNN extracts the local spatial features of the movements, BLSTM captures the temporal dependencies, and through the fusion of deep and shallow features and the batch normalization technique, the generalization ability of the model is taken into account along with the classification accuracy, which finally realizes the dynamic recognition of the dance emotions.

II. Construction of Dance Emotion Recognition Model Based on Action Feature Extraction and Schema Theory

II. A. Motion Feature Extraction

In order to extract the features such as orientation and force effect in the action, this paper converts the rotation angle in the motion capture data to the position information of the joint nodes based on the Euler angle principle. Firstly, the Euler angles are matrix converted, and the matrices into which the rotation angle data are converted are respectively:

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

$$P = P_X(p) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos p & \sin p \\ 0 & -\sin p & \cos p \end{pmatrix} \tag{2}$$

$$Y = R_Y(y) = \begin{pmatrix} \cos y & 0 & -\sin y \\ 0 & 1 & 0 \\ \sin y & 0 & \cos y \end{pmatrix} \tag{3}$$

where R, P, Y are the rotation matrices of roll angle, pitch angle, yaw angle, respectively, rotating around z, x, y axes, the rotation angles are r, p, y, based on the nature of orthogonal matrices, the following rotation matrix N can be obtained:

$$N(RPY)^{-1} = \begin{pmatrix} \cos r\cos y - \sin r\sin p\sin y & -\sin r\cos p & \cos r\sin y + \sin r\sin p\cos y \\ \sin r\cos y + \cos r\sin p\sin y & \cos r\cos p & \sin r\sin y - \cos r\sin p\cos y \\ -\cos p\sin y & \sin p & \cos p\cos y \end{pmatrix} \tag{4}$$

The rotation matrix calculates the node position from the relative orientation of the requested node to the intermediate node of the root node and obtains the position information in the world coordinate system of that joint through the de-rotation and de-translation process.

The de-rotation process is as follows:

$$J_0 = N_r \cdot N_{r-1} \cdot N_{r-2} \cdot \cdots \cdot N_2 \cdot N_1 \cdot \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} \tag{5}$$

where: (x_0, y_0, z_0) is the initial offset of the node in the BVH data; $N_r, N_{r-1}, N_{r-2}, \cdots, N_1$ represents the rotation matrix of all predecessor nodes of node N to the root node Root.

The de-translation process, which calculates the positional offsets of each predecessor node of node N, which have positional offsets of $O_1, \cdots, O_{r-2}, O_{r-1}, O_r$ with respect to their parent nodes, gives the world coordinates of node J_0 as

$$P = O_1 + O_2 + \cdots + O_r \tag{6}$$

Since the movement of the human body in space causes the deviation of the world coordinates of the skeletal nodes in the x and z directions from the initial acquisition position, it is not possible to carry out spatial orientation analysis, so after obtaining the world coordinates it is necessary to compute the relative coordinates of each coordinate with the Root node Root coordinates (x_r, y_r, z_r) in the BVH data as the reference point, i.e.

$$P_0 = P(x, y, z) - P(x_r, y_r, z_r) \tag{7}$$

After obtaining the coordinates of each joint point, the characteristic parameters describing the biomechanics of human movement, such as joint-to-distance, bone-to-pinch angle, body orientation, spatial orientation, and lightness and smoothness in force effect, need to be calculated from the motion capture data.

The eigenvalues joint-to-distance and bone-to-pinch angle are considered from the perspective of limb structure. Joint pair distance v reflects the movement speed, and the distance between joint pairs is calculated by Euclidean distance, and its characteristic equation is expressed as equation (8). Bone pair pinch angle ∂ indicates the state of bending between neighboring bones, and its characteristic equation is expressed as equation (9), that is

$$v = \sqrt{(x_j - x_{j-1})^2 + (y_j - y_{j-1})^2 + (z_j - z_{j-1})^2} \tag{8}$$

$$\partial = \arccos\left(v_a \cdot v_b \big/ \left(\|v_a\| \times \|v_b\|\right)\right) \tag{9}$$

Eigenvalues body orientation, spatial orientation, and lightness and smoothness in force effects Considering from the perspective of spatial orientation, body orientation n can be expressed as

$$n = s\vec{l} \times s\vec{r} = \left(\beta_1, \beta_2, \beta_3\right) \tag{10}$$

where $\vec{l}$ and $\vec{r}$ are plane normal vectors composed of different skeletal points, respectively.

Spatial orientation s is divided into vertical and horizontal orientation. Vertical orientation divides the human skeleton vertically into {upper, middle, lower}, i.e., it divides the space vertically into upper, middle, and lower 3 components. Horizontal orientation divides the horizontal components into {front, left front, left, left-right, back, right back, right front, in situ}, and each horizontal component is divided in intervals of 22.5°.

The eigenvalue force effect F includes lightness and smoothness. Lightness w represents the integration of the joint point y axis coordinate value curve over the start to end point of the time axis, and this parameter indicates the degree of rise and fall of the movement, and its characteristic equation is

$$w = \int_s^e f(y)dt \tag{11}$$

Fluidity f represents the expansion and contraction of the space through distances and is characterized by the equation

$$f = 2p(p - d_{12})(p - d_{13})(p - d_{23}) / d_{23} \tag{12}$$

where: d_{ij} denotes the distance between joints; $p = \left(d_{12} + d_{13} + d_{23}\right)$.

