

# A Study on the Visual Construction and Pathways for the Transmission of Cultural Memory in the Yao Long Drum Dance, Empowered by Digital Humanities

Yang Zhang<sup>1</sup> and Dong Shu<sup>1,\*</sup>

<sup>1</sup> International College, Krirk University, Bangkok, 10220, Thailand

\* Correspondence author: shudong8701@163.com

**Abstract:** In 2008, the Yao Long Drum Dance was included in the second batch of the National Intangible Cultural Heritage List, bringing to an end its long and tortuous history of marginalization and marking its transition from a regional dance to a national one. This study constructed a Yao Long Drum Dance dataset. To address the issue of self-occlusion of skeletal points during data acquisition, existing algorithms for restoring motion information in self-occlusion scenarios were improved, thereby recovering the position information of occluded skeletal points. Skeletal data of professional Yao Long Drum Dance performers were captured and stored using a Kinect, and a 3D CNN was designed for the recognition of typical movements in the Yao Long Drum Dance. Test results demonstrate that the dance motion capture method proposed in this paper provides significant support for dance motion acquisition and training. The 3D CNNs achieved a recognition accuracy of up to 96.45%, thereby facilitating the digital preservation and transmission of the Yao Long Drum Dance.

**Keywords:** 3D CNNs; Kinect; Yao Long Drum Dance; cultural heritage

## 1. Introduction

The Yao people possess a rich and long-standing history and culture, forming an important part of China's outstanding traditional culture. Over the course of history, the Yao culture has undergone exchanges, conflicts, and fusions resulting from internal developments within the ethnic group or interactions with other ethnic groups. This process is not merely a replication of cultural traits, but rather an integration and evolution at the sociocultural level [1–3]. The Yao Long Drum Dance is a symbol of Yao culture [4]. Due to historical, environmental, and cultural factors, most of the Yao people's traditional culture has been passed down orally, with very few written records. Therefore, exploring the evolution of the Yao Long Drum Dance culture holds significant value for understanding Yao culture.

Cultural change is regarded by anthropology as a normal phenomenon; stability and equilibrium during a given period are relative, while change and development are absolute [5]. The essence of cultural change lies in the transformation of cultural structures. Generally, research on cultural structure encompasses two-level, three-level, and multi-level theories [6]. For the Yao Long Drum Dance culture, changes at the material level—such as tools and artifacts [7]—constitute the observable and perceptible aspects of culture. Changes at the institutional level include shifts in institutions and behaviors, such as variations in performance forms and types, while changes in the functional value of the Long Drum Dance reflect its spiritual dimension [8].

German scholars Jan Assmann and his wife Aleida Assmann first introduced the concept of “cultural memory” in the 1980s [9] and conducted research and discussions on its conceptual characteristics, internal structure, and historical and social functions. From the perspective of the media and carriers of cultural memory, some scholars argue that the transmission and reinforcement of



historical images and memories are often achieved through ritualistic activities [10–11]. The symbolic repetition unique to commemorative activities plays a crucial role in shaping collective memory. The accumulation of bodily experience is primarily manifested at two levels: embodiment and inscription [12]. Alida Assmann’s wife has conducted an in-depth analysis of the functions, carriers, and modes of preservation of cultural memory. Through numerous examples from literary works, she has highlighted the crucial role of memory in shaping identity, as well as the complexity of memory selection and reconstruction [13]. For a form of cultural memory to possess vigorous and enduring vitality, it must take into account the emotions and interests of all members within the relevant collective [14]. It is well known that the Yao are a migratory people; the hardships endured by their ancestors during their migrations resonate with the collective emotions of the Yao people, constantly reminding them to remain vigilant even in times of peace [15–17]. The Yao people have wisely condensed the history of their migrations into the movements of the Yao Long Drum Dance; the ritual, celebrations, the continuous, cyclical movements of the Long Drum Dance evoke the image of their ancestors—constantly oppressed, pursued, and forced to flee—thereby ensuring that the historical and cultural memory of the Yao people’s migration, along with their spirit of bravery and fearlessness in overcoming adversity, is subtly and continuously perpetuated through the dance.

This study uses optical motion capture equipment to precisely capture the skeletal movements of the Yao Long Drum Dance, thereby enriching the modalities of the data. To address the issue of self-occlusion of skeletal points during data acquisition, existing algorithms for restoring motion information in cases of self-occlusion were improved, successfully recovering the position information of occluded skeletal points. Using a Kinect, 15 representative dance movements performed by 10 volunteers were recorded to create a Yao Long Drum Dance movement dataset, and a 3D CNN was proposed for recognizing the constructed dataset. Motion capture using the method described in this paper was employed to facilitate dance training. A 3D CNN was utilized to automatically recognize and analyze the movements of the Yao Long Drum Dance, helping users learn and master the dance’s postures and movements. This study holds significant importance for both the digitization of the Yao Long Drum Dance and the preservation of intangible cultural heritage.

