Received August 20, 2019, accepted October 31, 2019, date of publication November 5, 2019, date of current version November 15, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2951588

Dance Movement Learning for Labanotation Generation Based on Motion-Captured Data

MIN LI, ZHENJIANG MIAO, (Member, IEEE), AND CONG MA

Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China

Corresponding author: Min Li (16112066@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 2012BAH01F03.

ABSTRACT Labanotation is a widely used notation system for recording body movements, especially dances. It has wide applications in choreography preservation, dance archiving, and so on. However, the manual creation of Labanotation scores is rather difficult and time-consuming. Therefore, research on the generation of Labanotation scores is of great interest. In this paper, we aim to generate Labanotation scores based on the motion-captured data obtained from real-world dance performances. First, to deal with challenges such as various dance movement patterns, different dancer shapes, and noises in the motion-captured data, we propose a novel feature that is invariant to anthropometric variation and body orientation. Then, we generate the notations of both lower-limb movements and upper-limb gestures. On the one hand, we utilize the hidden Markov model (HMM) to analyze the temporal dynamic characteristics of limb movements and map each lower limb movement to a corresponding dance notation. On the other hand, for upper limbs, we train a multi-class classifier based on the extremely randomized trees (Extra-Trees) to identify the notations for arm gestures. Finally, we generate the Labanotation symbols based on the above movement analysis and thus create Labanotation scores. The proposed methods can generate the spatial symbols describing directions and levels in both the support column and arm column based on motion-captured data. The generated scores are clear and reliable. Experimental results show an average recognizing accuracy of over 92% for the generated notations, which is significantly better than previous work.

INDEX TERMS Labanotation, motion-captured data, hidden Markov models, extremely randomized trees, dance movement analysis.

I. INTRODUCTION

Folk dances and other performing arts are a valuable part of cultural heritage. The recording and archiving of traditional folk dances have attracted considerable research interests in the recent decade [1]–[7]. Due to the difficulty of art reproduction, preserving these precious cultural properties for the future generation is rather challenging. Specifically, in order to preserve and archive the performance of traditional dances, we need to record, process and reproduce human body motion by both quantitative and qualitative approaches.

Similar to music composers that record their compositions with music scores, the choreographers use dance notations to express their minds and talents, since professional scores are more accurate, convenient and easy to maintain and spread. Labanotation is one of the most widely used dance movement notation systems. It is highly accepted that the Labanotation [8], [9] is a powerful tool for choreography, dance training, and the recording and analysis of the human body motion [10]–[14]. Dance practitioners also utilize Labanotation to record traditional folk dances. Compared with other recording forms such as videos and images, Labanotation is saved in paper or digital form which needs little storage space, and it is better resistant to damages and easier to repair. Moreover, compared with written words, Labanotation is a vivid and scientific analysis and record system for human body motion. Dancers are able to read it conveniently.

However, manually creating the scores is difficult. It is quite time-consuming to carefully observe the dance movements and draw all the notations down, and it requires a

The associate editor coordinating the review of this manuscript and approving it for publication was Jenny Mahoney.

good grasp of specific domain knowledge to identify the features of body movements and convert them to Labanotation scores. In order to alleviate the work of drawing notations, some computer-aided tools have been created to simplify the manual process. For example, Laban Writer [15], LED & LINTEL [16] and LabanEditor [17] can be used to draw Labanotation scores and save them in a digital form. Nevertheless, it is still difficult to find expert users for these tools. Professional practitioners that are able to record dance with Labanotation are very rare due to the practical difficulties in dance recording. This is similar to the fact that many people can easily understand and perform a music score but cannot precisely record music with notations. In addition, it can take a lot of time for a skilled person to write down and proofread the correct notations for a piece of choreographic presentation, even if he has plenty of experience in composing Labanotation. Consequently, it is necessary to research on automatically generating Labanotation based on the digital data captured from real-world dances.

A natural solution for automatically generating Labanotation from dances is to analyze human motions based on the dance movement data obtained by the motion-capturing technique and thus work out the correct Labanotation scores. Therefore, the data analyzing method is vital for generating the correct Labanotation for a dance. Challenges in this procedure include noises in motion-captured data, different viewpoints of motion-capturing devices, the variety of dance performers with different anthropometric measurements (sizes of dancer bodies), and so on.

In this paper, to address the previously mentioned problems, we first propose a novel feature for a robust representation of motion-captured data for dance movement recognition. This feature not only handles the data variety in heights and shapes of different persons but also eliminates the body rotation effect and normalizes the various orientations in data caused by different viewpoints of devices. Then, based on the proposed feature, we utilize statistical learning methods to find corresponding notations for each motion segment. For lower limb movements, we take advantage of the hidden Markov model to categorize different kinds of movements and assign them to corresponding Labanotation symbols. For upper limb gestures, we propose to use the extremely randomized trees [18] algorithm to identify the corresponding notations. As shown in experiments, the proposed methods have better tolerance for the variances and noises in motion-captured data compared with previous methods. Contributions of this paper are summarized as follows:

- We propose an anthropometric and orientation invariant dance movement feature based on the *orientation-normalized relative joint position* (ON-RJP) vector extracted from motion-captured data to represent the movement properties.
- To generate the Labanotation symbols of support movements, we learn the dynamic characteristics of different dance movements on the lower limbs with the hidden Markov models. The observation probabilities in HMMs are estimated with Gaussian mixture models.
- To generate the spatial direction symbols in the arm column of Labanotation scores, we identify the upper limb gestures based on the extremely randomized trees algorithm. Compared with the spatial analysis based methods, our approach offers more flexibility and shows higher accuracies.
- Based on the above feature and models, we implement the dance movement learning and Labanotation generation for both lower limb movements and upper limb gestures. Experiments show favorable results against existing methods for generating corresponding notations.

This paper is an expanded version of our previous work [19]. On the one hand, to generate a more complete Labanotation score, we add the recognition for arm gestures based on upper limb analysis using an ensemble learning algorithm and obtain high accuracies in the generation of corresponding notations. On the other hand, for lower limbs, we construct a more complete dataset containing all the 48 symbols for directions and levels in the support columns and thus achieve a more fine-grained analysis for dance support movements. Therefore, this work makes a further step forward for the Labanotation score generation and contributes to applications such as dance step analysis and dance archiving.

The rest of this paper is organized as follows. First, in order to explain the Labanotation generation task clearly, we introduce the relevant background knowledge of Labanotation in Section II. The related work is reviewed in Section III. Then, in Section IV, we describe the details of feature extraction, dance movement modeling, and Labanotation symbol recognition. In Section V, we evaluate the proposed method on two large motion capture datasets and discuss the results. Finally, the paper concludes in Section VI.