II. B.Movement Concepts and Schemas Empowering Dance Aesthetic Education

After completing the extraction of physical features of dance movements, the mechanism of their cognitive level expression needs to be further explored. Movement is not only a collection of data, but also an integrated expression of culture, psychology and physiology. This section starts from the concept of movement and schema theory to reveal the intrinsic connection between body language and cognitive logic in dance aesthetic education.

The deep value of dance aesthetic education lies in the fact that it starts from the body, obtains cognition and aesthetics through embodied experience, and then conveys them to other cognitive structures through the brain's processing and transformation of the information to promote the development of the individual's comprehensive qualities, which is also where the fundamental value of aesthetic education lies. As a language, the information conveyed by movement not only reflects emotion and aesthetic culture, but also includes knowledge and understanding of the world, which often fails to reflect its profound content by only cutting from the shape or material. Concepts, on the other hand, can embody condensed and profound connotations, and the concepts and descriptions of movements that are traced back to the genealogy of movements fully demonstrate the unique logic and wisdom of body movements.

The conceptual cluster of descriptions of movement occurrence, process and result proposed by movement spectrum is essentially based on the practice of dance aesthetics education, which penetrates into the mechanism of movement generation and digs out the underlying logic of body expression, so as to enhance the depth and breadth of individual's understanding of movement principles and aesthetics. In this sense, the construction of movement concepts can actually empower dance aesthetic education to promote the development and innovation of individual cognitive models. “Modern cognitive psychology research shows that there are two different types of information processing: data-driven processing and concept-driven processing. Information-driven processing is bottom-up processing, which is initiated by external stimuli and events, and concept-driven processing is top-down processing, which is initiated by higher-level knowledge structures in people's minds. A schema is one of the higher-level knowledge structures in people's minds.” Similarly, the difference between an aesthetic dance practice with movement description as the core pathway compared to a practice based on material and routines is the difference between inside-out and outside-in. The formation of movement schemas through the combination of experience and concepts is an advanced mode of control for movement learning. The formation of movement schema is schematically shown in Figure 1.

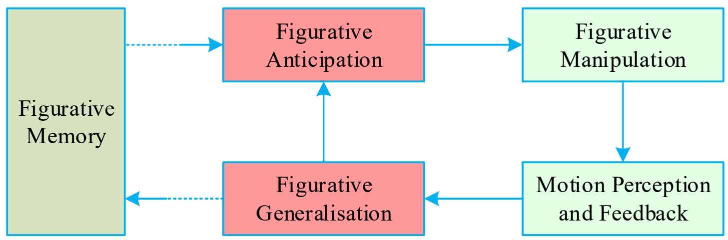

Figure 1: The formation illustration of the action schema

Movement concepts can strengthen understanding and thus quickly establish schemas; and from the perspective of movement memory and lasting educational traces, movement concepts can strengthen the stability of movement schemas. In the practice of movement practice and appreciation analysis, learners master the logic of movement language, and create more movement combinations and aesthetic styles through the understanding and accumulation of concepts; at the same time, the formed schema develops a variety of spectral schemas based on a single movement skill and serving multiple goals through the practice of variations, and the correlation between schemas and the refinement of movement skills are synchronized with each other in terms of empirical and logical constructions. The association between schemas and the refinement of motor skills are synchronized in the construction of experience and logic. Movement is a complex, integrated phenomenon, and its occurrence involves the participation of physiology, psychology, kinesiology, and other disciplines, as well as ethnicity, culture, religion, social development, and other multiple factors. In the modern scientific concept, "people's interpretation of action learning has transitioned from the behaviorist viewpoint of 'stimulation-reinforcement' to the perspective of cognitive processing, and from the discussion of the external characteristics and laws of action learning to the internal mechanism and process of action learning, and the explanation of action learning has become more and more comprehensive and reasonable." Action schema provides a path for individuals to construct mental exercises and observational learning in action learning, and action spectra provide refined, accurate action language and concepts and correspond one to one with symbols, which can form a logical and rigorous knowledge form in the learner's thinking loop and be preserved to form a memory by means of action symbolic language.

II. C.Emotion Recognition Model for Dance Movements Based on CNN-BLSTM with Deep and Shallow Feature Fusion

Based on the foundation of action feature quantization and schema theory, it becomes crucial to transform the abstract cognitive process into a computable emotion recognition model. In this section, we propose a CNN-BLSTM model that fuses deep and shallow features to achieve dynamic parsing and classification of dance emotions through joint modeling of spatio-temporal features.

II. C. 1) Bidirectional long- and short-term memory networks

Compared with RNN, LSTM allows to retain or forget information, which is more suitable for long-term dependency problems, and can better solve the gradient disappearance and gradient explosion problems that tend to occur in long sequence training, but its memory capacity is also limited, while bidirectional long-short-term memory network

(BLSTM) consists of two LSTMs in the opposite direction, which has a stronger time series processing capacity.BLSTM The structure of BLSTM is shown in Fig. 2.