## **2. Construction of a Yao Long Dance Movement Dataset Incorporating Cultural Memory**

### *2.1. Motion Capture of the Yao Long Drum Dance*

To create a motion dataset for the Yao ethnic group’s long drum dance, this study utilized three DJI OSMO Action 3 cameras and an OptiTrack passive optical motion capture system. The Action 3 cameras were used to capture RGB video from different angles, while the OptiTrack motion capture system was used to capture motion data. The data collection took place in a room measuring approximately  $7\text{ m} \times \text{m}$ . Eight high-speed infrared motion capture cameras from the OptiTrack system were mounted around the perimeter of the room, and the floor was marked with non-reflective tape to delineate key points and the dance area. One side of the room served as the performance area, separated from the motion capture area by a white curtain. The dancer, wearing a motion capture suit, performed within the motion capture area while surrounded by three Action 3 cameras, which simultaneously captured RGB video and motion data. A total of 11 data acquisition devices were used: eight infrared motion capture cameras were mounted on the walls surrounding the room, and three cameras were positioned around the dancer at a distance of 3 meters.

**Dance Categories:** This study collected various categories of the Yao Long Drum Dance, specifically the Pangu Long Drum Dance, the Gong and Sheng Long Drum Dance, the Reed Sheng Long Drum Dance, and the Sheep Horn Short Drum Dance.

**Camera Perspectives:** This study used three cameras simultaneously to capture RGB video from three different perspectives. The three cameras were positioned at the same height and evenly distributed around the dancer, specifically on a circle with a radius of 3 meters centered at the dancer’s initial position, with each pair of adjacent cameras forming a  $120^\circ$  angle with the line connecting them to the center of the circle.

### *2.2. Preprocessing of Skeletal Data*

Motion extraction is based on the 3D coordinate data of 25 human skeletal joints captured by the Kinect v2. However, during training, factors such as clothing and body movements can obscure the finger skeletal points, leading to inaccurate data capture. Additionally, considering that finger skeleton points have a minimal impact on changes in motion features, finger skeleton data is no longer collected during the skeleton data acquisition process. Instead, only the data for skeleton points that play a

critical role in the movement is tracked and captured to improve motion capture efficiency.

Differences in trainees' body types and variations in their distance from the Kinect during practice can lead to errors in the skeleton data obtained each time. Therefore, before the data captured during motion capture can be applied to motion matching, the acquired skeletal data model must first be normalized.

Using the 3D coordinate information of the 20 key skeletal points obtained from the Kinect, the position of each skeletal point at a given frame can be represented as [18]:

$$P_{20} = (X_1, Y_1, Z_1 \cdots X_{20}, Y_{20}, Z_{20}) \quad (1)$$

During Kinect-based dance motion capture, the distance between the left and right shoulders changes very little, so their center can be used as a reference point. The coordinates of reference point  $A$  can be calculated using the following formula:

$$A(X_a, Y_a, Z_a) = \frac{1}{2}(X_1 + X_2, Y_1 + Y_2, Z_1 + Z_2) \quad (2)$$

Once the reference points are obtained, the 3D coordinates of the 20 retained human body joints are normalized.

$$P' = \frac{P - A}{|L - R|} \quad (3)$$

Here,  $A$  represents the 3D coordinates of the reference skeletal point,  $P$  represents the position coordinates of the skeletal point before normalization,  $P'$  represents the position coordinates of the skeletal point after normalization,  $L$  represents the position coordinates of the left shoulder, and  $R$  represents the position coordinates of the right shoulder. The standard length is the distance between the left and right shoulder skeletal points. After normalization, the original coordinate system is transformed into a coordinate system with the center between the two shoulders as the reference point. This process reduces errors in the experimental results caused by differences in the trainee's body type.

### 2.3. Reconstruction of Occluded Skeletal Points and Motion Information

The Yao Long Drum Dance features complex movements. Due to factors such as costumes and movement patterns, certain joints may be obscured, resulting in some discrepancies between the captured data and actual data. Obscured points are reconstructed based on the continuity of human movement and the constancy of bone lengths. While Kinect is relatively accurate in capturing the human torso, errors may occur during the acquisition of skeletal points in the limbs due to occlusion. Assuming that the skeletal points captured by Kinect for the head, neck, and shoulders are reliable, it is only necessary to determine whether the skeletal points for the wrists, elbows, knees, and ankles are accurate and to correct any inaccurate joint points. This section uses an improved self-occlusion motion information restoration algorithm to identify and correct occluded skeletal points during dance movements.