II. BACKGROUND

In order to clarify the goal of our task, we first give a brief introduction of Labanotation scores in this section. Labanotation [8] is one of the most popular standardized symbol systems for recording and analyzing human body movements in dances such as ballet. It was named after its inventor Rudolf Von Laban, who is an Austro-Hungarian dancer and choreographer. Initially designed for dance recording, Labanotation is applied to many areas such as choreography preservation, dance education, movement analysis and dance archiving. From the notations, we can recognize and reproduce the recorded dance movements beyond the boundaries of times, languages and cultures. Therefore, we aim to use Labanotation for the preservation and reproduction of traditional folk dances. However, the writing of Labanotation scores is time-consuming and requires many manual efforts. We hope to develop algorithms to generate the major Labanotation symbols from motion-captured data, thus alleviating the workload of dance recorders.

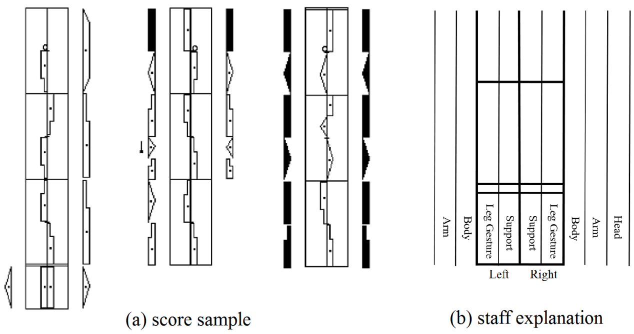


FIGURE 1. An example of the target Labanotation score and the illustration of score stuffs.

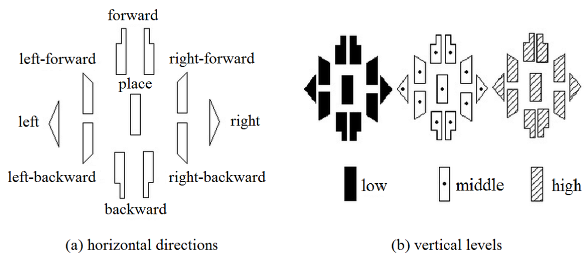


FIGURE 2. Spatial Labanotation symbols for directions and levels.

Here are some figures from [8] for an illustration of what we are aiming to generate. Figure 1 (a) shows an example of a Labanotation score, which consists of two parts: vertical staff and geometrical symbols. As Figure 1 (b) shows, in the score, there are nine basic columns and each column corresponds to a body part. In the middle of the staffs, there are two support columns specially designed to record the transference of weight, which is very important in dance theory. The underlying time axis is vertical and the score is to be read from the bottom to the top of the page.

Figure 2 (a) and (b) shows the fundamental spatial symbols of Labanotation. They describe the horizontal and vertical positions of dance movements in the space and are quite important for spatial Labanotation analysis. There are 9 horizontal direction symbols and 3 vertical level symbols and they can be combined to form a specific notation. These descriptions for horizontal orientations and heights divide the movement direction into 27 sub-spaces, leading to 27 notations. These notations are drawn for each body part to build up an entire record. Excluding the 3 *place* symbols that do not describe a movement, there are 24 spatial Labanotation notations that need to be drawn for each body part. In the Labanotation generation, we need to recognize movements from the data, correspond them to these correct notation symbols, and write them in the corresponding staff columns.

Based on whether the body weight is supported by the involved body part, the dance movements can be divided into two categories: support movements and non-support movements. Figure 3 explains the process of weight shifting during dance performing, where the symbols of right-right-middle and right-right-low represent the different weight transferring

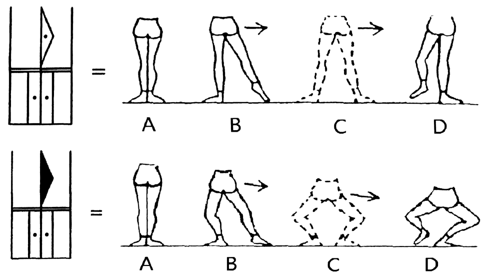


FIGURE 3. Examples of the body weight transference in a step and the corresponding spatial symbols.

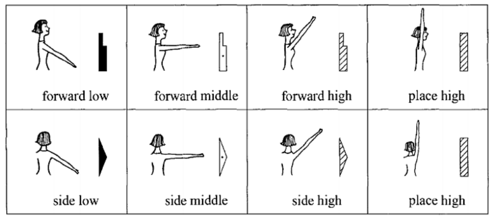


FIGURE 4. Illustration of several symbols for arm gestures in different directions and levels.

procedures from A to D. In the analysis and reproduction of dance, the transference of body weight plays an important role. This procedure is described by the symbols written in the two support columns of the Labanotation scores that focus on the weight-supporting part of the body. Therefore, a general weight transferring procedure completed by lower limbs can be described by the 48 spatial symbols representing 8 directions and 3 levels in the 2 columns. We aim to generate these notations for dance movements based on the motion-captured data. When both lower limbs are supporting the body weight, it usually can be described with the *place* symbols as a static pose. Furthermore, special support movements or postures such as kneeling, sitting, lying, and jumping are specifically described by more advanced symbols in the Labanotation system. These are beyond the scope of this work.

In addition, among non-support movements, the gestures and moving characteristics of arms contain important information on dance styles. The spatial symbols written in the arm columns describe the directions and levels of the upper limbs. Styles of the symbols are the same as previously stated, while the position in staff columns decides the body part that the symbols are describing. Some examples are shown in Figure 4 (Figures 3 and 4 are both from [8]). According to the Labanotation convention, the arm gestures are described in terms of the movement towards a specific destination. The symbols written in the arm columns are determined by the finishing point of a motion segment. Therefore, in order to generate the correct symbols, we will focus on features in the static ends of motion segments, which is different from analyzing lower-limb movements based on moving processes.

In summary, based on the Labanotation theory, for a piece of motion-captured data containing a dance movement, we aim to generate the notations in both support columns and arm columns. Therefore, we propose new features to deal with data variety and utilize statistical learning methods to analyze the dynamic process of body movements. We will present a detailed description of our method in Section IV.

III. RELATED WORK

A. MOTION-CAPTURED DATA

In order to digitize and analyze human dance movements, researchers have tried to acquire accurate human motion data from real-world dances. Motion-capturing is a technique first proposed by Gunnar Johansson [20], which has gotten rapid and maturely development in recent decades. This technique uses a specially-designed motion data format to record the spatial positions of major joints and has been applied to recording human body motion in the fields of CG-animation and movie production [21]. Since motion-capturing can provide precise 3D position data of human body motions across time, it is widely used to record, archive and evaluate performing arts, especially folk dances. For example, Hachimura [1], [2] used the motion-capturing technique in the digital archiving of dances. Stavrakis [3] recorded the performances of expert dancers with high-quality motion-captured data. In addition, some existing projects aiming at archiving and protecting intangible cultural heritages [4], [6], [22] also use Kinect and other multi-modal sensors to acquire the motion data of human dances.