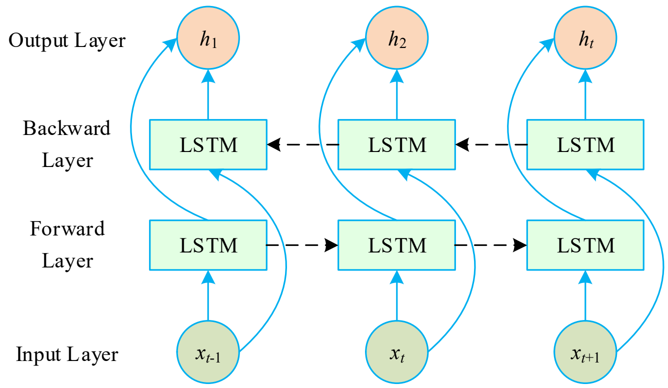

Figure 2: Bidirectional LSTM structure diagram

In a bidirectional LSTM, the output at time t depends not only on previous frames in the sequence, but also on upcoming frames. Since the emotion recognition of dance movements may depend on both past and future contexts at a given time step, this chapter uses a bi-directional LSTM, which consists of two independent hidden layers, and each node of the input layer is connected to these two independent hidden layers, and is computed once in the Forward layer along the 1 moment to the t moments in the forward direction to obtain and save the output of the forward hidden layer for each moment. In the Backward layer, the backward computation is performed from moment t to moment 1, and the output of the backward hidden layer is obtained and saved for each moment. Both hidden layers are connected to the same output layer, and finally the values from the forward and backward hidden layers are summed up at each moment to merge the past and future contexts.

II. C. 2) CNN-BLSTM model based on deep and shallow feature fusion

In this section, a convolutional neural network based on deep and shallow feature fusion is combined with a bidirectional long and short-term memory network and used for emotion recognition of dance movements. The CNN with deep and shallow feature fusion is used to extract the local features of the dance to be sent to BLSTM for temporal modeling, which in turn adequately captures the emotional features and contextual information of the dance action. Whereas, BLSTM captures the temporal dependencies, the dance emotion recognition model based on deep and shallow feature fusion CNN-BLSTM is shown in Fig. 3.

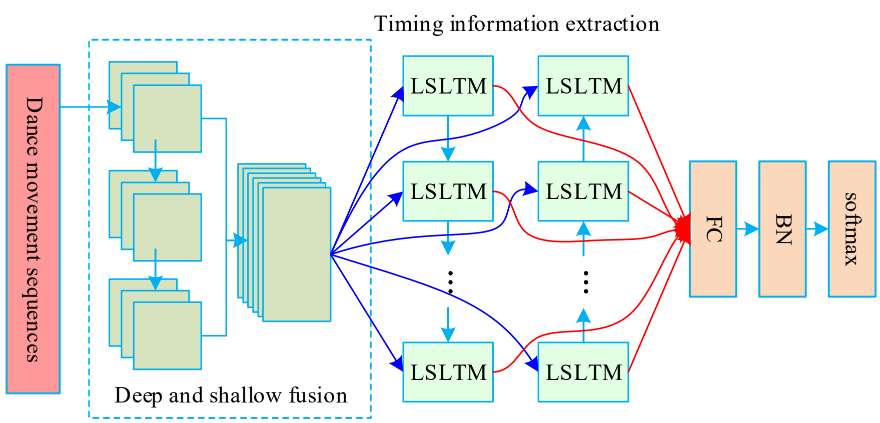

Figure 3: The dance emotion recognition model of CNN-BLSTM

After the dance action data sequence has passed through the deep and shallow feature fusion of the CNN, instead of directly using the selection of the fully connected layer, the shape of the fused feature vector is transformed, and then input into the subsequent BLSTM for temporal processing, which, on the one hand, avoids the loss of the representative features, and makes the extraction of the time-domain information more accurate, and on the other hand, prevents the occurrence of a situation in which the number of participants in the network decreases drastically due to the output features of the convolutional layer, and thus improves the generalization ability of the model. On the other hand, it also prevents the occurrence of a significant reduction in the number of parameters in the network due to the reduction of the output features of the convolutional layer, which in turn improves the generalization ability of the model and reduces the probability of the risk of overfitting. Subsequently, a fully connected layer is used and batch normalization (BN) is introduced to avoid possible gradient vanishing problems, and finally the emotions of the dance movements are classified in the Softmax layer.

III. Experiment and validation of dance movement emotion recognition based on CNN-BLSTM modeling

In Chapter 2, a CNN-BLSTM dance emotion recognition model is constructed based on action feature extraction and schema theory. In order to validate its actual performance, this chapter will carry out experiments around multimodal datasets, and comprehensively evaluate the effectiveness and robustness of the model through the comparison of accuracy, optimization of key frame extraction, and clustering analysis of emotion vectors.

III. A. Experimental setup

III. A. 1) Experimental environment

In order to verify the feasibility of the dance movement emotion recognition model based on CNN-BLSTM with deep and shallow feature fusion, the experimental environment is selected as CPU model Intel (R) Core (TM) I5-4460@3.20GHZ and memory size 8GB. The experimental operation is carried out under 64-bit Ubuntu. The simulation experiment platform is MATLAB2020b.

III. A. 2) Data sets

The datasets were selected from the commonly used DanceDB dance dataset and FolkDance dance dataset.

Among them, DanceDB contains 10 kinds of dance movements, which are represented by emotion labels, mainly including fearful, annoyed, bored, happy, painful, and tired movements, etc.; the performances are divided into 10 categories according to the emotional characteristics. Among the dance videos Afraid3, Angry4, Annoyed2, Bored5, Excited2, Happy3, Miserable5, Pleased4, Relaxed2, and Sad3 will be used for video retrieval, while the rest of the videos act as a retrieval database.