#### (1) Detection of Occluded Skeletal Points

Occluded skeletal points can be detected based on the constancy of joint lengths and the continuity of motion. This chapter uses the constancy of joint lengths to detect occluded skeletal points. Specifically, when the body is unobstructed, the Kinect can capture the exact coordinates of each skeletal point. It then calculates the length of the body's joints based on the distance between two points, using this as evidence that the skeletal points are unobstructed.

#### (2) Restoration of Obstructed Skeletal Points

The correction of occluded skeletal points includes the correction of intermediate skeletal points  $B$  (elbow, knee) and terminal skeletal points  $C$  (wrist, ankle).  $Q_A$  is a sphere with center  $A$  and radius  $L_{AB}$ , and  $Q_C$  is a sphere with center  $C$  and radius  $L_{BC}$ . If the intermediate skeletal point  $B$  is determined to be unreliable, then the range of skeletal point  $B$  should lie on the circle  $S_b$ , which is the intersection of spheres  $Q_A$  and  $Q_C$ . Therefore, for a point  $(x, y, z)$  on circle  $S_b$ , the following holds:

$$\begin{cases} L_{AB} = \sqrt{(x - x_a)^2 + (y - y_a)^2 + (z - z_a)^2} \\ L_{BC} = \sqrt{(x - x_c)^2 + (y - y_c)^2 + (z - z_c)^2} \end{cases} \quad (4)$$

The velocity of a skeletal point does not change significantly over a short period of time. When the animation reaches frame  $K$ ,  $P_{b,K}(x_{b,K}, y_{b,K}, z_{b,K})$  represents the position coordinates of point  $B$ . Assuming the interval between two frames is  $t$ , the average velocity  $V_{B,K-1} = (V_{Bx,K-1}, V_{By,K-1}, V_{Bz,K-1})$  of point  $B$  at frame  $K-1$  can be calculated using the motion information in Equation (5).

$$\begin{cases} V_{Bx,K-1} = \frac{x_{b,K-1} - x_{b,K-1-i}}{i \cdot t} \\ V_{By,K-1} = \frac{y_{b,K-1} - y_{b,K-1-i}}{i \cdot t} \\ V_{Bz,K-1} = \frac{z_{b,K-1} - z_{b,K-1-i}}{i \cdot t} \end{cases} \quad (5)$$

From this, we can calculate the average velocity  $V_{B,K} = (V_{Bx,K}, V_{By,K}, V_{Bz,K})$  of skeleton point  $B$  prior to frame  $K$ .

$$\begin{cases} V_{Bx,K} = \frac{x_{b,K} - x_{b,K-i}}{i \cdot t} \\ V_{By,K} = \frac{y_{b,K} - y_{b,K-i}}{i \cdot t} \\ V_{Bz,K} = \frac{z_{b,K} - z_{b,K-i}}{i \cdot t} \end{cases} \quad (6)$$

The average speeds of bone point  $B$  at frames  $K$  and  $K-1$  are equal. According to the above formula, we can calculate that:

$$\begin{aligned} & (x_{b,K} - x_{b,K-1})^2 + (y_{b,K} - y_{b,K-1})^2 + (z_{b,K} - z_{b,K-1})^2 \\ &= V_{Bx,K-1}^2 + V_{By,K-1}^2 + V_{Bz,K-1}^2 \end{aligned} \quad (7)$$

At frame  $K$ , the coordinates  $P_{b,k} = (x_{b,K}, y_{b,K}, z_{b,K})$  of skeleton point  $B$  lie at the intersection of sphere  $Q_O$  and circle  $S_b$ , as shown in Equation (8), where sphere  $Q_O$  has  $O$  as its center and  $R$  as its radius.

$$\begin{cases} L_{AB}^2 = (x_{b,K} - x_{a,K})^2 + (y_{b,K} - y_{a,K})^2 + (z_{b,K} - z_{a,K})^2 \\ L_{BC}^2 = (x_{b,K} - x_{c,K})^2 + (y_{b,K} - y_{c,K})^2 + (z_{b,K} - z_{c,K})^2 \\ (x_{b,K} - x_{b,K-1})^2 + (y_{b,K} - y_{b,K-1})^2 + (z_{b,K} - z_{b,K-1})^2 \\ = V_{Bx,K-1}^2 + V_{By,K-1}^2 + V_{Bz,K-1}^2 \end{cases} \quad (8)$$

#### 2.4. Recording a Dance Movement Dataset

Ten physically healthy professional dance volunteers (age  $(23 \pm 3)$  years) participated in the recording of the Yao ethnic group's long drum dance movement dataset. The participants had not engaged in strenuous physical activity within 12 hours prior to the experiment, were in good health, and each had studied various dance styles, with 4 to 5 years of experience in traditional folk dance.

