

Received July 15, 2020, accepted July 28, 2020, date of publication August 4, 2020, date of current version September 1, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3014157

Labanotation Generation From Motion Capture Data for Protection of Folk Dance

JIAJI WANG^{ID}, ZHENJIANG MIAO, (Member, IEEE), NINGWEI XIE, WANRU XU, AND ANG LI

School of Computer and Information Technology, Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China

Corresponding author: Jiaji Wang (12112069@bjtu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61672089, Grant 61273274, and Grant 61572064, and in part by the National Key Technology Research and Development Program of China under Grant 2012BAH01F03.

ABSTRACT Labanotation is one of the well-known notation systems for the documenting and archiving of human motion. It plays a powerful role in dance protection, choreography analysis, and so on. Recently, researchers are committed to using computer technology to automatically generate Labanotation rather than manually drawing. However, the existing generation methods cannot deal with the various changes in motion data, such as different scales, angles, motion modes and limbs. In this paper, we aim to generate Labanotation from motion capture data acquired through real folk dance performances. The main steps include feature extraction, motion segmentation and unit movement analysis. Firstly, a normalized feature named Lie group feature is extracted, which can cope with the challenges of different scales and angles in motion data. Secondly, in order to divide motion with different modes into unit fragments for further recognizing, we propose a segmentation method that combines the speed threshold and the region partition. Thirdly, to generate Laban symbols of unit movements for different limbs, two kinds of neural networks are used for the analysis. On the one hand, LieNet, a powerful network for analyzing time series data based on Lie group structure, is utilized to recognize the lower limb movements. On the other hand, extreme-learning machine, a single hidden layer feedforward neural network, is used to identify the upper limb postures. Experimental results demonstrate that our method of feature extraction, motion segmentation and unit movement analysis achieves better results than the previous works, which makes the generated Labanotation score more reliable.

INDEX TERMS Labanotation, motion segmentation, unit movement analysis, LieNet, extreme-learning machine.

I. INTRODUCTION

Folk dance is not only the crystallization of culture, but also an important part of intangible cultural heritage. With the passage of time, many of these dances are facing the risk of transmission loss, so the protection of folk dance has become an urgent task. Compared with videos, dance notation is an effective approach for recording human motion. Therefore, the generation of dance notation by computer technology is an important way to protect folk dance.

When it comes to recording methods of dance movements, people generally think of using intuitive video. However, it often happens that part of the body cannot be seen in the video. Therefore, the video cannot record the three-dimensional human motion completely. Similar to the function of music score, dance score is an appropriate recording method for three-dimensional human motion. Unlike video,

dance score can record human motion completely. In addition, the scores are saved in the form of paper or digital pictures, which not only takes up less storage space, but also popularizes the way of reading and communication.

However, it is difficult to obtain dance score. Taking Labanotation score, which is widely used in the world, as an example, its main acquisition method is manual record. Compared with a large number of folk dances in the world that need to be protected and recorded, this way of artificial acquisition is too slow to meet the demand. Therefore, the generation of Labanotation by computer technology has shown great advantages due to its high efficiency. Naturally, this topic has gradually entered the vision of researchers [1]–[8].

The automatic generation of Labanotation is an interdisciplinary subject of dance and computer technology, which is still in its infancy. Although video recording has existed for a long time, the two-dimensional video cannot be used as a bridge between the computer technology and Labanotation score for analyzing the three-dimensional human motion.

The associate editor coordinating the review of this manuscript and approving it for publication was Joewono Widjaja^{ID}.

Therefore, before the technology of 3D human motion capture is mature, there are no sufficient conditions for the research of automatic generation of Labanotation.

In general, the steps of generating Labanotation are as follows: first, process human motion capture data and extract feature; second, carry out motion segmentation to get unit movements that can be represented by Labanotation symbols; third, conduct operation of unit movement analysis to recognize upper and lower limb movement; finally, generate Labanotation score. The flow chart is shown in Figure 1. Due to the various changes in 3D human motion capture data, the existing generation methods cannot deal with the motion data robustly. Therefore, we put forward a new scheme for the three steps of generation procedure.

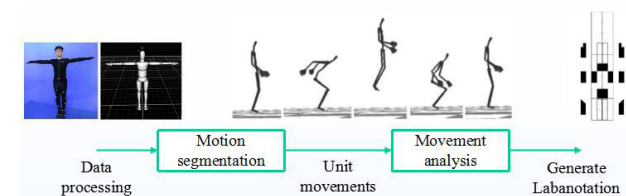


FIGURE 1. Flow chart of Labanotation generation from motion capture data.

In order to solve the problem that existing feature representation is sensitive to scale and angle, we propose to use lie group feature to represent motion data. The previous works [3], [6] used three dimensional coordinates of human joints as the feature. However, the coordinates can neither deal with the data variety of people with different shapes (tall, short, fat and thin) nor with the influence of different directions in data caused by various shooting angles. To address this issue, we introduce the Lie group feature. This feature describes the angular relationship between joints, so it not only standardizes the data variety of people, but also eliminates the influence of various shooting angles. In addition, the Lie group feature in Riemannian space is also useful to distinguish the subtle differences between similar motions.

In order to solve the problem that the existing segmentation methods are sensitive to the motion mode, we propose a method that combines the speed threshold and the region partition. For motion segmentation, to obtain the unit movements of the upper and lower limbs from human motion, a challenge is how to deal with various motion modes, such as fast changing and gentle motion. In previous studies [3], [4], [6], the methods based on the change of velocity is used to segment human motion. However, these methods cannot deal with smooth and slow motion because there is no significant change in the speed. In order to solve this problem, we combine speed threshold and region partition to analyze motion from time and space respectively, which makes the segmentation algorithm more robust. Furthermore, we match the segmentation results with the time units in Labanotation to ensure that the starting and ending positions of the unit movements are standard and accurate.

In order to solve the problem that the existing movement analysis methods are sensitive to different limbs, we propose to use two kinds of neural networks to deal with the motion data respectively according to the different properties of the upper and lower limb movements. In Labanotation, movement analysis of the lower limb focuses on the process and the upper limb focuses on the final posture. The previous works [3], [4], [6] just used one method to process both upper and lower limb motion, which is not appropriate. In this work, we propose to utilize two neural networks to recognize the corresponding Laban symbols of two kind of unit movements according to their different properties. For lower limb movements, we use the powerful spatiotemporal analysis function of LieNet [9] to recognize different categories of movements and assign them to relevant Laban symbols. For upper limb postures, we use extreme-learning machine as a multi-class classifier to identify the corresponding symbols.

Experiments show that compared with the previous works, our methods of feature extraction, motion segmentation and unit movement analysis have achieved better results. Therefore, the generated Labanotation scores will be more reliable. In summary, the contributions of this paper are as follows:

- Lie group feature is introduced to the Labanotation generation to deal with the challenge of both different appearances of persons and various directions of movements.
- A robust motion segmentation method based on speed threshold and region partition is proposed. It analyzes the problem from two aspects of time and space. Furthermore, the time units in Labanotation is used to refine the segmentation result.
- According to the different characteristics of upper and lower limb movements defined in Labanotation, we propose to use LieNet to recognize lower limb movements, and use extreme-learning machine to identify upper limb postures.
- Based on the above architecture, we realize the idea of protecting folk dance by generating Labanotation from motion capture data. Experimental results show that our method performs better than the previous works for generating Labanotation scores.