FolkDance contains 4 types of dance movements, namely heel double flower combination, rikatana combination, hand towel flower combination and katana combination. Both datasets have a frame rate of 20fps and a frame section size of 480×360.

The above datasets are all live recorded video images. A large amount of noise occurs in the process of converting video to image, which leads to poor dance movement feature extraction. To solve this problem, the two datasets are preprocessed. The specific processing methods are background elimination and median filtering methods. Among them, the main function of background elimination is to extract the foreground, and then separate the human movement region; the median filtering method filters the noise in the dataset, so as to reduce the impact of noise on feature extraction.

III. A. 3) Experimental methods

In order to better validate the proposed algorithm, the cross-validation method is used for testing. The commonly used methods are K-fold crossover and leave-one-out cross-validation, where leave-one-out crossover achieves more realistic and accurate results, i.e., it takes the average of 10 validation results, so leave-one-out crossover is chosen for algorithm validation.

III. B. Comparative Analysis of Dance Movement Recognition Accuracy and Key Frame Extraction by Multimodal Feature Fusion

Based on the experimental environment and data preprocessing settings, this section further compares and analyzes the accuracy difference between the proposed CNN-BLSTM model and the traditional single-feature approach in complex dance movement recognition, and explores the optimization effect of key frame extraction.

III. B. 1) Comparison of Dance Movement Recognition Accuracy

In order to verify the effectiveness of the proposed algorithm, the DanceDB and FolkDance datasets are used as test sets for the experiment, so as to realize complex dance movement recognition. And the recognition rate of the proposed method is compared with three other single features (directional gradient histogram feature, optical flow directional histogram feature and audio signature feature) in the two datasets respectively, and the comparison results of the experimental results on the DanceDB and FolkDance datasets are shown in Figs. 4 and 5, respectively.

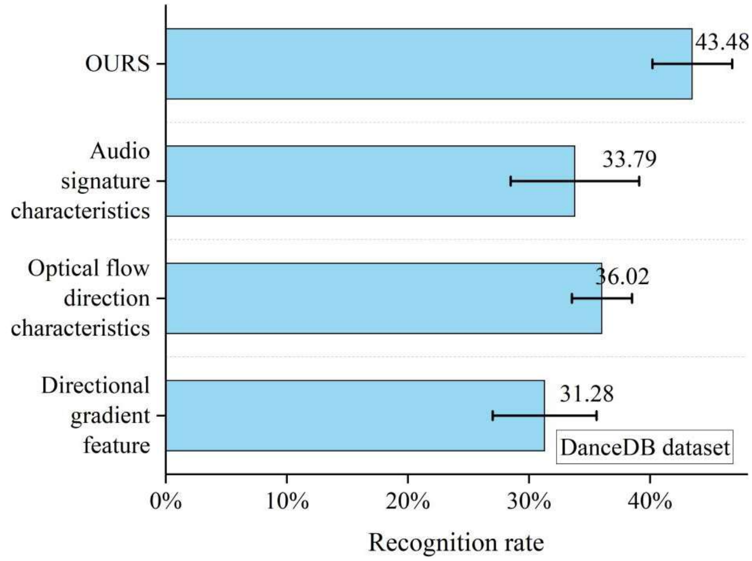

Figure 4: The comparison results of experimental results on the DanceDB dataset

From Fig. 4, it can be visualized that compared to the other three single methods, the recognition accuracy of this paper's method is the highest, with a recognition rate as high as 43.48%, which is 12.20%, 7.46% and 9.69% higher than that of the directional gradient, the optical flow direction, and the audio features, respectively. This shows that compared to the single feature recognition methods, the method in this paper has higher recognition accuracy and better recognition effect for complex dance movements.

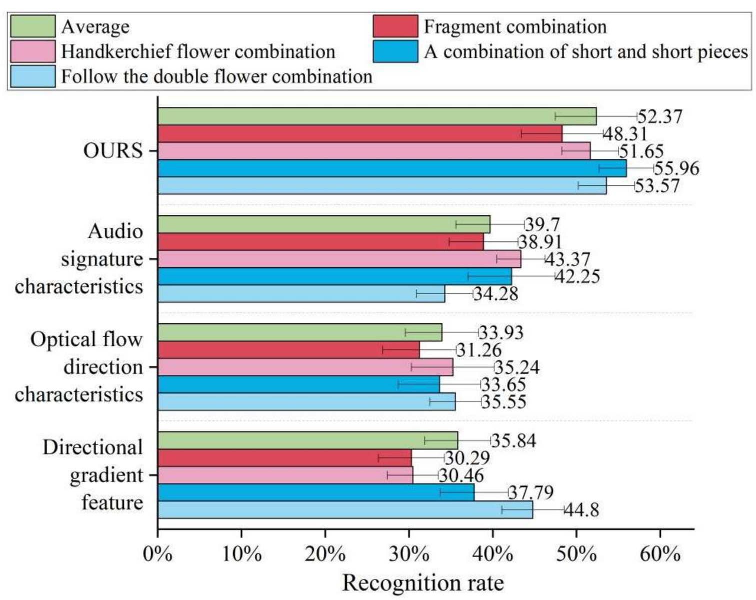

Figure 5: The comparison results of experimental results on the FolkDance dataset