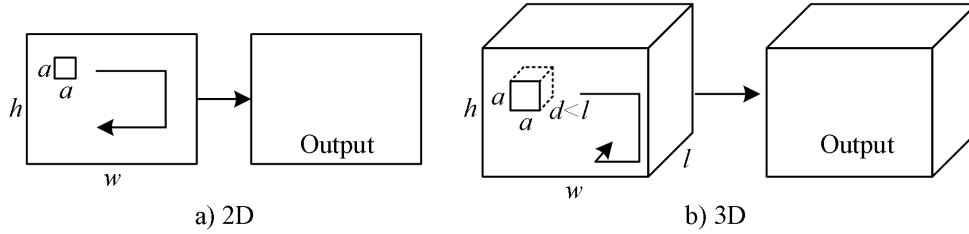
Before the experiment began, participants were required to sign the written study protocol and informed consent form. During the experiment, the following constraints were established based on the actual laboratory environment: (1) The Kinect should be positioned 1–1.2 m above the ground and 2–3 m horizontally from the trainee. Within this range, all skeletal joints of the human body can be fully detected by the Kinect camera, ensuring optimal data capture and tracking performance. (2) The relative angle between the human body and the Kinect must not exceed 45 degrees. If the relative angle is too large, the Kinect may detect joint positions with distortion. (3) When capturing human motion data, the trainee should stand directly in front of the Kinect and ensure that the testing area is open and free of obstructions.

The experiment required professional dancers to learn 15 sample dance movements from the Yao Long Drum Dance by studying the instructional videos and movement descriptions 3–5 days prior to data collection, and to be able to perform them fluently and in sequence to the rhythm of the music. During the experiment, participants were instructed to avoid pauses, excessive speed, or excessive slowness in their movements as much as possible. Throughout the experiment, participants performed the 15 movement sequences in accordance with verbal instructions and rhythmic cues, executing each sequence three times. A motion capture program was used to record the coordinates of skeletal points corresponding to the professional dancers’ standard dance movements. After the experiment concluded, the 600 recorded samples were added to the dataset.

### 3. Deep Learning-Based Recognition of Typical Movements in the Yao Long Drum Dance

#### 3.1. Principles of 3D Convolutional Neural Networks

Three-dimensional (3D) convolution is an improvement over two-dimensional (2D) convolution. Unlike 2D convolution, 3D convolution can extract features not only in the spatial dimension but also in the temporal dimension. Figure 1 shows a schematic diagram illustrating how 2D and 3D convolutions extract features. As shown in the figure, the features extracted by 2D convolution remain 2D feature maps, whereas those extracted by 3D convolution are 3D feature maps.



**Figure 1.** The difference between 2D convolution and 3D convolution

(1) Convolution layer. Similar to 2D convolution layers, 3D convolution layers also consist of multiple convolution kernels, with each kernel extracting different features. As shown in Equation (9), this represents the value of the element at position  $(x, y, z)$  in the  $j$ th feature map of the  $i$ th layer.

$$v_{ij}^{xyz} = f\left(b_{ij} + \sum_r \sum_{l=0}^{l_i-1} \sum_{m=0}^{m_i-1} \sum_{n=0}^{n_i-1} \omega_{jlr}^{lmn} v_{(i-1)r}^{(x+l)(y+m)(z+n)}\right) \quad (9)$$

(2) Pooling layer. Unlike pooling operations on two-dimensional images, temporal features—in addition to spatial features—also require pooling. Pooling reduces the feature dimension of the previous layer, accelerates model computation, and facilitates multiple training iterations. The calculation for max pooling is given by Equation (10):

$$V_{x,y,z} = \max_{0 \leq i \leq s_1, 0 \leq j \leq s_2, 0 \leq k \leq s_3} (\mu_{x \times s + i, y \times s + j, z \times s + k}) \quad (10)$$

#### 3.2. Recognition of Yao Long Drum Dance Movements Based on 3D CNNs

##### (1) Network Overview

Compared with other traditional deep learning methods, 3D CNNs are not limited to 2D single-frame image inputs; they can extract features from both the temporal and spatial dimensions and are capable of capturing motion information from multiple consecutive frames [19]. Therefore, this paper employs 3D CNNs to recognize the skeletal motion information of typical movements in the Yao Long Drum Dance. Since skeletal motion information is single-channel data, it requires less computational effort and results in better model recognition performance. The 3D deep convolutional neural network used in this study consists of 4 convolutional layers, 2 downsampling layers, 2 fully connected layers, and 1 Softmax classification layer. The downsampling layers use max-pooling with a kernel size of  $3 \times 3 \times 3$  and a stride of 1.