The rest of this paper is arranged as follows. Section II introduces the background of Labanotation and motion capture. Section III describes the related works of combining Labanotation with computer technology. In Section IV, we describe the proposed method in detail, including feature extraction, motion segmentation and unit movement analysis. In Section V, we conduct experimental analysis, including quantitative evaluation based on two motion capture datasets and qualitative evaluation from users. Section VI contains the conclusion and future work.

II. PRELIMINARY

A. LABANOTATION

Labanotation is a notation system designed by Rudolf von Laban, an Austro-Hungarian choreographer and dancer in the

20th century, which is used to record and analyze human motion [10]. It is made up of two parts: structure and notation. Figure 2 shows a four-page instance of Labanotation. The reading rules are as follows: in the page, read from the bottom up; between the pages, read from left to right.

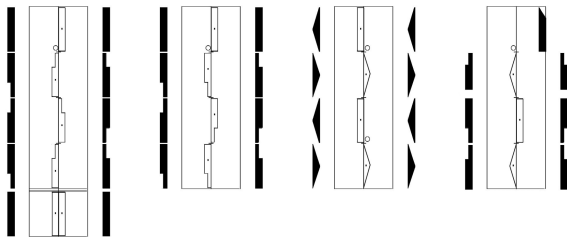


FIGURE 2. A 4-page Labanotation instance.

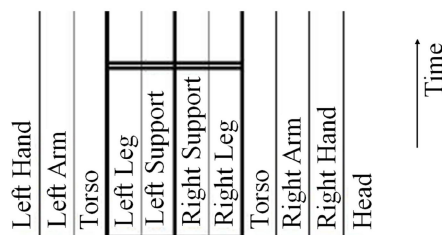


FIGURE 3. Structure of Labanotation.

Figure 3 displays the structure of Labanotation. The vertical dimension shows the time. The horizontal dimension consists of 4 to 11 columns, each corresponding to a part of the human body. The number of columns can be changed according to user requirements. If a notation appears in a column, it indicates that the corresponding body part is moving.

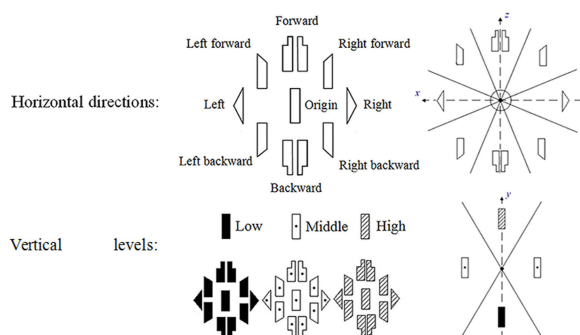


FIGURE 4. Basic notations and the space division.

Figure 4 shows 27 basic notations and corresponding space division. Each notation defines a class of unit movement. For notations, the form represents the horizontal direction and the decorative pattern represents the vertical level. Please refer to [10] for more information.

B. MOTION CAPTURE

The motion capture device can record the motion track of human body in three-dimensional space. In 1973, Gunnar

Johansson first proposed the technology of motion capture [11]. At present, the technology has been widely used in human-computer interaction, sport analysis, film production and other fields.

The format of capture data includes Bio-vision Hierarchy (BVH), Hierarchical Translation Rotation (HTR), Acclaim Skeleton File/ Acclaim Motion Capture data (ASF/ AMC), etc [12]. In this work, BVH format is chosen because of its universality. The BVH document consists of two parts. The first part is the structural information of human skeleton joints. The second is to use Euler angel data to save the motion of all joints [13].

III. RELATED WORK

The application of combining Labanotation with computer technology mainly includes three aspects.

The first is to write Labanotation through software. In recent years, the most widely used editing software is Laban Writer [14]. It is promoted by International Council of Kinetography Laban (ICKL). The similar software includes Labanatory [15], Calaban [16] and LED&LINTEL [17]. Compared with drawing on paper, the software can improve the convenience and save time. However, the professionals still need to spend a lot of time. Therefore, this kind of research cannot solve the problem of low efficiency of writing Labanotation.

The second is to drive the human model through Labanotation. Laban Editor [18], Laban Dancer [19] and Life Forms [20] are the tools to convert Labanotation into human animation. However, the significance of this research lies in visual display, rather than improving the efficiency of Labanotation generation.

The third is to research the automatic generation of Labanotation. Hachimura and Nakamura [1] first proposed an approach to generate Labanotation through human motion. On the basis of spatial analysis, the Labanotation of upper limb movement is generated. This method can only handle the simple upper limb motion for unit movement analysis. Chen *et al.* [2] analyzed human motion by designing rules. However, they ignored the pause in human motion. So, its motion segmentation is not good. Guo *et al.* [3] added reserved notations to their spatial rule-based approach. Its feature of motion data is three dimensional coordinates, which is not robust. Worawat *et al.* [4], [5] built an instrument named “GenLaban” which can generate Labanotation automatically. GenLaban chose the key frames of human motion, and then analyzed human posture with curvature. However, the rhythm of Labanotation is not taken into account when selecting key frames. So, the result of motion segmentation is unsatisfactory. These methods attempt to formulate rules to map movements to Laban symbols. We call them rule-based methods. However, the human movement is very flexible, and it is difficult to be described with rules. If the motion data is not standard or contains noise, these methods may not produce the correct results. In the approach presented by Zhou *et al.* [6], [7], the unknown motion is compared

with the established template library by template matching to identify the category. However, the feature of motion data is coordinates, which lacks robustness. This affects the generality of template and limits the practicability. Li *et al.* [8] proposed a method to model human motion based on Hidden Markov chains. However, not all movements are consistent with the observation independence hypothesis of Hidden Markov Model. Besides, it can only handle the lower limb motion for unit movement analysis. These two methods are more flexible and superior than rule-based ones, but their application is limited due to their shortcomings in the feature extraction and unit movement analysis.

Compared with the previous methods, we have made improvements from the three aspects of feature extraction, motion segmentation and unit movement analysis. In order to make the feature of motion data robust, we use the Lie group feature to deal with the data variety of people and different shooting angles from the capture device. For motion segmentation, we use speed threshold and region partition to segment the motion from two aspects of time and space. In addition, the rhythm is used to refine the segmentation result. For unit movement analysis, we recognize the movements by two neural networks according the different characteristics of upper and lower limb motion. In experiments, we obtain better performances for the feature, segmentation and movement analysis.

IV. PROPOSED METHOD

In this work, the procedure of generating Labanotation consists of three main steps: data processing, motion segmentation and unit movement analysis. Data processing includes data transformation and feature extraction. The purpose of motion segmentation is to divide a long period of human motion into unit fragments. Unit movement analysis recognizes the corresponding Laban symbols of each movement. Finally, the Labanotation score of motion capture data is generated. The flow chart of our method is shown in Figure 5.

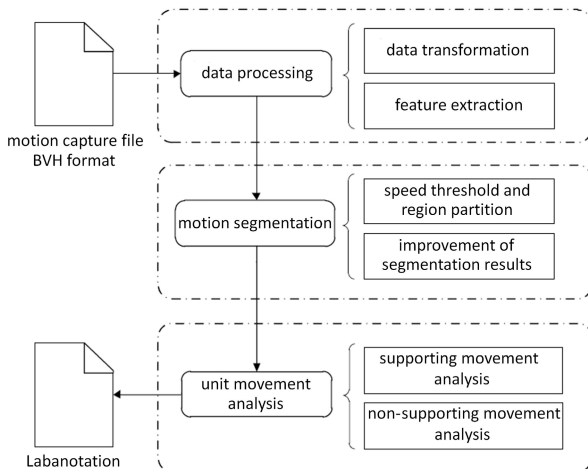


FIGURE 5. The Labanotation generation flow chart of our method.