As can be seen from Figure 5, among the four methods, the recognition accuracy of the proposed method is higher than the other three methods. Among them, the recognition effect of each method is different for different dance movements. In the heel-step double flower movement, the recognition accuracy of the proposed method reaches 53.37%, and the recognition rate of the directional gradient histogram is 44.80%, which are both higher than that of the optical flow feature and the audio feature; in the lilac flower, the recognition rate of the proposed method is 55.96%, and the recognition rates of the directional gradient feature and the audio feature are 37.79% and 43.37%, respectively, and the recognition rate of the optical flow feature is the lowest, which is 33.65%; in the combination of Hand Towel Flower and Piece Flower, the recognition rate of this method is 51.65% and 48.31%, respectively, which are higher than the other three methods, but the difference is that the recognition rate of the audio feature is higher than that of the directional gradient and the optical flow feature, which are 43.37% and

38.91%, respectively. Comprehensive analysis shows that in the FolkDance dataset, the dance movement emotion recognition model based on CNN-BLSTM with deep and shallow feature fusion has the highest recognition accuracy for complex dance movements, with an average accuracy of 52.37%, which has a certain degree of validity, while the other three methods have an average accuracy of 35.84%, 33.93%, and 39.70%, respectively.

III. B. 2) Comparison of key frame extraction results

In this paper, dance video retrieval is performed by extracting key frames from music and dance videos. In video retrieval, key frames can be utilized to represent the main content of a video, reducing the amount of computation. Key frame extraction of dance video based on depth feature extraction is compared with the three feature extraction methods mentioned above. In the experiment, since the dance movements of the dance video and its accompanying music mirror each other, the corresponding entropy value changes when the accompanying music keeps on changing over time, so that heavy beats, and beats with a relatively large amount of movement, can be effectively differentiated. However, based on the directional gradient feature and the optical flow direction feature, the candidate keyframes are mainly selected by clustering, and then ranked to select the more representative keyframes. The set of keyframes selected by this method is less redundant, but the selected keyframes are not accurate enough and differ from the user evaluation, mainly because the method ignores the audio features. The threshold selection in the experiment, on the other hand, is made in accordance with the evaluation criteria, and the selection is carried out through continuous iteration to select a collection of keyframes with high evaluation coefficients. A comparison of the key frame extraction results of different methods on the DanceDB dataset is shown in Table 1.

Table 1: Comparison of key frame extraction results by different methods

		Afraid	Angry	Annoyed	Bored	Excited	Happy	Miserable	Pleased	Relaxed	Sad
	Total number of frames	1337	496	1376	2579	615	3102	1345	2656	2191	1044
Directional gradient feature	Key frame count	33	20	65	169	22	193	36	167	141	51
	Compression ratio (%)	3.9	5.4	2.9	4.5	3.1	2.1	2.6	5.8	5.8	6.5
	Precision	68.95	61.53	67.37	80.85	78.04	82.08	78.73	71.89	75.45	72.11
	Recall	76.26	74.7	62.91	70	77.82	71.41	70.28	70.36	64.93	72.77
	F1	72.42	67.48	65.06	75.03	77.93	76.37	74.27	71.12	69.80	72.44
Optical flow direction characteristics	Key frame count	35	21	70	176	24	200	34	154	144	60
	Compression ratio (%)	3.6	5.2	2.5	4.1	2.7	2.1	2.7	6.1	5.8	6.6
	Precision	86.15	84.63	85.17	82.81	79.39	88.65	84.88	79.05	77.73	81.92
	Recall	76.47	83.59	70.86	72.71	75.54	77.1	73.04	76.43	84.62	82.96
	F1	81.02	84.11	77.36	77.43	77.42	82.47	78.52	77.72	81.03	82.44
Audio signature characteristics	Key frame count	28	14	58	141	14	149	24	133	126	44
	Compression ratio (%)	2.6	4.2	1.9	4.0	2.1	1.6	1.7	4.7	5.1	5.7
	Precision	90.77	86.52	90.29	90.77	94.4	89.01	95.74	89.97	90.99	89.53
	Recall	86.96	87.39	84.63	78.85	83.4	79.37	83.54	82.32	86.14	85.76
	F1	90.77	86.52	90.29	90.77	94.4	89.01	95.74	89.97	90.99	89.53
Our method	Key frame count	25	12	61	136	9	148	20	123	123	42
	Compression ratio (%)	2.1	3.7	1.6	3.4	1.8	1.3	1.6	4.4	4.7	5.0
	Precision	95.04	94.12	93.20	93.15	98.21	98.73	98.59	97.80	93.57	92.97
	Recall	86.07	85.31	91.71	86.3	92.07	93.57	89.17	85.29	85.91	89.86
	F1	90.33	89.50	92.45	89.59	95.04	96.08	93.64	91.12	89.58	91.39

In the key frame extraction comparison experiments on the DanceDB dataset, the CNN-BLSTM model based on deep and shallow feature fusion proposed in this paper significantly outperforms the traditional method in several metrics. For different emotional categories of dance movements (e.g., fear, anger, happiness, etc.), the directional gradient features and optical flow direction features select key frames by clustering, but their compression rates are 4.26% and 4.14% on average, which have a low match with the user assessment, and the ignoring of the audio features leads to the poor performance of both the key frame precision rate and the recall rate. Although the audio signature feature performs better in some of the actions, its keyframe redundancy is higher, with a compression rate of 3.36% on average and an F1 score of 90.79% on average, which is lower than that of this paper's method. In contrast, the method in this paper optimizes the key frame selection by fusing multimodal features (visual and audio), which reduces the compression ratio to an average of 2.96%, and at the same time achieves a precision and recall rate of 95.54% and 88.53%, respectively, with an average F1 score of 91.87%. For example, in the “happy” movement, the keyframe precision rate of 98.37% and recall rate of 93.57% of this method significantly exceeds that of other methods, which indicates that it can capture the synergistic changes of the dance movement and the music beat more efficiently, reduce redundancy, and improve the semantic representativeness of the keyframes. This result validates the effectiveness of multi-feature fusion strategy in complex dance sentiment analysis.