##### (2) Detailed Model Structure

To capture motion information across multiple consecutive frames, features are computed in both the spatial and temporal dimensions. The value of the unit at position  $(x, y, z)$  in the  $j$ th feature map

of layer  $i$  is given by the formula in (11):

$$V_{ij}^{xyz} = f \left( b_{ij} + \sum_r \sum_{l=0}^{l_i-1} \sum_{m=0}^{m_i-1} \sum_{n=0}^{n_i-1} \omega_{ijr}^{lmn} v_{(i-1)r}^{(x+l)(y+m)(z+n)} \right) \quad (11)$$

In this case, the temporal dimension of the 3D convolution kernel is  $n_i$ , the spatial dimension is  $(l, m, n)$ , and the weights of the kernel are  $\omega_{ijr}^{lmn}$  when applied to the  $r$ th feature map.

The ReLU function is the most commonly used activation function in deep learning models. This function enables the model's parameters to become sparse, thereby reducing overfitting. In addition, it reduces the computational load of the model. The ReLU activation function is defined by Equation (12):

$$f(x) = \max(0, x) = \begin{cases} 0, & (x \leq 0) \\ x, & (x > 0) \end{cases} \quad (12)$$

The calculation of max-pooling in the model is shown in Equation (13):

$$V_{x,y,z} = \max_{0 \leq i \leq s_1, 0 \leq j \leq s_2, 0 \leq k \leq s_3} \left( \mu_{x \times s + i, y \times t + j, z \times r + k} \right) \quad (13)$$

Here,  $\mu$  represents the 3-dimensional input vector,  $V$  represents the output after the pooling operation, and  $s, t, r$  represents the sampling stride in the direction. The Softmax function is commonly used in the final layer of classification tasks; it maps an  $n$ -dimensional vector  $x$  to a probability distribution such that the probability of the correct class approaches 1, the probabilities of other classes approach 0, and the sum of the probabilities across all classes equals 1.

## 4. Experiments and Analysis

### 4.1. Laboratory Equipment and Preparations

#### (1) Experimental Equipment

All training methods used in the experiment employed the same experimental equipment. The computer specifications are as follows: CPU: i5 4690; RAM: Hacker DDR4 2666 8GB; Hard Drive: WD10EZEX 1T; Graphics Card: GTX950 2G. A Kinect was used to capture 2D and depth images of the human body, connected to the computer via a USB port. Since the Kinect can determine the most accurate motion data by tracking multiple sensors worn on the subject's body and performing computational corrections and spatial localization—making it currently the most accurate method for motion capture—the Kinect results were used as the reference standard. An Ordoro AC5 camera was used for video recording.

#### (2) Experimental Preparation

To correct trainees' incorrect movements, standard dance routines are required for comparison, and instructional videos of standard movements must be recorded. In the experiment designed for this paper, a professional dance instructor was invited. The instructor's movement data was captured using the Kinect, and the movements were recorded using the Ordoro AC5. Kinect preserves each dance movement using an occlusion point recovery algorithm. The motion data includes 30 dance movements, which are divided into four decomposed movements to meet the diverse needs of trainees.

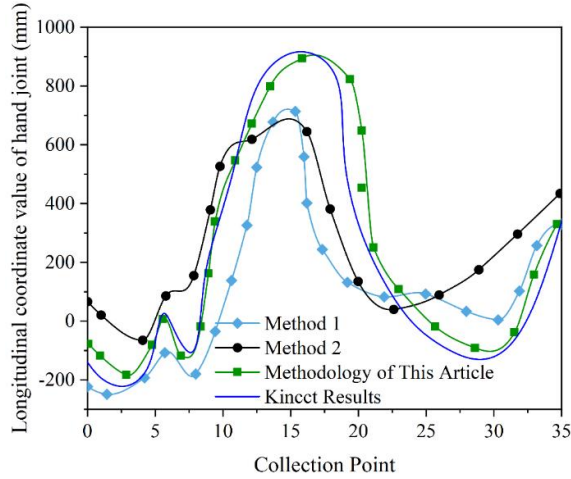
### 4.2. Motion Capture Experiment Results and Analysis

The assisted training method collects the joint coordinates of each trainee's movements and compares them with the standard movements. By using Kinect to capture the trainee's movement coordinates, the method compares the joint coordinate errors derived from the three dance assistance methods with the data collected by Kinect. The results are shown in Table 1. As can be seen from the table, the proposed method calculates the differences between the trainee's movements and the standard movements with greater accuracy.