A. DATA PROCESSING

For motion data in BVH format, the human skeleton is a tree structure, and the motion information is recorded by the rotation of joints as Euler angle in each frame. In human skeleton, the nodes represent the joints in human body, and each node has a definition of the initial position offset with its parent node. The skeleton contains 26 nodes in this work, as shown in Figure 6. This is the same as that used in our previous work [21].

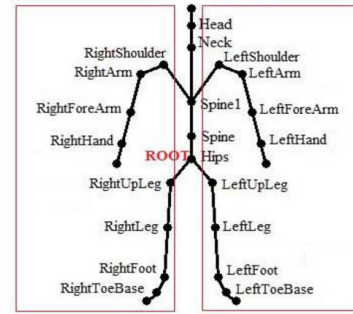


FIGURE 6. Human skeleton [21] (since the skeleton is facing us, the left and right sides are opposite).

1) DATA TRANSFORMATION

In BVH file, Euler angle represents the relative rotation of the node and its parent node, excluding the intuitive position. However, motion segmentation needs to analyze the location and speed of nodes. Therefore, we transform the Euler angle into three-dimensional coordinates for further segmentation.

For a non-leaf node J_p , its child node J is located by the information of J_p . Assuming the original position offset of child node J is (x_0, y_0, z_0) , then after a rotation of node J_p , the new offset (x_c, y_c, z_c) of child node J can be calculated by (x_0, y_0, z_0) and the angular displacement matrix $\mathbf{M}$ of node J_p , as shown in Equation (1).

$$\begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} = \mathbf{M} \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} \quad (1)$$

For a node chain, $J, J_1, J_2 \dots J_r$ (the former node is the child of the latter one and J_r is the root node), the position offset of node J in the chain can be obtained by calculating Equation (1) layer by layer until the root node. As a result, the offset of any node can be calculated. In this way, we convert the Euler angle into coordinates. For more information about the format conversion, please refer to [21].

2) FEATURE EXTRACTION

Considering the problem that the same human movement will result in dissimilar data because of people with different shapes and various directions caused by random shooting angles, we extract features for unit movement analysis to deal with this problem. In order to achieve robust representation of motion data, we introduce a normalized feature, Lie group feature, in the field of Labanotation generation.

For two adjacent nodes in human skeleton, the Lie group feature records the relative angle relationship between nodes. A set of 3×3 rotation matrices in $\mathbf{R}^3$, the matrix of Lie group feature, constitutes a special orthogonal group SO_3 . Then, the motion sequence of a series of skeletons can be seen as a curve on the Lie group $SO_3 \times \dots \times SO_3$. On the one hand, because the relative angle between joints is independent of the length of bone, this feature normalizes different body shapes. On the other hand, the relative angle is completely determined by the joints of human body, so it also addresses the issue of various shooting angles.

Assuming the Euler angle of node J is (r, p, y) , the corresponding rotation matrix $\mathbf{R}_J$ representing the Lie group feature is calculated as follows:

Based on the human skeleton in Figure 6, assuming that the motion sequence contains N frames, then the Lie group feature of this sequence is represented by $(26-1) \times N$ rotation matrices, which has a differentiable Riemannian manifold structure [9].

B. MOTION SEGMENTATION

In the long-term motion sequence, there are a series of unit movements, each corresponding to a Laban symbol. In order to generate the Labanotation score, it is necessary to segment the motion into unit fragments. We propose to use speed threshold and region partition to segment the motion data from the perspective of time and space respectively. After segmentation, the results are refined according to the rhythm of Labanotation to ensure that the generated symbols are standard.

1) SPEED THRESHOLD AND REGION PARTITION

Considering that the change of motion speed is sometimes very fast, sometimes very slow, we propose to divide the human motion (leg and arm motion) by speed threshold and region partition. The treatment of the lower and upper limbs is the same.

When the motion speed changes rapidly, i.e. the motion has obvious pauses, the speed threshold has a good effect. Take the swing of the arm as an example: let's say that the arm first moves to the left and then to the right. When the direction of motion changes, the arm will pause with a local minimum speed, and the cutting point of the motion is at the position where the speed is the minimum. The same goes for leg motion. Whenever the speed of the limb drops below the threshold, it means that a new movement will be started. For the selection of speed threshold, considering that the limb length of people with different shape is quite different, so in this paper, the angular speed is chosen to replace the speed of the limb end. According to the empirical data in references [7] and [23], the angular speed threshold in this paper is 0.18 rad/s. Furthermore, to reduce noise interference, fragments are ignored if they last less than 0.1 seconds.

However, the speed threshold method is not applicable to gentle motion, such as Taijiquan (another name is Tai Chi, a kind of traditional Chinese shadow boxing) [24]. In this

case, we use region partition to deal with the motion with approximate uniform velocity. The Laban direction consists of 9 horizontal directions and 3 vertical levels (shown in Figure 4), so it is combined into 27 subspaces. In Figure 4, for horizontal and vertical division, the angles between two adjacent solid lines are 45 degrees and 60 degrees, respectively. In human skeleton, we use vectors to represent the arms and legs. For a limb, we obtain the vector $\mathbf{V}$ by subtracting the coordinates of its root node from the endpoint. Suppose that vector $\mathbf{V}_1$ represents the arm before a movement and vector $\mathbf{V}_2$ after the movement, and then generates angle θ between two vectors. If angle θ exceeds 180 degrees and there is still no segmentation by speed threshold, we will use region partition to cut the motion when vector $\mathbf{V}$ moves from one subspace to another.

2) IMPROVEMENT OF SEGMENTATION RESULTS

After segmentation by speed threshold and region partition, we get a set of unit movements. However, the segmentation result does not take rhythm into account. For Labanotation, rhythm is a very important property. So we modify the result to match the rhythm.

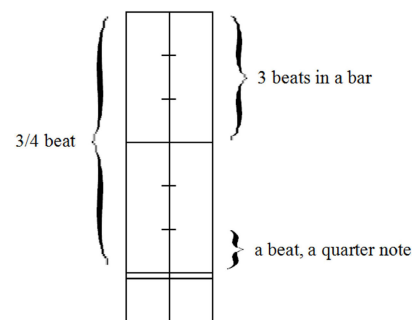


FIGURE 7. Diagram of 3/4 beats rhythm.

Similar to the rhythm of a music score, the bar line and beat line are used to represent the rhythm in Labanotation, too. Take the rhythm of 3/4 beats as an example, see Figure 7. The 1/4 note is a beat, and a section contains three beats. Generally speaking, in Labanotation a quarter of beat is the smallest unit. In order to make the symbols consistent with the time unit, we choose the closest 1/4 beat, 1/2 beat or beat line as the starting or ending position for each symbol. In this way, the generated symbols no longer have irregular positions, thus becoming more standard and accurate.

In addition, as a dance notation, Labanotation is not to record movements for any specific person, but to write down the basic idea of movements [4]. Therefore, our conception of “modify segmentation result to match the rhythm” improves the generality of the generated Labanotation, which is quite appropriate.