III. C. Analysis of clustering and distribution results of sentiment vectors

Based on the verification of the model recognition accuracy, this section analyzes the distribution of emotion vectors by means of the support vector clustering method to reveal the emotional relevance and cluster topology of different dance movements.

In order to reveal the characteristics between emotions, this section clusters the emotion vectors in the FolkDance dataset using the support vector clustering method, which divides the data points into different clusters, making the data points within the same cluster more similar and the data points between different clusters more different. Where the distance metric uses cosine distance. The clustering is visualized using the mean-variance coordinates, and the distribution of the emotion clusters is shown in Figures 6, 7, 8, and 9, respectively. Each point therein corresponds to the emotion vector for each dance movement. The horizontal axis represents the mean value and the vertical axis represents the variance value.

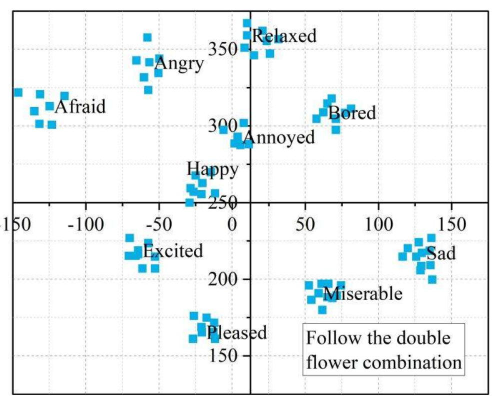

Figure 6: A cluster formed by following the double flower combination

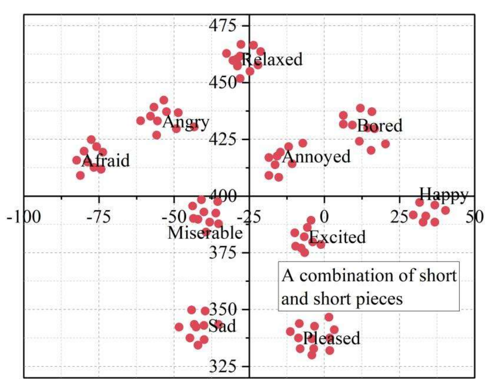

Figure 7: A cluster of interior flower combinations

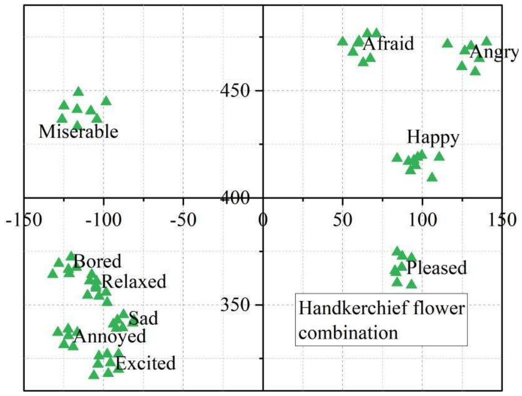

Figure 8: Clusters of handkerchief flowers

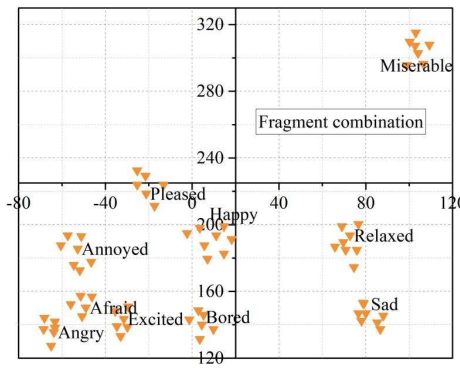

Figure 9: Clusters of floral combinations

Observing the emotional characteristics of different dance movements in Figs. 6-9, it can be concluded:

In the heel-step double-flower combination movement, Afraid and Happy are basically distributed on the left and right sides, and their relative positions are stable.Relax is distributed near the y=0 line, which is located in the middle of the ten emotions. Therefore, Afraid→Relax→Happy forms a relatively stable topology, with Afraid and Happy distributed on both sides and Relax in the middle.

In the Ri-katana combination movement, it is similar to the heel-step double-flower combination movement, and also presents a relatively stable topology of Afraid→Relax→Happy, but Excited is closer to afraid in the heel-step double-flower combination movement, and closer to Happy emotion in the Ri-katana combination movement, so Excited is more similar to Happy than the other emotions. .

In the clusters of hand towel flower combination, the distribution of Bored, Sad, Annoyed, Excited and Relaxed is unstable and varies a lot with the dance movements.Relaxed is sometimes close to Bored and sometimes close to Sad.So Relaxed can be expressed like Bored and Sad.

The clusters of Katana combination actions, Angry→Happy→Miserable can also form a topological structure around which other emotions are distributed.

Based on the appeal observation, the experiment shows that both anchor point emotions and emotion vectors can be well clustered, and the clusters of anchor point emotions and FolkDance dataset have similar distributions. This suggests that emotion vectors can represent not only anchor point emotions but also emotion categories from the FolkDance dataset.