**Table 1.** Trainer's Motion Error Coordinates

Position	Kinect	Methodology of This Article	Method 1	Method 2
Head	1.26	1.23	2.27	1.92
Neck	-0.21	0.00	0.00	-1.17
Left shoulder	3.24	3.17	3.82	4.08
Left elbow	1.58	1.58	1.15	1.89
The left hand	11.57	10.46	13.58	6.93
The right hand	15.78	15.25	17.31	10.09
Right shoulder	3.76	3.17	2.25	3.58
Right knee	-4.78	-4.29	1.09	0.00
Right foot	-27.63	-26.35	-11.59	-36.59
Left knee	3.82	3.71	2.64	2.18
Left foot	16.62	15.43	12.64	20.71

To further demonstrate the advantages of the training method, a comparison of motion trajectories in dance movements is presented in Figure 2. As shown in the figure, the method described in this paper yields motion trajectories that are closer to those collected by the Kinect than other methods, indicating that the proposed method can accurately capture the trainee's motion trajectories.



**Figure 2.** Trainingee's Motion Trajectory Detection Diagram

Further analysis of joint angle tracking in motion capture yielded the experimental results shown in Table 2. The method proposed in this paper achieves higher accuracy in angle tracking compared to other comparison methods. Based on the combined results in Table 1, Figure 2, and Table 2, it can be concluded that the dance motion capture method proposed in this paper provides higher accuracy in tracking the trainee's motion and position, enabling more precise comparison with the standard motion and facilitating error correction during assisted training.

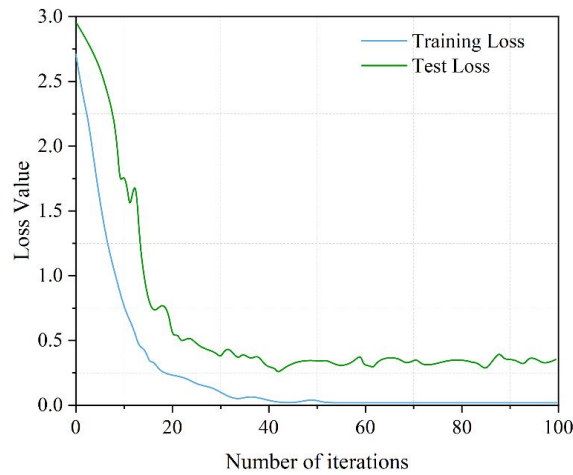
**Table 2.** Comparison of joint angle values between trainees

Position	Kinect	Methodology of This Article	Method 1	Method 2
Left wrist-Left elbow-Left shoulder	175.1	169.9	206.2	134.8
Right wrist-Right elbow-Right shoulder	106.2	104.1	134.0	190.0
Neck-Right shoulder-Right elbow	145.9	144.8	179.2	142.9
Neck-Left shoulder-Left elbow	184.3	186.2	166.9	173.7
Lumbar region-Left knee-Left ankle	165.8	167.7	179.1	190.1
Lumbar region-Right knee-Right ankle	113.1	112.5	116.3	124.1

### 4.3. Motion Recognition Experiment Results and Analysis

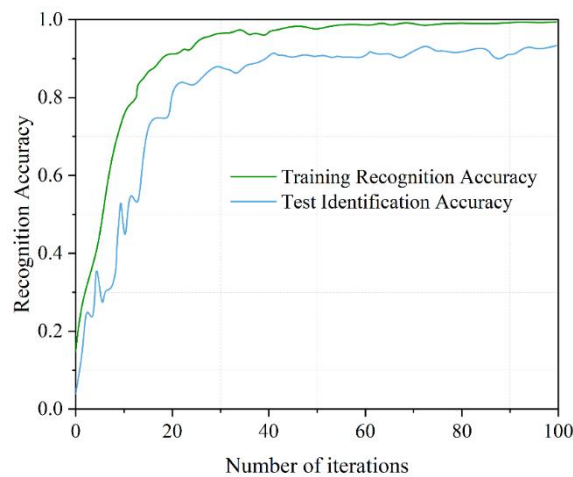
To evaluate the performance of the proposed 3D CNN model in automatically recognizing the dance movements of the Yao Long Drum Dance, the collected dataset of Yao Long Drum Dance movements was divided into a training set and a test set in a 3:1 ratio, and the proposed 3D CNN model was used for recognition. Figure 3 shows the evolution of the model's loss values on the training and test sets as the number of iterations increased. As shown in the figure, after 45 iterations, both the

training and test loss values remained essentially stable, indicating that the model had achieved the desired training results.



**Figure 3.** The curve showing how the loss value changes with the number of iterations

Figure 4 shows the curves of recognition accuracy for the proposed 3D CNNs model on the training and test datasets as a function of the number of training iterations. It can be seen that after 45 training iterations, the model’s training and test recognition accuracies remain essentially stable, reaching 98.31% and 89.91%, respectively. This demonstrates that the proposed 3D CNN model exhibits excellent recognition performance for the dance movements of the Yao Long Drum Dance.