One of the popular applications of motion-captured data is guiding humanoid robots to perform human-like motions. This requires the feature extraction and motion pattern recognition based on motion-captured data. Since the data obtained via different motion-capturing systems are in various formats, some researchers [23], [24] propose unified representations for motion visualization, recognition, and reproduction on actual robots. Inamura *et al.* [25] used HMMs to integrate the behavior cognition and generation. Takano and Nakamura [26], [27] also proposed to learn and symbolize motion patterns using HMMs. In addition, recognizing human actions such as run, jump, drink, and sit down from motion-captured data is also an important research topic. Recent methods proposed for action recognition on motion-captured data include approaches based on multi-class SVMs [28], histogram comparison [29], and so on. However, the challenges in generating Labanotation scores based on motion-captured data are different from these tasks. We need to analyze the positions, gestures and moving trajectories of each specific body part based on Laban theories. In this paper, we propose a more compact and task-specific feature, use a three-state HMM to model the movements of different body parts, and assign notations to the movement segments. Method details are described in Section IV.

B. LABANOTATION

Besides motion-captured data, another method for archiving and preserving dances is the dance notation system. Labanotation [8] is one of the most widely used dance notations. Except for manually drawing and reviewing the Labanotation scores on paper, computer-aided tools including Laban Writer [15], LabanEditor [17], LED & LINTEL [16], and LabanDancer [30] have been developed to assist drawing, visualizing and analyzing Labanotation scores. However, these tools still require considerable manual efforts and are difficult for non-experts of Labanotation. In addition, there are also some methods proposed to capture the semantics of Labanotation and represent them with open text formats, such as LabanXML [31], MovementXML [32], and DanceOWL [33], [34]. These formats are more human-readable, but there is still a big gap between real-world dance movements and Labanotation representations.

Labanotation generation aims to analyze motion-captured data of human dance movements and generate Labanotation scores accordingly. Since the related research is interdisciplinary, it is still an initiative and the solutions are limited to a few kinds of methods. The generating procedure usually consists of two steps: motion data acquisition and dance movement analysis. Different from the human action recognition task in the pattern recognition field, the analysis and identification for dance movements focus on the positions, gestures and moving trajectories of specific body parts, rather than recognizing semantical actions based on motions of the whole body. In addition, the motion-captured data of dances are usually in specialized formats that require additional processing methods.

In the literature, there are several methods developed to generate Labanotation scores from motion-captured data. Hachimura and Nakamura [35] proposed to divide the 3D coordinate space and map the subspaces to Labanotation symbols for upper limb movements. Later on, Chen *et al.* [36] exploited more features including turns and steps. They also considered the dynamic analysis from multiple frames. Guo et al. [37] studied the automatic generation for the reserved notations. Furthermore, Choensawat *et al.* [38], [39] took the weight support and bending of body parts into account and analyzed the jump movements. These methods are based on spatial analysis of body motions. They deal with raw data and use fixed rules that are specially designed to map different dance movements to Labanotation. However, when the real-world motion-captured data contain noises or deviate from the expected values, the fixed-rule-based methods may fail to generate the correct notations. In addition, Sankhla *et al.* [40] proposed a method based on SVMs to recognize postures from Kinect recording data and generate Labanotation scores and XML representations. This is a model-based method that paves the way for more general applications. However, it is limited to recognizing static postures rather than movements.

Recently, Zhou *et al.* [41], Zhou [42] proposed a template matching method to generate Labanotation scores for dance movements. They matched the test dance movement segments with templates in a manually constructed reference motion dataset. In this way, they obtained the best-matched notation for the dance movement and thus generated the

Labanotation score. Their sequential key-frame matching is based on the Euclidean distance and dynamic time warping. This method avoids defining fixed rules for symbols and shows better flexibility compared with the rule-based methods. Furthermore, Wang and Miao [43] proposed to combine rhyme alignment, spatial analysis on upper limbs, and DTW matching for lower limbs to form a more comprehensive solution for Labanotation generation. Their work shows favorable accuracy in the generation of standard Labanotation symbols.

However, the real-world data may deviate from handcrafted standard templates when there are many dancers with various heights and shapes. In order to achieve better generalization performances, we propose to build a hidden Markov model (HMM) for each category of movement segments that correspond to a certain dance notation on the lower limbs and then the corresponding notations of dance movement segments can be generated based on searching among all the trained hidden Markov models. In addition, we put forward to train a classifier based on the extremely randomized trees to identify the notations for arm gestures. In our experiments, we can obtain better performances than the method described in [42] and [43]. More detailed descriptions are in the following sections.

IV. DANCE MOVEMENT LEARNING

Given the segments of motion-captured data, in order to correctly recognize the movements of individual body parts and obtain corresponding notations, we extract representative features from the raw motion data. Based on the proposed feature vectors, we determine the corresponding Labanotation symbol for a movement segment with statistical learn methods. For the weight-supporting movements of lower limbs, we learn temporal dynamic characteristics of movement segments with the hidden Markov models. For upper limbs, we utilize the extremely randomized trees algorithm to identify arm gestures from the final states of arm movements. In this way, we recognize the unknown dance movement in a data segment and find the corresponding symbol for it. Finally, a Labanotation score is generated. Details of our method are described below.

A. MOTION DATA PROCESSING

First of all, we use a labeled optical motion-capturing system provided by OptiTrack^{TM} [44] to obtain the dance motion data. This system can precisely capture the body movements of the dance performer by recording the locations of 41 reference points on the body. The obtained data are fitted to a 26-joint skeleton model as shown in Figure 5. The captured data is saved in the BVH (Bio-vision hierarchical) format for further analysis with computer algorithms.

The BVH format is designed and widely used for recording captured human body motions. A BVH file is a plain text file consisting of two parts, one part beginning with keyword *hierarchy* defines the structure of human skeletal joints, the other part beginning with keyword *motion* records the motion data, in which each row is a frame data that describes

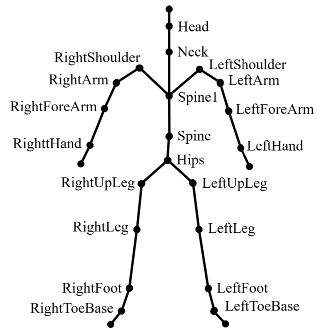


FIGURE 5. The skeleton model of 26 joints.

the pose of a human body at a certain moment. BVH file uses the Euler angle to represent the relative rotation data of a skeletal joint to its parent joint. In this way, the orientation information of joints in each movement is recorded. However, when generating Labanotation, it is difficult to directly use the raw data in such a form to analyze the motions of body parts. Therefore, appropriate data processing and feature extraction need to be carried out.

In order to get a clear and well-organized representation of the raw captured data, we first convert the BVH-formatted Euler angle data to a sequence of the 3D Cartesian world coordinates. For each joint J_p, there is a set of Euler angles in the captured data which are assumed to be r, p and y obtained by rotating according to the rule of "roll-pitch-yaw". Assuming that the initial position offset of the joint J_p's child joint J_c is (x_0, y_0, z_0). After the rotation of joint J_p, a rotation matrix $\mathbf{M}$ can be computed based on the recorded Euler angles as shown in Equation (1), where $\mathbf{R}$, $\mathbf{P}$ and $\mathbf{Y}$ are rotation matrices of joint J_p around the axes for roll, pitch, and yaw respectively. They can be computed by the Equation (2), (3) and (4). Therefore, a new offset (x_c, y_c, z_c) of the joint J_c after rotating with the matrix $\mathbf{M}$ is obtained by Equation (5) (1)–(5), as shown at the bottom of the next page.