III. D. Validation of Dance Emotion Distribution Based on Emotion Vector Clustering and Standard Parameter Matching

To further validate the physical meaning of the sentiment vectors, this section combines the matching analysis of the joint motion features of the dance movements with the standard sentiment parameter library AnimStreet to explore the consistency of the model outputs with the manually labeled sentiment labels.

Table 2 lists the standard sentiment data corresponding to the synchronized FolkDance data for the dance gestures, with values generated based on bvh in AnimStreet after manual discrimination of labels. Since six of the ten feature vectors are left-right symmetric (Angry and Annoyed; Pleased and Happy; Sad and Miserable), only seven data items are listed in the standard set for each gesture. Three pieces of data are listed for each cell for illustrative purposes: period corresponds to the angle of the center of the circle, the number of frames experienced, and the final rate characterization value.

Table 2: Standard emotional parameters of dance movements

Joint		Angry	Pleased	Sad	Afraid	Bored	Excited	Relax
Arm	Central Angle	21.89°	23.59°	14.63°	22.3°	15.36°	18.59°	12.80°
	Frame rate	7.42fs	5.11fs	5.06fs	10.19fs	5.40fs	8.54fs	5.68fs
	Velocity	3.421	1.802	0.413	1.372	1.052	2.557	0.868
Hand	Central Angle	16.09°	8.59°	11.48°	6.44°	9.12°	10.55°	6.26°
	Frame rate	7.64fs	5.00fs	5.81fs	10.38fs	9.83fs	7.78fs	5.06fs
	Velocity	2.802	1.218	0.276	1.922	0.921	2.894	0.590
Thigh	Central Angle	20.65°	14.83°	10.14°	17.26°	8.43°	19.4°	7.09°
	Frame rate	10.13fs	7.12fs	5.90fs	8.44fs	6.75fs	10.13fs	6.09fs
	Velocity	2.154	1.896	0.826	1.680	0.229	2.480	0.644
Foot	Central Angle	13.87°	8.95°	6.1°	6.88°	4.63°	12.1°	4.01°
	Frame rate	10.41fs	5.49fs	8.35fs	10.18fs	7.61fs	8.51fs	6.89fs
	Velocity	1.770	2.206	0.664	1.858	0.355	2.612	0.370

Table 2 shows the characteristics of the motion parameters of the key joints (arms, hands, thighs, and feet) of the human body in different dance emotional states, including the central angle, frame rate, and rate values. From the data distribution, it can be seen that there are significant differences in the characteristics of joint movement among different emotions: Angry: the rate of each joint is generally higher (for example, the arm rate is 3.421, the hand is 2.802), and the central angle is larger (arm 21.89°), indicating that the movement amplitude is large and rapid, which is in line with the explosive characteristics of anger; Pleased: the central angle of the arm is large (23.59°), but the velocity value is low (1.802), and the movements are smooth and stretched, reflecting the elegant expression of emotion; Sad (Sad) has the lowest velocity value of each joint (e.g., arm 0.413), the central angle is small (arm 14.63°), and the movement is slow and small, reflecting the low and restrained emotion. Excited: The rate value is significantly higher than that of other emotions (e.g., thigh 2.480), the frame rate is higher (arm 8.54fs), and the action is fast-paced and energetic, which is consistent with the performance of the excited state. In addition, there is a symmetry between emotions, such as anger and annoyed (Annoyed) are similar in parameters, indicating the similarity of emotional expression. The data verified the effectiveness of the quantitative analysis of dance emotion based on biomechanical features, and provided a physical basis for model recognition.

Taking the 50 frames of data of the dance action of a Ri film combination as an example, the emotion recognition results are shown in Table 3.

Table 3: Emotion recognition result

	L-Arm	R-Arm	L-Hand	R-Hand	L-Thigh	R-Thigh	L-Food	R-Food
Central Angle	24.39°	17.49°	11.45°	10.89°	13.2°	14.52°	10.92°	14.23°
Frame rate	4.67fs	9.06fs	6.4fs	8.33fs	7.46fs	8.54fs	8.57fs	8.41fs
Velocity	1.899	2.504	2.830	2.240	1.820	1.785	2.797	2.614
Emotion	Pleased	Excited	Excited	Excited	Pleased	Pleased	Excited	Excited

The rate values are obtained from the joint motion data after hemispherical modeling and normalization. The rate of other nodes is mostly Excited, so it can be concluded that the dance performer is more likely to be “excited”.

IV. Conclusion

In this study, the quantitative analysis and dynamic recognition of dance body language and emotion expression are realized by computer vision technology. The experimental validation of the CNN-BLSTM model constructed based on biomechanical feature extraction and schema theory on the DanceDB and FolkDance datasets shows that its emotion recognition accuracy reaches 43.48% and 52.37%, respectively, which is a significant enhancement over the traditional methods. In keyframe extraction, the multimodal feature fusion strategy optimizes the semantic representation and reduces the compression rate to 2.96%, while the accuracy and F1 score are improved to 95.54% and 91.87%, respectively. The emotion vector clustering analysis further reveals the emotion distribution patterns of different dance movements, such as the differences in joint rates (e.g., arm rate of 3.421 and 2.557) and frame rates (7.42 fs and 8.54 fs) between “anger” and “excitement”, which validate the biomechanics of the dance movements. The difference between “excited” and “excited” in joint rates, such as arm rates of 3.421 and 2.557, and frame rates of 7.42 fs and 8.54 fs, respectively, validates the strong correlation between biomechanical parameters and emotional states. Matching experiments with standard emotional parameters showed that the model outputs were in high agreement with the manually labeled labels, e.g., the rate values of “excited” actions were 2.480-2.894, which were significantly higher than those of “sad” (0.276-0.826).