C. UNIT MOVEMENT ANALYSIS

Movement of the lower limbs causes the moving of human body. In Labanotation, its analysis focuses on the whole pro-

cess. In contrast, movement of upper limbs only changes its own position, which focuses on the final posture. Therefore, different analysis methods are needed for lower limb and upper limb movements. According to the different definitions for the supporting movements of lower limb and the non-supporting movements of upper limb in Laban theory, we adopt two analysis methods. For one thing, LieNet, a neural network with powerful spatiotemporal analysis function, is used for supporting movements. For another, extreme-learning machine is for non-supporting ones.

1) SUPPORTING MOVEMENT ANALYSIS

LieNet, a novel neural network architecture, is designed for analyzing the process of 3D human motion. Similar to convolution network, LieNet has one or more fully connected convolution layers called RotMap and corresponding pooling layers called RotPooling. Based on the Lie group feature, Equation (3) and (4) represent the functions of RotMap and RotPooling respectively.

$$\begin{aligned} f_r^{(k)}((R_1^{k-1}, R_2^{k-1}, \dots, R_K^{k-1}); W_1^k, W_2^k, \dots, W_K^k) \\ = (W_1^k R_1^{k-1}, W_2^k R_2^{k-1}, \dots, W_K^k R_K^{k-1}) = (R_1^k, R_2^k, \dots, R_K^k) \end{aligned} \quad (3)$$

$$\begin{aligned} f_l^{(k)}((R_1^{k-1}, R_2^{k-1}, \dots, R_K^{k-1})) \\ = (\log(R_1^{k-1}), \log(R_2^{k-1}), \dots, \log(R_K^{k-1})) \end{aligned} \quad (4)$$

In the equations, k represents the k -th layer; K is the number of rotation matrices; $(R_1^{k-1}, R_2^{k-1}, \dots, R_K^{k-1})$ is the Lie group feature; and $(W_1^k, W_2^k, \dots, W_K^k)$ are transformation matrices. The calculation of logarithm mapping $\log(\cdot)$ is detailed in [9].

In this work, the LieNet structure is composed of 3 RotMap layers, a logarithm mapping layer, and other regular layers in the neural network, including a rectifier linear unit layer (ReLU), a fully-connected layer (FC) and a SoftMax output layer, shown in Figure 8. The dimension of weight in FC layer is $c \times d_{k-1}$, where c is the category number of lower limb unit movements and d_{k-1} is corresponding to the output in logarithm mapping layer. Specially, there is no RotPooling layers because the scale of features of a lower limb is much less than that of the whole body.

In our human skeleton, one leg contains 5 nodes. The relationship between two adjacent nodes is expressed by a rotation matrix. So, for one frame in the motion sequence, the Lie group feature of one leg contains 4 rotation matrices. We train the LieNet with tagged data. After training, given an unknown movement e to network, the category label of movement e can be recognized as:

$$\text{label}(e) = \arg_j \max h_j(e) \quad (5)$$

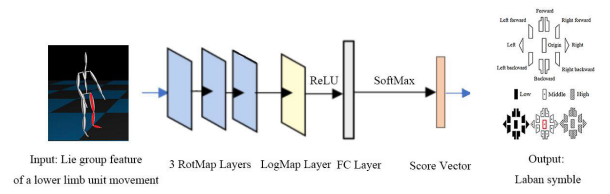


FIGURE 8. The architecture of LieNet.

where $j = 1, 2, \dots, 27$, which represents the j -th output node as well as the type number of movements; and $h_j(e)$ indicates the probability of movement e belonging to the j -th category. In this way, we obtain the category of each supporting movement and confirm the Laban symbol accordingly.

2) NON-SUPPORTING MOVEMENT ANALYSIS

Extreme-learning machine (ELM) is a single hidden layer feedforward neural network [25], which is capable of classifying the posture of human motion data.

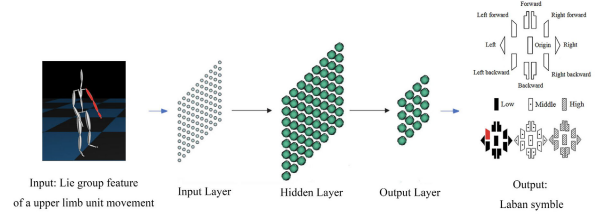


FIGURE 9. The architecture of ELM.

The architecture of the ELM contains a input layer, a hidden layer with 300 neurons and a SoftMax output layer, shown in Figure 9. The goal of training for ELM network is to obtain the minimum training error and the minimum parameter matrix weight norm. For L training samples $\{x_i, y_i\}$, where $i = 1, \dots, L$; x_i is the Lie group feature; y_i is the category label, then the output of the n -th hidden neuron is $g(x; w_n, b_n) = g(x \cdot w_n + b_n)$, where $g(\cdot)$ is a nonlinear activation function, multivariate quadratic function; w_n is the weight vector between the input layer and the hidden layer; b_n is the offset. Therefore, the mapping matrix from the input layer to the hidden layer is:

$$\text{Mapping}(\mathbf{x}) = [g(\mathbf{x}; w_1, b_1), \dots, g(\mathbf{x}; w_{300}, b_{300})] \quad (6)$$

where the activation function is:

$$g(\mathbf{x}; w_n, b_n) = \sqrt{1 + (\mathbf{x} \cdot w_n + b_n)^2} \quad (7)$$

Assuming that u_{nm} is the weight between the n -th hidden neuron and the m -th output node, then the relationship between

$$\begin{aligned} \mathbf{R}_J &= \mathbf{R}_{z(-r)} \mathbf{R}_{x(-p)} \mathbf{R}_{y(-y)} \\ &= \begin{pmatrix} \cos r \cos y - \sin r \sin p \sin y & -\sin r \cos p & \cos r \sin y + \sin r \sin p \cos y \\ \sin r \cos y + \cos r \sin p \sin y & \cos r \cos p & \sin r \sin y - \cos r \sin p \cos y \\ -\cos p \sin y & \sin p & \cos p \cos y \end{pmatrix} \end{aligned} \quad (2)$$

the output node and the hidden neuron is as follows:

$$d_m(\mathbf{x}) = \sum_{n=1}^{300} u_{nm} g(\mathbf{x}; \mathbf{w}_i, b_i) \quad (8)$$

Finally, the mapping matrix of neural network from the input layer to output layer is expressed as:

$$\mathbf{D}(\mathbf{x}) = \text{Mapping}(\mathbf{x})\mathbf{U} \quad (9)$$

In the ELM network, the output matrix of the hidden layer is uniquely confirmed after the weight vector $\mathbf{w}_n$ and offset b_n are randomly determined. The training of single hidden layer neural network ELM can be transformed into the solution of linear system. Please refer to literature [26], [27] for the detailed training process.

After the training, the process of judging the category of unknown upper limb posture and corresponding Laban symbol is the same as that of lower limb.

V. EXPERIMENTS AND EVALUATIONS

In order to conduct a comprehensive analysis, both objective and subjective evaluation are used to evaluate the effectiveness of the method. At present, most of the existing Labanotation scores are recorded by experts. However, for a motion sequence, sometimes even the results recorded by two experts are not exactly the same. Therefore, the performance measures need both objective and subjective ones.