Due to the hierarchical dependency of human joints, the joint J_c is not only affected by the rotation of its immediate precursory joint J_p, but also by all of its precursory joints. Denoting the rotation matrices of all the precursory joints from near to far till the root joint with $M_1, M_2, \cdots, M_r$, the genuine offset T_c of the joint J_c relative to its all precursory joints can be calculated by:

$$T_c = \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \mathbf{M_r}\left(\mathbf{M_{r-1}}\left(\cdots\left(\mathbf{M_2M_1}\begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix}\right)\right)\right). \quad (6)$$

Thus, for all the precursory joints $J_1, J_2, \cdots, J_r$ (J_r denotes the root joint without parent joint) of the joint J_c, we can obtain their translating offsets denoted by $T_1, T_2, \cdots, T_{r-1}$. Therefore, the position coordinate data P are computed by:

$$P = P_{root} + T_{r-1} + \cdots + T_2 + T_1 + T_0, \quad (7)$$

where P_{root} indicates the root joint position in the world coordinate system, which is given in the raw BVH data.

B. FEATURE EXTRACTION

A series of dance movements can be represented by a sequence of 3D position coordinates of body joints. However, challenges including viewpoint changes, different body orientations, and various dancer shapes make it difficult to recognize the body movements simply based on the 3D coordinate data sequences.

On the one hand, The Labanotation is an orientation-encoded movement description system with an implicit default orientation of the whole human body. But due to viewpoint changes and different relative locations of the dancer body and the recording device, the same movement in various body orientations may be recorded with fairly different spatial coordinate data. This can seriously affect recognition results. Therefore, we propose an orientation-normalized feature (ON) to eliminate the disorder caused by body rotation and achieve a better generalization performance.

In order to normalize the human body orientations, we organize the motion data in the world coordinate system in a body-forward oriented way and convert the data to another right-handed Cartesian coordinate system. In this representation, we use the forward orientation (the face direction) of the body as the Z-axis. In each frame of data, in order to find out the direction that the dancer is facing, we first compute the forward vector that is perpendicular to the coronal plane of the body. In the human skeleton model, this plane can be represented with three key joint points *LeftUpLeg*, *RightUpLeg*, and *Spine1* among the 26 joints, as the triangle shown in Figure 6. The forward vector is denoted as $\vec{v}$.

The computing procedure of the forward vector is as follows. In the world coordinate system, the body coronal plane P is determined with three joint points (x_l, y_l, z_l), (x_r, y_r, z_r) and (x_s, y_s, z_s). We can compute the normal vector $\vec{v}$ of plane P with a simple geometry method. Defining two intersecting vectors $\overrightarrow{SR}$ and $\overrightarrow{SL}$ based on the three joint points, the normal vector can be calculated by the cross-product of intersection vectors, as follows:

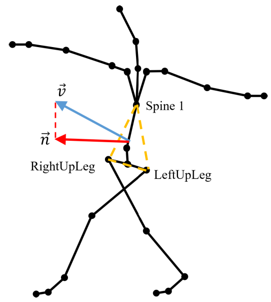


FIGURE 6. Representation of the human body orientation vector in the 26-joint skeleton model.

$$\begin{aligned}\vec{n} &= \overrightarrow{SR} \times \overrightarrow{SL} \\ &= ((x_s, y_s, z_s) - (x_r, y_r, z_r)) \times ((x_s, y_s, z_s) - (x_l, y_l, z_l)).\end{aligned} \tag{8}$$

As such, we get the vector that indicates the forward orientation of the human body. However, considering that $\vec{v}$ has an elevation angle, we take the horizontal component $\vec{n}$ of $\vec{v}$ to represent the straight-ahead direction of the body. Assuming that vector $\vec{v}$ is (x_f, y_f, z_f) in the world coordinate system, the vector $\vec{n}$ is $(x_f, 0, z_f)$, as shown in Figure 6.

With the vector $\vec{n}$ computed, we can apply a 3D rotation operation around the Y-axis to convert the motion data from the world coordinates to the body forward-oriented coordinates. Denoting the joint J as (x, y, z) in the world coordinate system, and (x', y', z') in the body forward-oriented system, the rotation operation can be described as:

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \tag{9}$$

where θ is the rotation angle computed from $\vec{n}$ and the X-axis.

$$\mathbf{M} = \mathbf{RPY} = \begin{bmatrix} \cos r \cos y - \sin p \sin y \sin r & -\sin r \cos p & \cos r \sin y + \sin r \sin p \cos y \\ \sin r \cos y + \cos r \sin p \sin y & \cos r \cos p & \sin r \sin y - \cos r \sin p \cos y \\ -\cos p \sin y & \sin p & \cos p \cos y \end{bmatrix} \tag{1}$$

$$\mathbf{R} = R_z(-r) = \begin{bmatrix} \cos r & -\sin r & 0 \\ \sin r & \cos r & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

$$\mathbf{P} = R_x(-p) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos p & -\sin p \\ 0 & \sin p & \cos p \end{bmatrix} \tag{3}$$

$$\mathbf{Y} = R_y(-y) = \begin{bmatrix} \cos y & 0 & \sin y \\ 0 & 1 & 0 \\ -\sin y & 0 & \cos y \end{bmatrix} \tag{4}$$

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \mathbf{M} \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} \tag{5}$$

On the other hand, different dance performers are different in anthropometric measures such as height, shape and limb lengths. This also makes it difficult to identify the same dance movements from various motion data. By observing hundreds of dances, we find that the category of body movements tend to be more easily distinguished by the relative motions of limbs and joints, rather than the absolute positions. Therefore, the above problem can be alleviated by computing the relative position differences of joints. Consider an arbitrary joint J_c with coordinate (x_c, y_c, z_c) and the first joint J_0 as (x_0, y_0, z_0) in one frame, we can compute a relative position vector between them as follows:

$$\overrightarrow{J_0 J_c} = (x_c, y_c, z_c) - (x_0, y_0, z_0). \tag{10}$$

For the 26 joints in each frame, we can compute the difference vectors as a relative joint position (RJP) feature to obtain a more flexible representation of body posture and movements. When we focus on particular limbs or other body parts, we can also compute the difference vectors only among the related joints to obtain a more compact representation.

Furthermore, we can combine the two features and obtain a series of feature vectors for each frame of motion-captured data. The combined feature is named as Orientation-Normalized Relative Joint Positions (ON-RJP). As a better representation of the body movements based on the motion-captured data, this feature can be applied to multiple dance movement analysis methods including motion distance measuring and statistical modeling. In addition, this feature is flexible and suitable for the application to Labanotation generation, since it can be extracted on a subset of joints when we focus on the movements of a particular body part.