References

[1] Cox, L., & Youmans-Jones, J. (2023). Dance is a healing art. Current Treatment Options in Allergy, 10(2), 184-195.

[2] Lu, Y. (2022, February). Analysis of body and emotion in dance performance. In 2021 Conference on Art and Design: Inheritance and Innovation (ADII 2021) (pp. 46-50). Atlantis Press.

[3] Saumaa, H. (2022). Dance emotions. Integrative and Complementary Therapies, 28(3), 134-137.

[4] Beauquel, J. (2013). Physical and aesthetic properties in dance. Thinking Through Dance: The Philosophy of Dance Performance and Practices, 17-35.

[5] Laland, K., Wilkins, C., & Clayton, N. (2016). The evolution of dance. Current Biology, 26(1), R5-R9.

[6] Peng, X. (2024). Historical development and cross-cultural influence of dance creation: Evolution of body language. Herança, 7(1), 88-99.

[7] Brooks, L. M., & Meglin, J. A. (2015). Language and dance: Intersection and divergence. Dance Chronicle, 38(2), 127-133.

[8] Shikanai, N., & Hachimura, K. (2015, October). Evaluation of impressions and movements related to negative emotional expressions in dance. In 2015 15th International Conference on Control, Automation and Systems (ICCAS) (pp. 657-660). IEEE.

[9] Saumaa, H. (2022). Developing self-confidence through the body and dance. Journal name: Integrative and Complementary Therapies, 28(6), 281-285.

[10] Burger, B., & Toiviainen, P. (2020). See how it feels to move: relationships between movement characteristics and perception of emotions in dance. Human Technology, 16(3), 233-256.

[11] Yang, J. (2015, June). Analysis on Proper Application of Body Language in National Dance Teaching. In International Conference on Education, Management and Computing Technology (ICEMCT-15) (pp. 812-815). Atlantis Press.

[12] Uspuriene, A. B., & Sniras, S. (2019). Impact of dance sport on communication abilities of adolescents. Teoriya i Praktika Fizicheskoy Kultury, 14-16.

[13] Li, M., Miao, Z., & Ma, C. (2019). Dance movement learning for labanotation generation based on motion-captured data. IEEE Access, 7, 161561-161572.

[14] Theodorou, L., Healey, P. G., & Smeraldi, F. (2019). Engaging with contemporary dance: What can body movements tell us about audience responses?. Frontiers in psychology, 10, 71.

[15] Premelč, J. (2022). TIME MOTION AND PERFORMANCE ANALYSIS IN TANGO DANCE. Kinesiologia Slovenica, 28(1).

[16] Dieterich-Hartwell, R. M. (2019). Music, movement, and emotions: an inquiry with suggestions for the practice of dance/movement therapy. Body, Movement and Dance in Psychotherapy, 14(4), 249-263.

[17] Mateu, M., Garcías, S., Spadafora, L., Andrés, A., & Febrer, E. (2021). Student moods before and after body expression and dance assessments. Gender perspective. Frontiers in Psychology, 11, 612811.

[18] Borowski, T. G. (2023). How dance promotes the development of social and emotional competence. Arts Education Policy Review, 124(3), 157-170.

[19] German-Molina, L., Caballero-Julia, D., & Cuellar-Moreno, M. (2022). Dance as a tool for managing emotions. A systematic review. Research in Dance Education, 1-12.

[20] Amado, D., García-Calvo, T., Marreiros, J., Chamorro, J. L., & Del Villar, F. (2015). Analysis of students’ emotions in agreement with the dance teaching technique used. European Journal of Human Movement, 34, 123-138.

[21] Su, C. (2025). Exploring the Diverse Pathways of Emotional Expression in Dance Performance and Enhancing Its Artistic Appeal. Advances in Education, Humanities and Social Science Research, 13(1), 233-233.

[22] Sun, Q., & Wu, X. (2023). A deep learning-based approach for emotional analysis of sports dance. PeerJ Computer Science, 9, e1441.

[23] Roxas, J. C., Richards, D., Bilgin, A., & Hanna, N. (2018). Exploring the influence of a human-like dancing virtual character on the evocation of human emotion. Behaviour & information technology, 37(1), 1-15.

[24] Panagiotopoulou, E. (2018). Dance therapy and the public school: The development of social and emotional skills of high school students in Greece. The Arts in Psychotherapy, 59, 25-33.

[25] Çetin, Z., & Erdem Çevikbaş, P. (2020). Using creative dance for expressing emotions in preschool children. Research in Dance Education, 21(3), 328-337.

[26] Christensen, J. F., Bruhn, L., Schmidt, E. M., Bahmanian, N., Yazdi, S. H., Farahi, F., ... & Menninghaus, W. (2023). A 5-emotions stimuli set for emotion perception research with full-body dance movements. Scientific Reports, 13(1), 8757.

[27] Smith, R. A., & Cross, E. S. (2023). The McNorm library: creating and validating a new library of emotionally expressive whole body dance movements. Psychological research, 87(2), 484-508.