**Figure 4.** The curve showing how recognition accuracy changes with the number of iterations

To validate the recognition performance of this model, the experiment identified five classic dance movements from the Yao Long Drum Dance: “Lotus Crown,” “Crossing the Ocean,” “Journey Down the River,” “Golden Rooster Spreads Its Wings,” and “Turning Heaven and Earth.” The recognition results are shown in Table 3. The results show that the recognition rates for the five classic dance movements were 99.89%, 97.51%, 96.38%, 98.05%, and 97.58%, respectively. The recognition accuracy remained at or above 96% for all movements, indicating a high level of precision. This demonstrates that this model can accurately recognize different dance movements.

**Table 3.** Recognition rates for 5 classic actions

	<b>Choreographic elements</b>	<b>Recognition Accuracy</b>
Action 1	Lotus flower cap	99.89%
Action 2	travel far away across the sea	97.51%
Action 3	Xingjiang River	96.38%
Action 4	The Golden Rooster spreads its wings.	98.05%
Action 5	Through all the changes in time and space...	97.58%

To further validate the performance of this model, the experiment used a 3D CNN model to recognize 10 typical movements of the Yao Long Drum Dance. The confusion matrix for the recognition results is shown in Figure 5. As can be seen from the figure, although the model's recognition accuracy for typical movement 6 is relatively low, it can still recognize most of the typical movements of the Yao Long Drum Dance, with an overall recognition accuracy of 96%. This demonstrates that the proposed 3D CNN model has practical applicability for recognizing dance movements in the Yao Long Drum Dance.

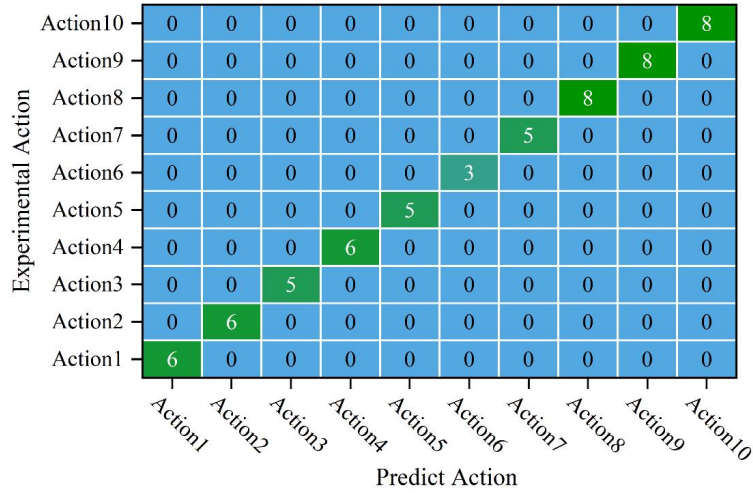


Figure 5. Identification Result Confusion Matrix

To further validate the effectiveness and superiority of the proposed 3D CNN model, we conducted experiments comparing the recognition accuracy of the proposed model with that of other commonly used recognition models—C3D and Pose3D—on the collected Yao ethnic group's Long Drum Dance pose dataset. The results are shown in Table 4. As shown in the figure, the proposed 3D CNN model achieved the highest recognition accuracy for the Jingbian Donkey-Running Dance poses, reaching 96.45%. Compared to the benchmark models, recognition accuracy was improved to varying degrees. This demonstrates that the proposed 3D CNN model exhibits a certain degree of effectiveness and superiority in recognizing the poses of the Yao Long Drum Dance.

Table 4. Comparison of identification accuracy among different recognition methods

Model	Recognition Accuracy
C3D	87.37%
Pose3D	90.17%
3D CNNs	96.45%

## 5. Conclusion

This paper addresses the challenges in preserving the Yao Long Drum Dance by collecting dance data. Using a Kinect sensor, data for 15 typical movements of the Yao Long Drum Dance were captured to construct a dataset of these movements, which was then analyzed using a 3D CNN method for recognition. Experiments demonstrate that the motion capture method described in this paper can effectively assist in dance training and provide timely error correction, proving to be reasonably effective. The 3D CNN model achieved accurate automatic recognition of the Yao Long Drum Dance movements, with a recognition accuracy of 96.45%. The method designed in this paper can capture and record the movements of the Yao Long Drum Dance in real time, and automatically recognize and analyze them. When dance movements are accurately recognized, classified, and stored, cultural memory gains the potential to be transformed from individual heritage into public cultural resources, thereby facilitating the construction of cultural identity and the strengthening of ethnic community consciousness on a broader societal level.