A. DATASETS

For experimental data, considering the low popularity of motion capture devices and the fewer existing motion capture databases, we construct the datasets with our motion capture device: OptiTrack [28]. It consists of 18 infrared cameras with a frame rate of 150 Hz. As a commercial equipment, its accuracy is enough to meet the experimental requirements. The construction environments is shown in Figure 10. Based on this device, our team has accumulated more than 200 pieces of motion capture data in the preliminary works [3], [6]–[8], [13], [21], [29]–[31]. Each piece of motion data lasts about 30 seconds to 2 minutes. The total number of frames for all motion data exceeds 1,500,000 frames. In this paper, we use a part of the data that has been collated. The three datasets are shown below.

Dataset A consists of 19,200 unit movements that are manually segmented and tagged. These unit movements are the supporting ones of both left and right legs. The movements of each leg contain 24 categories, including 8 horizontal directions and 3 vertical levels. Each category consist of 400 samples. For each sample, due to the different duration of time, the number of frames varies from 90 to 200. The original motion data is captured from four volunteers of different shapes (their heights range from 160 cm to 190 cm) and the shooting angles of the capture device are different, too. This dataset is available from the work [29].

Dataset B consists of 453 unit movements that are manually segmented and tagged. These unit movements are the non-supporting ones of two arms. For left and right arm,



FIGURE 10. The construction environments of database.

there are 231 and 222 unit movements respectively. Each movement contains about 200 frames, and there are 10 categories of common movements for each arm. The original motion data is captured from two volunteers, and this dataset is available from the work [7].

Dataset C consists of 116 pieces of motion data. The types of motion include: walking gait (vertical and horizontal changes of gravity center), jumping gait and arm waving, etc. This dataset covers most of the basic limb movements, including 464 unit movements of two arms and two legs (without segmented). In addition, these unit movements cover all the basic notations in Figure 4, including 9 horizontal ones (*forward, backward, left, right, left-forward, right-forward, left-backward, right-backward and origin*) and 3 vertical ones (*low, middle and high*). The original motion data is captured from two volunteers with dance basis during the work [21], and it can be opened to the public.

B. OBJECTIVE EVALUATION

In this section, we verify the effect of the proposed feature, motion segmentation and unit movement analysis respectively. At last, we compare the generated Laban symbols with the ground truth.

1) EVALUATION OF THE FEATURE

Based on dataset A, we compare our feature with the ones in other methods. The different features are input into the same neural network LieNet, and then we compare the recognition accuracy. In Table 1, the first feature represents the original 3D coordinates of the joint position. The second feature normalizes the orientation of human body based on the first feature. The purpose of the second one is to reduce the impact of various directions caused by shooting angles from the motion capture device. The third feature represents the relative vectors of the joint position based on the first feature. The goal of the third one is to weaken the influence of different shapes of people. The fourth feature normalizes the orientation of human body for the relative vectors of the

TABLE 1. Comparisons between the features in references [1], [3], [4], [8], [32], [36] and the proposed one on recognition accuracy based on dataset A.

Features	Left Leg	Right Leg	Average
Coordinates [1], [3], [4]	49.75%	51.54%	50.65%
Li et al. [8]	78.13%	77.61%	77.87%
Li et al. [8]	87.26%	90.39%	88.83%
ON-RJP [32]	91.78%	92.81%	92.30%
Euler angle [36]	93.45%	93.89%	93.67%
Proposed Feature	96.47%	96.12%	96.30%

TABLE 2. Comparisons between the segmentation methods in references [3], [31] and the proposed one on segmentation accuracy based on dataset A.

Segmentation Methods	Left Leg	Right Leg	Average
Speed Threshold [3]	40.36%	42.18%	41.27%
Rhythm Statistics [31]	58.08%	58.22%	58.15%
Spatial Segmentation [31]	64.14%	63.26%	63.70%
Proposed Method	69.83%	70.99%	70.41%

joint position. The fifth feature, Euler angle, is a set of three angle parameters used to determine the rotation position of a point. This kind of relative angle relationships can eliminate the influence of different orientation of human body and different shapes of people. However, this angle feature is too simple, and this description of data is not as comprehensive as the rotation matrix. Our Lie group feature normalizes both different shapes of people and various directions caused by shooting angles. In addition, the feature uses a set of 3×3 rotation matrices to describe data comprehensively. It can be seen from Table 1 that our feature achieves the best recognition accuracy. Our accuracy is about 96%, which is 2% higher than the Euler angle feature used in work [36]. This proves the validity of the Lie group feature for our task.

2) EVALUATION OF THE SEGMENTATION

We compare our motion segmentation method with other works, including speed threshold [3], rhythm statistics [31] and spatial segmentation [31]. Based on dataset A, segmentation accuracy is the measure. The unit movements in dataset A are manually segmented, and we use the segmentation results as the ground truth. If the segmentation error is less than 0.1 seconds, i.e. 15 frames, we consider that the segmentation is correct. The comparison results are shown in Table 2. From the table, our segmentation method achieves the best segmentation accuracy. The speed threshold and region partition in our method can segment the motion from the aspects of time and space. Therefore, our approach can do well in different kinds of motion, and this is better than the other segmentation methods.

3) EVALUATION OF THE MOVEMENT ANALYSIS

Based on dataset A and dataset B, we compare the average recognition accuracy of unit movements for lower and

TABLE 3. Comparison results of average recognition accuracy for lower limb movements based on dataset A.

Methods	Left Leg	Right Leg	Average
Rule-based [3]	64.29%	60.33%	62.31%
Template-based [6]	85.75%	83.61%	84.68%
HMM-based [32]	92.72%	92.58%	92.65%
JDM [33]	92.92%	92.90%	92.91%
JTM [34]	91.21%	89.30%	90.26%
SPF + TPF [35]	94.53%	96.82%	95.68%
Proposed Method	96.47%	96.12%	96.30%

TABLE 4. Comparison results of average recognition accuracy for upper limb movements based on dataset B.

Methods	Left Leg	Right Leg	Average
Laban space theory [1]	78.65%	77.82%	78.24%
SVM [30]	86.27%	85.89%	86.08%
Proposed Method	93.76%	94.48%	94.12%

upper limb respectively. The comparison results are shown in Table 3 and Table 4.

For lower limb, the compared methods include rule-based method [3], template-based method [6], HMM-based method [32] and three CNN- and LSTM-based methods [33]–[35]. The rule-based and template-based methods belong to the traditional ways. The accuracy of two traditional ones is lower than that of our method, which reflects the superiority of neural network LieNet. The HMM-based method uses chain structure model. However, the descriptive ability of the chain is inferior to that of LieNet. So its accuracy is lower than ours. The three CNN- and LSTM-based methods are originally designed for action recognition, which are similar to but different from our work. We applied them to our case for comparison. As shown in Table 3, our method achieves the best results in the average recognition accuracy of unit movements. This indicates that the proposed neural network architecture, LieNet based on Lie group feature, is suitable for the analysis of the unit movement data.

For upper limb, the compared methods include the ones based on Laban space theory [1] and based on SVM [30]. From Table 4, we can see that the proposed method based on extreme-learning machine is about 15% higher than the method based on Laban space theory in the recognition accuracy, which shows that compared with the fixed parameter method, the extreme-learning machine network can learn more comprehensive motion information. Compared with the method of support vector machine, our result is still about 7% higher. The extreme-learning machine is easy to train and suitable for the requirements of real-time computing tasks. Furthermore, the number of upper limb movements is small at present, the extreme-learning machine network fully plays its effectiveness in this task.

