面向民族动态艺术数字化的拉班舞谱自动生成

王嘉骥*, 苗振江, 张雪艳, 张强

北京交通大学计算机与信息技术学院信息科学研究所, 北京 100044

* 通信作者. E-mail: 12112069@bjtu.edu.cn

收稿日期: 2018-11-25; 接受日期: 2019-01-09; 网络出版日期: 2019-02-18

国家自然科学基金 (批准号: 61672089, 61273274, 61572064) 和国家重点技术研发计划 (批准号: 2012BAH01F03) 资助项目

摘要 拉班舞谱是一种用于记录人体动作的符号记录体系, 它是保护、继承和传播民族民间舞蹈及其他动作艺术的有力工具. 目前, 拉班舞谱的获取主要依靠专业人员手工记录, 效率较低. 本文基于三维人体运动捕捉数据, 通过识别与拉班符号相对应的元素动作, 提出了一种自动生成拉班舞谱的方法. 方法的主要部分包括运动分割和元素动作分析. 在运动分割中, 利用速度阈值方法将人体运动捕捉数据分割成与舞谱符号对应的元素动作片段. 然后, 将动作片段与拉班舞谱的节奏相匹配, 从而确保生成的舞谱符号更加规范和准确. 在元素动作分析中, 拉班舞谱将动作区分为支撑动作和非支撑动作, 考虑到两类动作性质不同, 因此本文采用两种方法分别对支撑与非支撑动作进行有针对性的分析. 实验表明, 本文的方法实现了将运动捕捉数据表示的舞蹈动作转换为数字化的拉班舞谱, 并且自动生成的效率远高于手工记录. 所以该方法可以快速记录濒临失传的民族舞蹈, 为保护和传承动作类非物质文化遗产做贡献.

关键词 运动捕捉, 拉班舞谱, 运动分割, 元素动作分析

1 引言

1.1 背景与意义

我国是一个多民族国家, 民族民间舞蹈艺术丰富多彩. 然而, 受到传承方式不稳定以及区域发展落后的影响^[1], 部分民间舞蹈正面临着失传的危险. 这些舞蹈是中华民族的宝贵财富, 是非物质文化遗产的重要组成部分, 需要被记录、保护和传承下来.

提及舞蹈动作的记录方法, 人们首先想到的是利用直观的视频. 但是, 在一个时间点, 视频中的画面只能从一个角度来记录动作. 由于人体自身的遮挡, 经常会发生看不到视频中部分肢体动作的情况, 所以视频作为二维数据无法将三维人体动作记录完整. 运动捕捉是一种精准的三维人体动作记录

引用格式: 王嘉骥, 苗振江, 张雪艳, 等. 面向民族动态艺术数字化的拉班舞谱自动生成. 中国科学: 信息科学, 2019, 49: 229–244, doi: 10.1360/N112018-00204
Wang J J, Miao Z J, Zhang X Y, et al. Automatic generation of Labanotation for national dynamic art digitalization (in Chinese). Sci Sin Inform, 2019, 49: 229–244, doi: 10.1360/N112018-00204

图 1 拉班舞谱示例

Figure 1 Exampel of Labanotation with 4 pages

图 2 拉班舞谱的谱面结构. “L”, “C” 和 “R” 分别代表左、中和右

Figure 2 Structure of Labanotation. The “L”, “C” and “R”, represent left, center and right, respectively

方法,但是播放这种数据需要专业的软件,现在并不普及.参照乐谱对于音乐的记录作用,舞谱是一种合适的三维动作记录方法,并且阅读方式大众化.与二维视频画面相比,舞谱这种三维记录方式能够将人体动作记录完整以及更直接地展现动作的风格特征.考虑到几种记录方式各有优点,因此将舞谱、视频及运动捕捉等多种方式结合起来可以更加全面地记录和保护民族民间舞蹈艺术.但是,以目前国际上广泛使用的拉班舞谱为例,其主要获取方法是手工记录.这种获取方式相对于我国数量巨大的民间舞蹈来说记录速度太慢,无法满足需求.因此,通过计算机技术自动生成舞谱因其高效性体现出了巨大的优势.于是,利用民族民间舞蹈的三维人体运动捕捉数据生成拉班舞谱渐渐进入了研究者的视野.

1.2 拉班舞谱和运动捕捉

拉班舞谱是一种为记录人体动作而设计的符号记录体系,是“现代舞理论之父”——匈牙利人 Rudolf Laban 于 20 世纪早期所创^{[2]1)},由谱面结构和舞谱符号构成.图 1 是一个拉班舞谱的例子,共有 3 页.页内从下至上阅读,页间从左至右阅读.

如图 2 所示,拉班舞谱的谱面结构在水平方向上分为 4 到 11 栏,每一栏对应人体的一个部分,舞谱中栏的数量可以根据使用者的需求而定.如果在某一栏中存在符号,则表示对应的人体部位有动作.

1) Labanotation. <https://en.wikipedia.org/wiki/Labanotation>.

图 3 拉班舞谱的 27 种基本符号与对应的空间划分

Figure 3 27 basic symbols of Labanotation and the corresponding spatial partition

在谱面中, 符号竖直方向上的长度表示该动作的时间. 图 3 是拉班舞谱的 27 种基本符号和对应的空间划分. 每种基本符号对应一类元素动作, 符号的形状代表不同的水平方向, 填充样式代表不同的竖直方向. 水平方向有 9 种, 竖直方向有 3 种, 组合后将空间划分成 27 个子空间. 更多关于拉班舞谱的介绍请参阅文献 [3].