C. SUPPORT MOVEMENT MODEL BASED ON HMMS

Beyond extracting static feature from each frame of motion-captured data, we also propose to model the dynamic characteristics of dance movements for the robust recognition of key Labanotation symbols. In order to describe both the temporal transition of body movements and the spatial distributions of limbs and joints, we take advantage of the hidden Markov model (HMM) and build the temporal model based on a sequence of processed dance motion-captured data. As mentioned previously, in this work, we focus on the movement of lower limbs for dance step analysis. To find out the correct support-column Labanotation symbols for a piece of dance, each body movement needs to correspond to a specific notation. For each lower limb, we consider the 24 spatial notation symbols constructed by 8 horizontal directions and 3 vertical levels. Therefore, we need to build 48 fine-grained dance movement models in total, which will be described in this section.

Figure 7 is a visual illustration of the body movement that corresponds to the Labanotation symbol of *left-forward-middle* to be drawn in the left support column. When we model a segment of dance movement, we extract the feature vectors of 10 joints on the two legs. As described in the

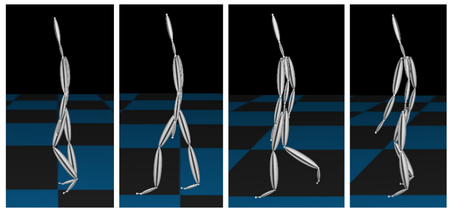


FIGURE 7. Visual illustration of the body movement that corresponds to the Labanotation support symbol of *left-forward-middle.*

subsection above, we compute the difference vectors for relevant joints based on the orientation-normalized coordinates in each frame as the data representation for model building.

The feature vector in a data frame describes the static state of the body pose at a particular moment. Thus the dance movement can be represented by a sequence of these features. Furthermore, in order to exploit the dynamic temporal information in the feature sequences, we use a hidden Markov model (HMM) with a three-state left-right structure. In this way, each category of human dance movements can be represented with a trained HMM. In frame t, we extract the ON-RJP feature from the motion-captured data and take it as the observation vector O_t. Therefore, a segment of dance movement can be denoted by the observation sequence $O = \{O_1, O_2, \ldots O_t\}$. We use a hidden Markov model denoted by $\lambda = (\pi, A, B)$, where A denotes the temporal state transition probability, B is the probability density of observation from hidden states, and π is the initial state distribution. Specifically, the observation probability density for each state is represented with the Gaussian mixture models (GMMs) as defined in the Equation (11), where o_t is a continuous observation vector, and M is the number of Gaussian mixture components. μ_m, $\sum_m$, and w_m denotes the mean vector, covariance matrix, and weight of each component m, respectively. In order to learn dynamic transition information from motion data, we first train an HMM model for each category of dance movement. Given observation sequences $O_1, O_2, \ldots O_n$ for N movement categories, we need to optimize the parameters of model λ to maximize all the $P(O|\lambda)$. According to the common solution for HMM training given by [45], this problem can be solved by the *Baum-Welch* algorithm [46].

$$b_i(O_t) = \sum_{m=1}^{M} \left(w_m \frac{1}{(2\pi)^{\frac{d}{2}} |\sum_m|^{\frac{1}{2}}} \times exp \left(-\frac{1}{2}(O_t - \mu_m) \sum_m^{-1} (O_t - \mu_m)^T \right) \right) \tag{11}$$

At the testing stage, we recognize a candidate dance movement sequence O by finding the best-fitted model i that has the maximum value of $P(O|\lambda_i)$, as shown in equation (12) and in Figure 8. The calculation of probability $P(O|\lambda)$ for sequence O with model λ is solved by the *Forwards-Backwards* algorithm [46], [47]. Finally, the best state sequence can be given by the *Viterbi* algorithm [48].

$$movement(O) = \underset{i=1,\cdots,N}{\arg\max}\, P(O|\lambda_i) \tag{12}$$

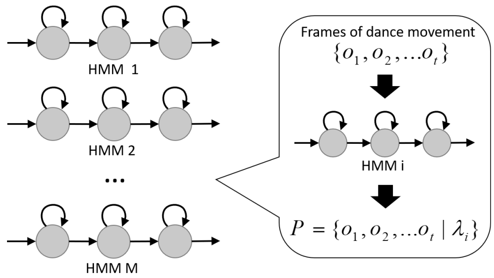


FIGURE 8. **Illustration of recognizing a body movement based on three-state left-right HMMs.**

As such, the category of a dance movement segment could be identified. Finally, we map the recognized labels to the corresponding Labanotation symbols in the support columns.

D. ARM GESTURE CLASSIFIER BASED ON TREES

In order to make a further step to a complete Labanotation score, we continue to work on the columns for arms. As explained in Section II, the choosing of Labanotation spatial symbols for arm gestures are decided by the final states of the movement. Different from lower-limb movements, when we generate the symbols in the arm columns, we concentrate on the static states at the end of upper limb movements, rather than the moving processes. Therefore, we take the last N frames of each movement segment and extract relative position features on arm joints as described above. In this case, HMMs are no longer the best choice for recognition.

In this task, given the upper-limb features extracted from each segment representing an individual dance movement, we aim to identify the correct Laban symbols for arm gestures in these segment samples. This can be treated as a multi-class classification problem. Each spatial symbol for the arm gesture can be seen as a class label since the symbols are mutually exclusive. Therefore, we propose to train a multi-class classifier based on an ensemble of decision trees. Specifically, we use the extremely randomized trees (Extra-Trees) algorithm [18], which has shown good performance in numerous classification scenarios.

Different from the Random Forest method [49] or other bagging approaches based on the tree classifiers, this algorithm builds the decision trees from the original dataset rather than randomly sampled subsets. The randomization characteristics lie in the way of choosing the best split. When building decision trees, the samples are split by randomly selected features at each node. Multiple decision trees are built from the training set and the label prediction are decided by combining the results of all the weak classifiers. Therefore, this method based on ensemble decision trees naturally fits for multi-class problems. In practice, we build 100 decision trees and use the Gini criterion to make splits. Finally, we predict a label for each movement segment and choose the corresponding symbols.

In summary, based on the above-described methods, the Labanotation score for a whole piece of dance can be generated. For the result presentation, we draw the score stuff and symbols with the computer graphics technique.

V. EXPERIMENTS

A. DATASETS

Due to the fact that the research on the generation of Labanotation scores is rare and the motion-capturing devices are expensive, few public-available datasets are labeled with Labanotation symbols. In order to analyze the support movements of lower limbs, we extend the work in [19] and construct a more comprehensive motion-captured dataset covering all the direction and level Labanotation symbols in the support columns. The data is recorded with the OptiTrack^{TM} motion-capturing system [44].

This dataset consists of 19, 200 manually segmented and labeled body movement segment data of 4 different persons with heights ranging from 160 cm to 190 cm. In the recorded and segmented data, all the 8 horizontal directions of the lower limb movements are labeled, which are: *left-forward, forward, right-forward, left, right, left-backward, right-backward,* and *backward*. Furthermore, beyond these 8 directions, we also label the 3 vertical movement levels including *high, middle,* and *low*. Therefore, we have all the 24 support movement directions labeled. These labels correspond to the 24 spatial symbols for directions and levels in the Labanotation support columns, as introduced in Section II. Since there are 2 support columns for the left and right leg respectively, our labels are of 48 categories in total.