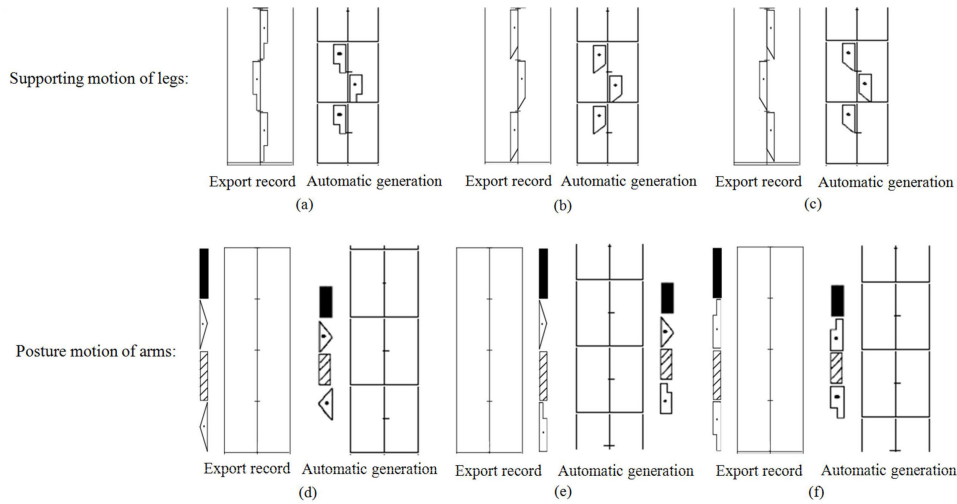


FIGURE 11. Comparison of expert records and automatic generation of Labanotation based on dataset A. The meaning of Laban symbols are as follows. (a) “backward, middle”; (b) “left-backward, middle”; (c) “right-backward, middle”; (d) “left, middle”, “origin, high”, “right, middle”, and “origin, low”; (e) “forward, middle”, “origin, high”, “right, middle”, and “origin, low”; (f) “backward, middle”, “origin, high”, “forward, middle”, and “origin, low”.

4) EVALUATION OF THE GENERATED LABAN SYMBOLS

Based on dataset C, we compare our generated Labanotation scores with those recorded by an expert. Bingyu Luo, a famous Labanotation expert in China, is invited to record the human motion in the dataset. Ms. Luo, the former director of Literary and Engineering Group of the Political Department of the Air Force of the People’s Liberation Army, has nearly 40 years of research experience, and she has translated the book Labanotation [10].

The Labanotation scores recorded by Ms. Luo and the generated ones by our method are shown in Figure 11. In the figure, there are 6 groups of comparison results, including 3 pieces of supporting motion of legs and 3 pieces of non-supporting motion of arms. We take the first and the fourth group as examples. The motion (a) is “walking backward” and there are three Laban symbols. From the bottom up in chronological order, the first symbol represents the unit movement “backward, middle” of left leg; the second symbol shows the unit movement “backward, middle” of right leg; and the third one is the same with the first. The motion (d) is “waving left arm” and there are four Laban symbols. From the bottom up, the first symbol represents the unit movement “left, middle” of left arm posture; the second symbol shows the unit movement “origin, high” of arm posture; the third symbol represents the unit movement “right, middle”; and the fourth one shows the unit movement “origin, low”. From the observation of Figure 11, it can be concluded that the generated Labanotation scores are consistent with those drawn by the expert. In other words, the generated results are correct.

C. SUBJECTIVE EVALUATION

For Labanotation scores, even if two experts record the same human motion, the results may not be exactly the

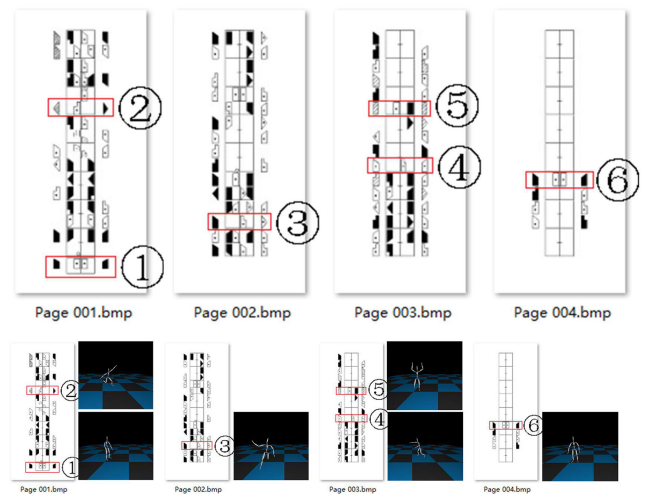


FIGURE 12. Generated Labanotation of a section of Guzi Yangge (a few places have manual modifications). From ① to ⑥ are 6 places corresponding to the original motion in Figure 13.

same. In other words, the recording task of Labanotation is partly subjective. Therefore, objective evaluation alone is not enough to comprehensively measure our results. Based on this consideration, we have done two more experiments from the subjective point of view. One is to generate Labanotation for a long-term motion sequence, and then compare the original motion with the generated Laban symbols. The other is to invite a group of graduate students with relevant research experience of Labanotation to evaluate the results of our method.

Firstly, we analyze the generated result for a long-term sequence. The description of this motion capture data is as follows:

1. Dance category: A section of Guzi Yangge, named “Xuanhaiyan”, a traditional Folk Dance in Shandong Province of China.
2. Details of data: Frame rate 150, a total of 6133 frames, BVH format.
3. Capture equipment: OptiTrack motion capture system.
4. Capture place and time: Center for Ethnic and Folk Literature and Art Development of Culture and Tourism of P.R.C., Beijing, China. October 11, 2016.

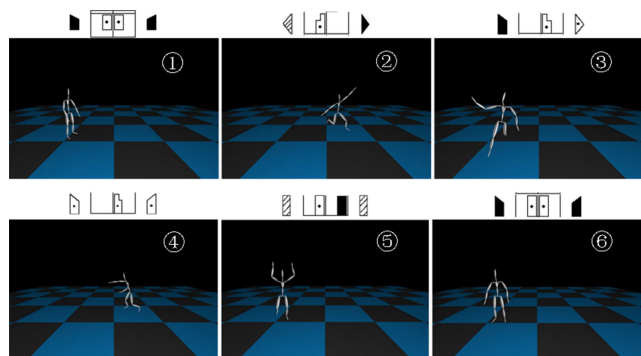


FIGURE 13. Generated Laban symbols with the corresponding original motion.



FIGURE 14. Video screenshots of Guzi Yangge that synthesized with video data (three channels) and motion capture data. There are 6 screenshots corresponding to Figure 13, each of which is a live video shot from three different angles on the left and motion capture data of two performers on the right (the upper right corner is the same with Figure 13).

For this section of folk dance, the generated Labanotation score contains 4 pages, as shown in Figure 12. The comparison between the original motion and generated notations is shown in Figure 13. The video screenshots synthesized with video data (three channels) and motion capture data (two performers) is shown in Figure 14. The analysis order of the limbs is as follows: first, analyze the supporting movement of leg, then analyze the arm posture. When analyzing the movement and posture, first say the horizontal direction and then the vertical level, for example: “*forward, middle*”, “*left-forward, high*”, “*origin, low*”, where “*forward*”, “*left-forward*” and “*origin*” represent the horizontal direction, “*middle*”, “*high*” and “*low*” indicate the vertical level. For the name and space division of all Laban symbols, please see Figure 4.