## References

1. Ning, B., Omar, R., & Ye, Y. (2025). Intergenerational transmission or cultural loss? The continuity of ethnic identity among the Yao minority in China. *Qualitative Research Journal*.
2. Yuan, X., Liang, Y., Yang, J., Ma, X., & Jiang, Y. (2025, June). Conceptual Definition and Contemporary Value of Ethnic-Specific Necessities for Zhuang, Dong, Miao, and Yao Ethnic Groups in the Nanling Corridor: A Belt and Road Perspective. In *International Conference on Human-Computer Interaction* (pp. 393-402). Cham: Springer Nature Switzerland.
3. Cen, X., Shen, W., Saensouk, P., Srisuk, P., Thammarong, W., Saensouk, S., & Jitpromma, T. (2026). Ethnobotanical study of wild edible plants used by the Yao people in Lingyun County, China. *Journal of Ethnobiology and Ethnomedicine*.
4. Li, R. (2025). Etiquette and dance—an analysis of the cultural phenomenon of the etiquette and custom dance of the ethnic minorities in Southwest China. *Mediterranean Archaeology and Archaeometry*, 25(2).
5. Xiaorong, H. (2022). Ethnic minorities in socialist China: development, migration, culture, and identity. In *Ethnic Minorities in Socialist China: Development, Migration, Culture, and Identity* (pp. 1-10). Brill.
6. Zhang, Y.. (2021). Entrepreneurs of the national past: the discourse of ethnic indigeneity and indigenous cultural writing in china. *positions: asia critique*, 29, 423 - 450.
7. Xu, M., Xin, Z., & Tan, Y. (2024, August). A generative approach to image generation of Yao Long Drums based on Cultural Gene analysis. In *2024 International Conference on Culture-Oriented Science & Technology (CoST)* (pp. 146-150). IEEE.
8. Zhao, X. (2025). The People's Spirit: The Source of Creation—A Study of Lingnan Dance Styles. *Journal of Sociology and Education*, 1(10).
9. Shershova, T., & Chaika, V. (2024). Cultural memory and memory culture. *Humanities Studios: Pedagogy, Psychology, Philosophy*, 1(12), 188-195.
10. Tucker, D. M., & Luu, P. (2022). The Evolution and Dissolution of Cultural Memory. In *Cultural Memory* (pp. 78-92). Routledge.
11. Crane, D. (2016). Culture and globalization: Theoretical models and emerging trends. *Global culture*, 1-25.
12. Shim, J. H. (2026). Localizing Sagehood: Cultural Memory of Yao, Shun, and Yu in Shanxi. *Acta Orientalia Academiae Scientiarum Hungaricae*, 062-2025.
13. Luo, Y. (2015). News translation as a site of framing Chinese identity: The case of Yeeyan Sport. *Ethnicities*, 15(6), 829-847.
14. Zhang, M., Chen, F., & Mo, Y. (2026). Analyzing Multimodal Teaching Behaviors in College Dance Education: A Technology-Integrated Approach with Lag Sequential Analysis. *Journal of Educational Technology and Innovation*, 8(1), 1-11.
15. Liu, W., Rattanachaiwong, N., & Zhang, Y. (2024). Reflections on the Cultural Characteristics and Stage Creation of Lingnan Folk Dance. *Journal of Dhamma for Life*, 30(3), 491-501.
16. Yu, Q., Vongphantuset, J., Geng, S., & Wang, Y. (2023). Creative application of Yao ethnic festival portraits in Souvenir design. *Asian Journal of Arts and Culture*, 23(1), e260962-e260962.
17. Varutti, M. (2010). The Politics of Imagining and Forgetting in Chinese Ethnic Minorities' Museums. *Outlines. Critical Practice Studies*, 12(2), 69-82.
18. Cai Laisi, Liu Dongwei & Ma Ye. (2021). Placement Recommendations for Single Kinect-Based Motion Capture System in Unilateral Dynamic Motion Analysis. *Healthcare*, 9(8), 1076-1076.
19. Fazliddin Anvarov, Dae Ha Kim & Byung Cheol Song. (2020). Action Recognition Using Deep 3D CNNs with Sequential Feature Aggregation and Attention. *Electronics*, 9(1).

#### **About the Author**

**Yang Zhang** was born in Shenhe District, Shenyang City, Liaoning Province, P.R. China, in 1984. He is currently

pursuing his Doctor of Arts degree at Krirk University, Thailand. His research interests include art theory, artistic anthropology, dance studies, Chinese ethnic and folk dance research, dance therapy, and dance aesthetic education.

**Dong Shu** was born in Jiangsu, China, in 1987. From 2010 to 2013, he studied in Xi'an Academy of Fine Arts and received his bachelor's degree in 2013. From 2013 to 2016, he studied in Northwest Minzu University and received his Master's degree in 2016. From 2021, he is pursuing a doctoral degree at Krirk University. he has published seven papers, one of which has been indexed by AHCI and one of which has been indexed by SCOPUS. His research interests are art and design studies, cultural dissemination and intangible cultural heritage.