Details of the dataset settings are described as follows. The original data is recorded based on the continuous movements of volunteers. For the ground truth label of each movement, the captured motion data sequences are manually segmented so that each segment includes one movement. Each movement sample is labeled with one of the 48 categories. In the dataset, support movement segments of different labels are equally distributed in each category. Meanwhile, the data are also equally captured from the performances of each person. We take 100 segment samples for each of the 48 movement categories of one person. Thus the total sample number is $100 * 48 * 4 = 19,200$. In addition, the segment data are uniformly distributed in different body orientations for better diversity. The sequence data of each segment are sampled at a fixed sampling speed of 150 frames per second and the lengths of these segments range from 90 to 200 frames due to different time durations. Challenges of the utilized dataset include different dance performers, changing viewpoints and other variances brought by a big data volume. The dataset will be made available to the public.

In addition, for the analysis of upper limbs, a big dataset containing arm movements is available from the work in [43]. This dataset includes 5310 movements of the left arm and 5220 movements of the right arm, which are already segmented to single movement samples and have the corresponding labeled symbols. In order to make a fair comparison with the method in [43], we use this dataset for the evaluation of Labanotation generation for the arm gestures.

TABLE 1. Comparisons between the approaches in references [28], [29], [42] and the proposed method on recognizing accuracy (%).

Methods	Left Leg	Right Leg	Average
DTW [42]	73.69	73.35	73.52
Multi-class SVMs [28]	86.16	85.52	85.84
Histogram-based [29]	89.36	90.28	89.82
Our method	**92.72**	**92.58**	**92.65**

TABLE 2. Ablation study on the two components of the extracted features.

Feature settings	Plain	ON	RJP	ON-RJP
Orientation normalization	-	✓	-	✓
Relative joint position	-	-	✓	✓
Average accuracy(%)	34.45	58.83	73.30	**92.65**

B. EXPERIMENTAL RESULTS

In order to generate the corresponding notation correctly, we need to learn from data and identify the categories of segment samples. Therefore, this problem can be seen as a specialized recognizing task. We evaluate the performance of our method with the average recognition accuracy. For the training-test split, we randomly select half of the samples in each movement category for training and the rest for testing. Since there is randomness in the algorithms, we repeat the experiment five times and averaged the results to obtain the final result.

1) SUPPORT MOVEMENTS

First, we evaluate the performance of HMMs for modeling dance movements and generating the spatial symbols in support columns. The results are compared with the key-frame matching method based on dynamic time warping (DTW) proposed by Zhou [42], the multi-class SVMs method [28] and histogram-based method [29].The latter two methods are originally designed for action recognition in motion-captured data and we adapt them to our case. As shown in Table 1, the accuracy of our method is significantly higher than those of the methods in [42], [28] and [29]. This indicates that the temporal dynamic modeling based on HMMs is suitable for motion-captured data analysis and can achieve favorable results.

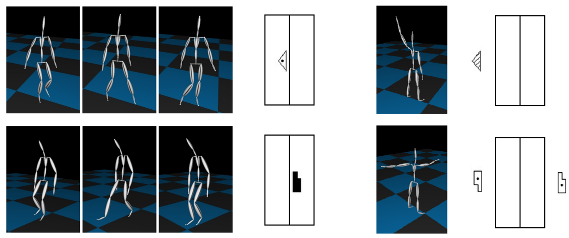
In addition, since the proposed feature consists of two parts, we conduct experiments on support movement recognition with four different feature configurations for an ablation study. The average accuracies are compared in Table 2. We can see that both the two components of the proposed feature contributes to the final performance.

2) ARM GESTURES

In order to evaluate the notation generation quality for arm gestures based on the upper limb analysis, we conduct experiments on the dataset from [43] and compare the proposed method with their spatial-analysis-based approach. The results are reported separately for the left arm and the right arm in that paper. Therefore, we evaluate our method in the same way and also compute the average results. The recognizing accuracies (in percentage) are shown in Table 3. From the table, we can see that the proposed method is over 10% higher than the previous work in terms of recognizing accuracy. This shows the advantage of a statistical learning method over the approach based on fixed spatial rules.

TABLE 3. Comparisons between the approach in reference [43] and the proposed method on recognizing accuracy (%) for arm gesture analysis.

Methods	Left Arm	Right Arm	Average
Proposed in [43]	83.61	82.73	83.17
Our method	**95.70**	**94.16**	**94.93**



(a) lower limb movements (b) arm gestures

FIGURE 9. Some examples of movements/gestures and the corresponding generated Labanotation symbols.

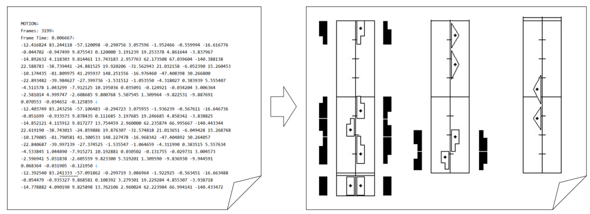


FIGURE 10. Illustration of the Labanotation generation for a piece of dance movements based on the corresponding motion-captured data.

3) PARAMETER DISCUSSION

In the analysis of lower-limb support movements, the motion data in the dataset are recorded at a speed of 150 frames per second (fps). There might be some data redundancy since this precision is more than enough to recognize the slightest human movements. Therefore, we down-sampled all the segment data to reduce the computation cost and speed up the model training procedure. In order to study the influence of sampling parameters, we conducted experiments and perform uniform down-sampling at different frame rates. The recognizing accuracies and training times (both averaged over the results of two legs) of different down-sampling settings are shown in Table 4. The table shows that lower frame rates can bring considerable reduction for the training time, while the accuracies are not affected significantly. Therefore, we can choose a low frame rate and save computation time.

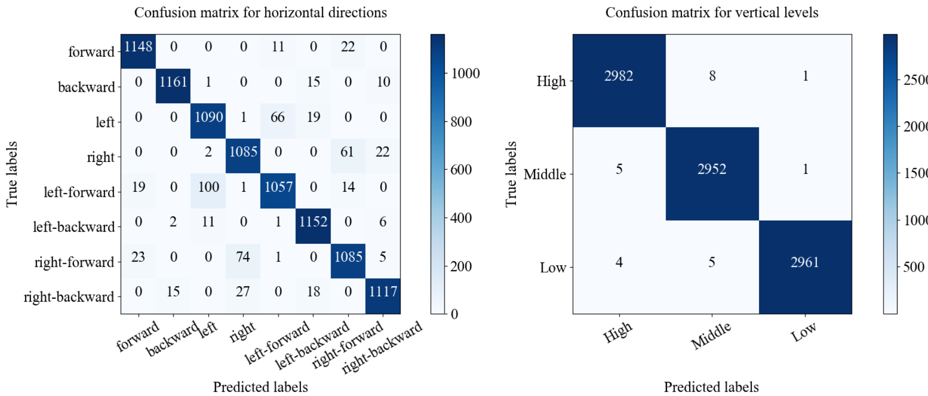


FIGURE 11. Confusion matrices for 8 horizontal directions and 3 vertical levels in support movement recognition.