The first motion ① is the initial state. The person is standing on two legs normally, i.e., the horizontal direction is “*ori-*

gin”, and vertical level of gravity center is “*middle*”. So, the Laban symbols “*origin, middle*” are applied to indicate the legs. For arm postures, the left one is “*left-forward, low*” and the other one is “*right-forward, low*”. Motion ② shows the left leg is towing the gravity center “*forward*”, and the vertical level of gravity center is “*middle*”. So, Laban symbol “*forward, middle*” is used to represent the supporting movement of left leg. At this time, the movement of right leg is accompanied by that of left leg, which does not need to be recorded according to the rules of Labanotation. For arm postures, the left one is “*left, high*”, and the other is “*right, low*”. Motion ③ indicates the right leg produces a “*forward, middle*” shifting of gravity center. For arms, left one is “*left-forward, low*” and other is “*right, middle*”. Motion ④ shows right leg produces a “*forward, middle*” shifting of gravity center. For arms, they are “*left-forward, middle*” and “*right-forward, middle*”. Motion ⑤ indicates a jumping. The left leg supports the gravity. Similar to motion ①, “*origin, middle*” is applied to represent the supporting movement of left leg. Both arms are “*origin, high*”. Motion ⑥ shows the ending posture. It is similar to the initial one. Two legs are “*origin, middle*”. Arms are “*left-forward, low*” and “*right-forward, low*”. Through the analysis of the original motion from ① to ⑥, it is concluded that the generated notations are consistent with the human motion. Therefore, the generated Labanotation score is basically correct.

Secondly, 9 graduate students with relevant research experience are invited to evaluate and give feedback on the generated Labanotation scores. The students have studied Labanotation from six months to three years. In addition, 8 of them have published papers in this field, and 6 of them have participated in the course of Labanotation offered by Beihang University (formerly Beijing University of Aeronautics and Astronautics).

After using our generation system, each student completes a questionnaire. The questions and answers are shown in the appendix. For question one, watch videos of motion capture data and write down the Labanotation. They can record about 60%-85% of the Laban symbols themselves. In contrast, with the assist of our software, the percentage has risen to 75%-95%. For question two, 8 out of 9 students agree that the software will help with the recording task. This proves that the generation system can provide ideas for recording Labanotation. With the help of our system, most of the students can improve their completion rate and writing efficiency. For question three, the students deem that the accuracy of the generation result for basic human motion is about 87%-95%. Therefore, the generated Labanotation scores can be used as a good reference.

VI. CONCLUSION AND FUTURE WORK

Labanotation, a scientific and logical recording system of human motion, is an effective approach for protecting folk dance. Based on the data obtained from motion capture device, we propose a method of automatically generating Labanotation to record and protect the intangible cultural

heritage of performing arts. First, the Lie group feature is extracted from the motion capture data for robust representation. Second, a segmentation method based on speed threshold and region partition is proposed to segment the long-term motion sequence. In addition, we match the segmentation results with the time units in Labanotation to ensure that the generated symbols are rhythmic. As far as we can see, it is difficult for computer technology researchers to consider the rhythm of Labanotation. Third, for unit movement analysis, we train the neural network LieNet to recognize the supporting movements of lower limbs, and use extreme-learning machine as a multi-class classifier to identify the non-supporting postures of upper limbs. Finally, we obtain the Laban symbols of all unit movements and generate the corresponding Labanotation score for the motion sequence. Experimental results show that our method has a better effect on the feature, motion segmentation and unit movements analysis compared with the previous works. Furthermore, our generated results can be used as a good reference for recording Labanotation scores and reduce the workload of recording personnel.

In the future work, given that there are many symbols used to decorate the details of human motion in Labanotation, we will expand our generation system to add these decorative symbols. After that, the system will be more practical for recording folk dance and contribute to the protection of performing arts for intangible cultural heritage.

APPENDIX

QUESTIONS AND ANSWERS IN THE QUESTIONNAIRE

Question 1: (watch a piece of motion capture data and then record the Labanotation) According to your experience, what is the accuracy of your own record? With the help of our generation system, what is the accuracy?

Answer 1: About 65%-85% (by themselves); About 75%-92% (with the help of the generation system).

Question 2: For the Labanotation recording task, are you willing to choose the generated results as a reference? If so, what is the help?

Answer 2: Among 9 students, 8 are willing to use it. The “help” is summarized as follows: it has auxiliary function, such as providing ideas and shortening the time for recording; in addition, it can provide reference for ambiguous movements.

Question 3: According to your experience, what is the accuracy of the generated Labanotation?

Answer 3: About 87%-95% for basic human motion. For the motion with large amplitude, the generated result is better. In contrast, for the slow rotating motion, the effect is not ideal.

ACKNOWLEDGMENT

Jiaji Wang would like to thank Prof. B. Luo for instruction on Labanotation, and for the support in data acquisition from the Center for Ethnic and Folk Literature and Art Development of Culture and Tourism of China.

REFERENCES

- [1] K. Hachimura and M. Nakamura, “Method of generating coded description of human body motion from motion-captured data,” in *Proc. 10th IEEE Int. Workshop Robot Hum. Interact. Commun.*, Sep. 2007, pp. 122–127.
- [2] H. Chen, G. Qian, and J. James, “An autonomous dance scoring system using marker-based motion capture,” in *Proc. IEEE 7th Workshop Multimedia Signal Process.*, Oct. 2005, pp. 1–4.
- [3] H. Guo, Z. Miao, F. Zhu, G. Zhang, and S. Li, “Automatic labanotation generation based on human motion capture data,” in *Proc. Chin. Conf. Pattern Recognit.* Berlin, Germany: Springer, 2014, pp. 426–435.
- [4] W. Choensawat, M. Nakamura, and K. Hachimura, “GenLaban: A tool for generating labanotation from motion capture data,” *Multimedia Tools Appl.*, vol. 74, no. 23, pp. 10823–10846, Dec. 2015.
- [5] W. Choensawat, M. Nakamura, and K. Hachimura, “Applications for recording and generating human body motion with labanotation,” in *Dance Notations and Robot Motion*. Berlin, Germany: Springer, 2016, pp. 391–416.
- [6] Z. Zhou, Z. Miao, and J. Wang, “A system for automatic generation of labanotation from motion capture data,” in *Proc. IEEE 13th Int. Conf. Signal Process. (ICSP)*, Nov. 2016, pp. 1031–1034.
- [7] Z. Zhou, “Research on automatic generation of labanotation based on dynamic programming,” M.S. thesis, Beijing Jiaotong Univ., Beijing, China, 2017.
- [8] M. Li, Z. Miao, and C. Ma, “Automatic labanotation generation from motion-captured data based on hidden Markov models,” in *Proc. 4th IAPR Asian Conf. Pattern Recognit. (ACPR)*, Nov. 2017, pp. 793–798.
- [9] Z. Huang, C. Wan, T. Probst, and L. Van Gool, “Deep learning on lie groups for skeleton-based action recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6099–6108.
- [10] A. H. Guest, *Labanotation: The System of Analyzing and Recording Movement*, 4th ed. Evanston, IL, USA: Routledge, 2014.
- [11] G. Johansson, “Visual perception of biological motion and a model for its analysis,” *Perception Psychophys.*, vol. 14, no. 2, pp. 201–211, Jun. 1973.
- [12] T. Li, “Research on human motion editing technology based on motion capture data,” Ph.D. dissertation, Huazhong Univ. Sci. Technol., Wuhan, China, 2008.
- [13] Q. Liang, “Research on key technology of motion capture in digital dynamic art,” Ph.D. dissertation, Beijing Jiaotong Univ., Beijing, China, 2016.
- [14] L. Venable and D. Ralley, “LabanWriter workshop (4.5.1),” Dept. Dance, Ohio State Univ., Columbus, OH, USA, Tech. Rep., Jul. 2004. [Online]. Available: <https://dance.osu.edu/sites/dance.osu.edu/les/resources-dnb-labanwriter-manual.pdf>
- [15] G. Misi. (Nov. 16, 2016). *Labanatory [EB/OL]*. [Online]. Available: <http://www.labanatory.com/>
- [16] A. Adamson, “Calaban,” Birmingham Univ., Birmingham, U.K., Tech. Rep., 1987.
- [17] F. Hunt, G. Politis, and D. Herbison-Evans, “Led & lintel: A windows mini-editor and interpreter for labanotation,” Basser Dept. Comput. Sci., Univ. Sydney, Sydney, NSW, Australia, Tech. Rep. 343, 2010.
- [18] K. Kojima, K. Hachimura, and M. Nakamura, “LabanEditor: Graphical editor for dance notation,” in *Proc. 11th IEEE Int. Workshop Robot Hum. Interact. Commun.*, Sep. 2002, pp. 59–64.
- [19] L. Wilke, T. Calvert, R. Ryman, and I. Fox, “From dance notation to human animation: The LabanDancer project,” *Comput. Animation Virtual Worlds*, vol. 16, nos. 3–4, pp. 201–211, 2005.
- [20] D. Maranan, M. Coyle, and T. Calvert, “A tool for translating dance notation to animation,” in *Proc. Western Comput. Graph. Symp.*, 2002, pp. 1–7.
- [21] J. Wang, Z. Miao, H. Guo, Z. Zhou, and H. Wu, “Using automatic generation of labanotation to protect folk dance,” *J. Electron. Imag.*, vol. 26, no. 1, Feb. 2017, Art. no. 011028.
- [22] M. Meredith and S. Maddock, “Motion capture file formats explained,” Dept. Comput. Sci., Univ. Sheffield, Sheffield, U.K., Tech. Rep., 2001.
- [23] H. Guo, “Research on automatic generation of labanotation based on human motion capture data,” Ph.D. dissertation, Beijing Jiaotong Univ., Beijing, China, 2015.
- [24] Taijiqian, [EB/OL]. (Mar. 22, 2020). [Online]. Available: https://en.wikipedia.org/wiki/Tai_chi
- [25] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, “Extreme learning machine: Theory and applications,” *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, Dec. 2006.

- [26] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst., Man, Cybern. B. Cybern.*, vol. 42, no. 2, pp. 513–529, Apr. 2012.
- [27] Z. Huang, Y. Yu, J. Gu, and H. Liu, "An efficient method for traffic sign recognition based on extreme learning machine," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 920–933, Apr. 2017.
- [28] OptiTrack. *OptiTrack Motion Capture Systems for Applications Ranging From Animation, Virtual Reality, Movement Sciences, Robotics and More*. Accessed: Aug. 20, 2019. [Online]. Available: <https://www.optitrack.com/>
- [29] X. Zhang, Z. Miao, Q. Zhang, and J. Wang, "Skeleton-based automatic generation of labanotation with neural networks," *Proc. SPIE*, vol. 28, no. 2, pp. 023026.1–023026.11, 2019.
- [30] X. Zhang, "Automatic generation of labanotation based on deep learning," M.S. thesis, Beijing Jiaotong Univ., Beijing, China, 2019.
- [31] Z. Li, "Design and Implementation of Genealogy Platform Based on Automatic Segmentation," M.S. thesis, Beijing Jiaotong Univ., Beijing, China, 2018.
- [32] M. Li, Z. Miao, and C. Ma, "Dance movement learning for labanotation generation based on motion-captured data," *IEEE Access*, vol. 7, pp. 161561–161572, 2019.
- [33] C. Li, Y. Hou, P. Wang, and W. Li, "Joint distance maps based action recognition with convolutional neural networks," *IEEE Signal Process. Lett.*, vol. 24, no. 5, pp. 624–628, May 2017.
- [34] P. Wang, W. Li, C. Li, and Y. Hou, "Action recognition based on joint trajectory maps with convolutional neural networks," *Knowl.-Based Syst.*, vol. 158, pp. 43–53, Oct. 2018.
- [35] C. Li, P. Wang, S. Wang, Y. Hou, and W. Li, "Skeleton-based action recognition using LSTM and CNN," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2017, pp. 585–590.
- [36] Y. Ning, C. Xiong, W. Xiong, H. Li, and G. Zhao, "Motion-recognition system with dual-processor using threshold-based algorithm," in *Proc. Int. Conf. Biol. Inf. Biomed. Eng.* Frankfurt, Germany: VDE, Jun. 2018, pp. 1–6.



ZHENJIANG MIAO (Member, IEEE) received the B.E. degree from Tsinghua University, Beijing, China, in 1987, and the M.E. and Ph.D. degrees from Northern Jiaotong University, Beijing, in 1990 and 1994, respectively.

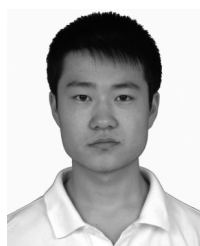
From 1995 to 1998, he was a Postdoctoral Fellow with the Ecole Nationale Supérieure d'Electrotechnique, d'Electronique, d'Informatique, d'Hydraulique et des Telecommunications, Institut National Polytechnique de Toulouse, Toulouse, France. From 1998 to 2004, he was with the Institute of Information Technology, National Research Council Canada, Nortel Networks, Ottawa, ON, Canada. He joined Beijing Jiaotong University, Beijing, in 2004. He is currently a Professor and the Director of the Media Computing Center, Beijing Jiaotong University, and also the Director of the Center for Ethnic and Folk Literature and Art Development, Institute for Digital Culture Research, Ministry of Culture, Beijing. His current research interests include image and video processing, multimedia processing, and intelligent human-machine interaction.



NINGWEI XIE received the B.E. degree from Beijing Jiaotong University, in 2018. She is currently pursuing the M.E. degree. She is the author of three conference papers. Her current research interests include image processing, human action recognition, multi-view computer vision, and analyzing 3D human motion data based on motion capture device.



WANRU XU received the B.S. degree in biomedical engineering and the Ph.D. degree in signal and information processing from Beijing Jiaotong University, Beijing, China, in 2011 and 2018, respectively. She is currently a Postdoctoral Researcher with the School of Computer and Information Technology, Institute of Information Science, Beijing Jiaotong University. Her current research interests include computer vision, machine learning, and pattern recognition.



JIAJI WANG received the B.E. degree from Beijing Jiaotong University, in 2012. He is currently pursuing the Ph.D. degree. He is the author of four conference papers and two journal articles. His current research interests include pattern recognition, image processing, multi-view computer vision, and analyzing 3D human motion data based on motion capture device.



ANG LI received the bachelor's degree from the Harbin Institute of Technology, in 2011. He is currently pursuing the doctor's degree with Beijing Jiaotong University. His research interests include compressed sensing, video processing, abnormal event detection, sparse reconstruction, and low-rank matrix reconstruction.

...