TABLE 4. Performance comparisons of different down-sampling settings.

Down-sampling settings	Recognizing accuracy (%)	Training time (s)
Frame rate = 5 fps	92.58	**7750.20**
Frame rate = 10 fps	92.62	8401.54
Frame rate = 50 fps	**92.65**	9404.23
Frame rate = 150 fps (original)	92.61	9819.89

TABLE 5. Comparisons of different frame counts in arm gesture analysis on recognizing accuracy (%). The best results are in bold.

Frame counts	Left Arm	Right Arm	Average
1	**95.87**	93.34	94.61
5	95.70	**94.16**	**94.93**
10	95.54	93.66	94.60
20	94.72	93.87	94.30

In addition, the analysis of arm gestures focuses on the final states of a motion segment. In practice, we use the last N frames to train the classifier. Additional experiments are conducted to verify how many frames are enough to obtain satisfactory results. We experiment on $N \in \{1, 5, 10, 20\}$ and show the recognizing accuracies in Table 5. From the table, we can see that the best overall performance is achieved when $N = 5$. The accuracy decreases when the frame count is over 10. In contrast, using very few frames is not a better choice, either. Therefore, we use the last 5 frames in the experiments.

4) SCORE GENERATION

Finally, with the notations recognized for all the movement segments in a piece of dance, we can draw the final Labanotation scores in both the support columns and the arm columns using the corresponding symbols as explained in Figure 2. Some examples of the dance movements and corresponding symbols are shown in Figure 9. After the recognition, notations for all the motion segments are sequentially concatenated along the time axis to form the entire score of a big motion sequence. Therefore, the Labanotation score generation is finished. An example of the scores generated based on a sequence of motion-captured data is shown in Figure 10.

5) FAILURE CASES AND LIMITATIONS

To further analyze the proposed method, we present the confusion matrices of the recognition results in Figure 11. For the convenience of analysis, the data for 8 horizontal directions and 3 vertical levels are summarized separately. From the figure we can see that, there are some recognition failures for horizontal directions between *left* and *left-forward* as well as *right* and *right-forward*. This is not surprising since natural movements are not always performed in an exact direction. A slight deviation of movement may confuse the two adjacent directions, especially for the forward movements which are more casual. In contrast, it is easier to recognize vertical levels.

In addition, despite the high recognition accuracies, our work still has a few limitations. The above-described methods are based on the pre-segmented movement sequences while each segment corresponds to an individual limb movement. Currently, the pre-segmentation step still requires manual efforts or automatic segmentation algorithms. Besides, this work aims to generate the major symbols for supporting lower limbs and upper limbs in Labanotation, while the study on more complicated movements and highly advanced symbols remains for future work.

VI. CONCLUSION

In this paper, we aim to generate Labanotation scores based on motion-captured data. First, we propose a robust feature to obtain an effective representation for dance movements based on the processed motion data. Then, given the dance movement segments, we train an HMM for each category of lower limb movements corresponding to the Labanotation support symbols. Furthermore, we propose to use an extremely randomized trees algorithm to learn the arm gestures from upper limb data. As such, we identify the notations in both support columns and arm columns based on the motion data. Finally, the Labanotation score for a piece of dance can be generated. Experiments show that our method achieves favorable accuracies for symbol generation against the previous methods. This can reduce the workload for dance recorders in writing Labanotation scores.

In the future, we hope to record and obtain more dance motion-captured data and test the proposed method on larger motion sequences. When a large amount of labeled data are available, we can consider using methods with more parameters such as deep neural networks to further improve the generation quality. In addition, we are also exploring ways to generate Labanotation scores directly from continuous motion sequences without pre-segmentation. These remain for future work.

REFERENCES

[1] K. Hachimura, "Digital archiving of dancing," *Rev. Nat. Center Digit.*, vol. 8, no. 1, pp. 51–66, 2006.

[2] K. Hachimura, "New directions in digital humanities for Japanese arts and cultures," in *Digital Archiving of Dance by Using Motion-Capture Technology*. Kyoto, Japan: Nakanishiya Publishing, 2008, pp. 167–182.

[3] E. Stavrakis, A. Aristidou, M. Savva, S. L. Himona, and Y. Chrysanthou, "Digitization of cypriot folk dances," in *Proc. Euro-Medit. Conf.* Berlin, Germany: Springer, 2012, pp. 404–413.

[4] A. Kitsikidis, K. Dimitropoulos, S. Douka, and N. Grammalidis, "Dance analysis using multiple Kinect sensors," in *Proc. Int. Conf. Comput. Vis. Theory Appl. (VISAPP)*, vol. 2, Jan. 2014, pp. 789–795.

[5] A. Aristidou, E. Stavrakis, P. Charalambous, Y. Chrysanthou, and S. L. Himona, "Folk dance evaluation using laban movement analysis," *J. Comput. Cultural Heritage*, vol. 8, no. 4, p. 20, 2015.

[6] A. D. Doulamis, A. Voulodimos, N. D. Doulamis, S. Soile, and A. Lampropoulos, "Transforming intangible folkloric performing arts into tangible choreographic digital objects: The terpsichore approach," in *Proc. VISIGRAPP*, 2017, pp. 451–460.

[7] I. Rallis, N. Doulamis, A. Doulamis, A. Voulodimos, and V. Vescoukis, "Spatio-temporal summarization of dance choreographies," *Comput. Graph.*, vol. 73, pp. 88–101, Jun. 2018.

[8] A. H. Guest, *Labanotation: The System of Analyzing and Recording Movement*, 4th ed. Evanston, IL, USA: Routledge, 2014.

[9] R. Laban and L. Ullmann, *The Mastery of Movement*, 4th ed. Binsted, U.K.: Dance Books, 2011.

[10] D. Chi, M. Costa, L. Zhao, and N. Badler, "The EMOTE model for effort and shape," in *Proc. 27th Annu. Conf. Comput. Graph. Interact. Techn.* Reading, MA, USA: Addison-Wesley, 2000, pp. 173–182.

[11] L. Torresani, P. Hackney, and C. Bregler, "Learning motion style synthesis from perceptual observations," in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, pp. 1393–1400.

[12] A. Aristidou, P. Charalambous, and Y. Chrysanthou, "Emotion analysis and classification: Understanding the performers' emotions using the LMA entities," *Comput. Graph. Forum*, vol. 34, no. 6, pp. 262–276, 2015.

[13] F. Durupinar, M. Kapadia, S. Deutsch, M. Neff, and N. I. Badler, "Perform: Perceptual approach for adding ocean personality to human motion using laban movement analysis," *ACM Trans. Graph.*, vol. 36, no. 1, p. 6, 2017.

[14] S. Dewan, S. Agarwal, and N. Singh, "Spatio-temporal Laban features for dance style recognition," in *Proc. IEEE 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 2911–2916.

[15] L. Venable and D. Ralley, "LabanWriter workshop (4.5.1)," Dept. Dance, Ohio State Univ., Columbus, OH, USA, Jul. 2004. [Online]. Available: https://dance.osu.edu/sites/dance.osu.edu/files/resources-dnb-laban-writer-manual.pdf

[16] F. Hunt, G. Politis, and D. Herbison-Evans, "Led & lintel: A windows mini-editor and interprepter for labanotation," Basser Dept. Comput. Sci., Univ. Sidney, Sydney, NSW, Australia, Tech. Rep. 343, 2010.

[17] 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.

[18] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Mach. Learn.*, vol. 63, no. 1, pp. 3–42, Apr. 2006.

[19] M. Li, Z. Miao, and C. Ma, "Automatic Labanotation generation from motion-captured data based on hidden Markov models," in *Proc. Asian Conf. Pattern Recognit. (ACPR)*, Nov. 2017, pp. 764–769.

[20] G. Johansson, "Visual perception of biological motion and a model for its analysis," *Perception Psychophysi.*, vol. 14, no. 2, pp. 201–211, 1973.

[21] R. Maiocchi, "3D character animation using motion capture," in *Interactive Computer Animation*. Upper Saddle River, NJ, USA: Prentice-Hall, 1996, pp. 10–39.

[22] K. Dimitropoulos, S. Manitsaris, F. Tsalakanidou, S. Nikolopoulos, B. Denby, S. Al Kork, L. Crevier-Buchman, C. Pillot-Loiseau, M. Adda-Decker, S. Dupont, J. Tilmanne, M. Ott, M. Alivizatou, E. Yilmaz, L. Hadjileontiadis, V. Charisis, O. Deroo, A. Manitsaris, I. Kompatsiaris, and N. Grammalidis, "Capturing the intangible an introduction to the i-Treasures project," in *Proc. Int. Conf. Comput. Vis. Theory Appl. (VISAPP)*, vol. 2, Jan. 2014, pp. 773–781.

[23] P. Azad, T. Asfour, and R. Dillmann, "Toward an unified representation for imitation of human motion on humanoids," in *Proc. IEEE Int. Conf. Robot. Autom.*, Apr. 2007, pp. 2558–2563.

[24] C. Mandery, Ö. Terlemez, M. Do, N. Vahrenkamp, and T. Asfour, "Unifying representations and large-scale whole-body motion databases for studying human motion," *IEEE Trans. Robot.*, vol. 32, no. 4, pp. 796–809, Jul. 2016.

[25] T. Inamura, I. Toshima, H. Tanie, and Y. Nakamura, "Embodied symbol emergence based on mimesis theory," *Int. J. Robot. Res.*, vol. 23, nos. 4–5, pp. 363–377, 2004.

[26] W. Takano and Y. Nakamura, "Humanoid robot's autonomous acquisition of proto-symbols through motion segmentation," in *Proc. 6th IEEE-RAS Int. Conf. Humanoid Robots*, Dec. 2006, pp. 425–431.

[27] W. Takano and Y. Nakamura, "Symbolically structured database for human whole body motions based on association between motion symbols and motion words," *Robot. Auto. Syst.*, vol. 66, pp. 75–85, Apr. 2015.

[28] C. Li, P. R. Kulkarni, and B. Prabhakaran, "Segmentation and recognition of motion capture data stream by classification," *Multimedia Tools Appl.*, vol. 35, no. 1, pp. 55–70, 2007.

[29] M. Barnachon, S. Bouakaz, B. Boufama, and E. Guillou, "Ongoing human action recognition with motion capture," *Pattern Recognit.*, vol. 47, no. 1, pp. 238–247, 2014.

[30] 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.

[31] M. Nakamura and K. Hachimura, "An XML representation of labanotation, labanxml, and its implementation on the notation editor labaneditor2," *Rev. Nat. Center Digit. (Online J.)*, no. 9, pp. 47–51, 2006. [Online]. Available: http://www.ncd.matf.bg.ac.rs/issues.html

[32] J. Hatol, "MOVEMENTXML: A representation of semantics of human movement based on Labanotation," Ph.D. dissertation, School Interact. Arts Technol., Simon Fraser Univ., Burnaby, BC, USA, 2006.

[33] K. El Raheb and Y. Ioannidis, "A labanotation based ontology for representing dance movement," in *Proc. Int. Gesture Workshop*. Berlin, Germany: Springer, 2011, pp. 106–117.

[34] K. El Raheb and Y. Ioannidis, "Dance in the world of data and objects," in *Proc. Int. Conf. Inf. Technol. Perform. Arts, Media Access, Entertainment*. Berlin, Germany: Springer, 2013, pp. 192–204.

[35] 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 Human Interact. Commun.*, Sep. 2001, pp. 122–127.

[36] 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.

[37] 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.

[38] 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, 2015.

[39] 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.

[40] A. Sankhla, V. Kalangutkar, H. B. Bhuyan, T. Mallick, V. Nautiyal, P. P. Das, and A. K. Majumdar, "Automated translation of human postures from kinect data to labanotation," in *Proc. Nat. Conf. Comput. Vis., Pattern Recognit., Image Process., Graph.* Berlin, Germany: Springer, 2017, pp. 494–505.

[41] 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.

[42] Z. Zhou, "Research on automatic generation of labanotation based on dynamic programming," M.S. thesis, Beijing Jiaotong Univ., Beijing, China, 2017.

[43] J. Wang and Z. Miao, "A method of automatically generating labanotation from human motion capture data," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 854–859.

[44] 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/

[45] L. Rabiner and B. Juang, "An introduction to hidden Markov models," *IEEE ASSP Mag.*, vol. 3, no. 1, pp. 4–16, Jan. 1986.

[46] L. E. Baum, "An inequality and associated maximization technique in statistical estimation of probabilistic functions of a Markov process," *Inequalities*, vol. 3, pp. 1–8, 1972.

[47] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[48] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 2, pp. 260–269, Apr. 1967.

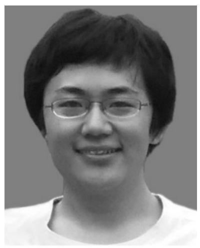
[49] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

MIN LI received the B.E. degree from Qiqihar University, in 2013, and the M.E. degree from Beijing Jiaotong University, in 2016, where she is currently pursuing the Ph.D. degree.

Her current research interests include pattern recognition, motion capturing, and human motion analysis.

ZHENJIANG MIAO (A'11–M'13) 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 of the Ecole Nationale Superieure 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, 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, the Ministry of Culture, Beijing. His current research interests include image and video processing, multimedia processing, and intelligent human–machine interaction.



CONG MA received the B.E. degree from Beijing Jiaotong University, Beijing, China, in 2013, where he is currently pursuing the Ph.D. degree.

In 2018, he was a Visiting Student with the University of California at Merced supported by the China Scholarship Council (CSC). His current research interests include pattern recognition, image and video processing, and visual tracking.