运动捕捉系统是一种用于准确记录运动物体在三维空间运动状况的技术设备 [4][2]. Johansson [5] 于 1973 年首次提出了运动捕捉, 现在运动捕捉技术已经广泛地应用在电影和动画制作 [6]、运动分析、人机交互等诸多领域. 运动捕捉数据的通用格式包括 BVH (bio-vision hierarchy), HTR (hierarchical translation rotation) 和 ASF/AMC (acclaim skeleton file/acclaim motion capture data) 等 [7]. 本文选择应用最为普及的 BVH 数据格式, 它由两部分组成: 第 1 部分是人体骨架结构的定义; 第 2 部分是每个骨架节点的运动数据 [8].

1.3 研究现状

拉班舞谱的自动生成是舞蹈与计算机技术的交叉学科研究方向, 目前处于刚起步的阶段. 对它的研究刚刚展开的原因是: 通过计算机自动生成拉班舞谱需要对三维数据进行分析. 虽然, 利用视频影像记录动作的方法早已存在, 但是, 视频作为二维数据无法充当计算机与拉班舞谱之间三维人体动作分析的桥梁. 所以, 在运动捕捉技术成熟之前, 拉班舞谱的自动生成研究没有正式开展的条件.

在国际上, Hachimura 和 Nakamura [9] 率先提出了通过运动捕捉数据生成拉班舞谱. 其思路是基于拉班空间分析, 研究人体上肢节点近端与远端的空间相对位置. 这种方法只能分割和分析简单的动作, 如挥动手臂. 另外, 该方法只对人体上肢建模, 对下肢和重心信息利用不足. Chen 等 [10] 通过划分空间, 判断每个时间点肢体所处的空间位置. 该方法在分析每帧数据肢体位置时, 没有考虑到帧与帧之间的联系, 导致该方法缺少对肢体运动过程的分析, 因此对涉及重心变化的下肢运动的分析不够准确. 此外, 该方法对所有时刻数据进行无差别分析的做法对噪声比较敏感, 使其容易受到坏点数据的影响. Choen-sawat 等 [11] 基于拉班空间先进行肢体运动片段提取, 然后结合弯曲度, 分析四肢的动作. Guo 等 [1, 12]

2) <https://baike.baidu.com/item/运动捕捉/4102164?fr=aladdin>.

图 4 (网络版彩图) 基于运动捕捉数据生成拉班舞谱的流程图

Figure 4 (Color online) Flow chart of generating Labanotation based on human motion capture data

用基于规则方法在自动生成拉班舞谱时考虑到了保留符号. 以上方法在进行动作分割时, 由于没有考虑到人体运动的节奏性, 所以对肢体运动的时-空信息建模不够充分. Zhou 等^[13,14]采用模板匹配的方法, 将未知动作与建立好的模板库相对比, 从而识别出未知动作的类别. 但是, 模板只能是某个人或某几个人的, 个体的差异会影响模板的普适性, 从而限制了该方法的实用性. 国内浙江大学的研究者实现了文献 [9] 中拉班舞谱的生成方法, 并将拉班符号作为三维人体运动捕捉数据的特征描述, 着重提出了基于拉班符号的运动捕捉数据的检索^[15]和编辑^[16]方法, 以及通过拉班舞谱重新生成运动捕捉数据, 所以其研究的重点在于运动捕捉数据的检索和编辑.

1.4 研究内容及文章结构

自动生成拉班舞谱的步骤如下: 首先, 输入三维人体运动捕捉数据; 其次, 进行运动分割, 得到可以运用舞谱符号表示的元素动作; 然后, 进行元素动作分析; 最后, 生成拉班舞谱. 其流程如图 4 所示.

与已有的工作^[1,9~14,17]相比, 本文对运动分割的方法进行了优化, 并且对元素动作识别方法提出了新的思路. 对于运动分割过程, 在利用速度阈值方法完成切分后, 已有方法没有考虑到利用舞谱的节奏信息规整运动分割结果. 这样会导致元素动作片段的起始点和结束点与拉班舞谱的节奏无关. 考虑到拉班舞谱是有节奏性的, 并且具有最小时间单位, 因此, 本文将分割得到的元素动作与拉班舞谱的节奏相匹配, 从而使生成的舞谱符号更加规范和准确. 对于运动分析过程, 根据拉班舞谱的记谱规则: 对于引起人体重心变化的动作, 需要侧重分析动作过程; 而对于与重心无关的动作, 则侧重分析最终姿态. 由此可见, 这两类动作的分析方法是不同的. 但在已有方法中, 均只用一种方法分析两种不同性质的动作, 这是不合理的. 因此, 本文根据两类动作的不同性质, 采用两种方法进行动作分析. 针对与重心相关的动作, 采用基于时序数据的超限学习机 (extreme learning machine, ELM) 进行着重过程的分析; 针对与重心无关的动作, 采用空间划分方法进行着重姿态的分析. 通过实验验证, 本文的方法得到了更加准确的结果, 从而将利用拉班舞谱、视频和运动捕捉数据共同记录民族民间动态艺术的构想向实际应用又推进了一步.

本文的结构安排如下: 第 1 节介绍研究背景及意义、简述拉班舞谱与运动捕捉、总结研究现状和提出研究内容; 第 2 节详细描述本文的方法, 包括运动捕捉数据分析和格式转换、运动分割和元素动作分析; 第 3 节进行实验分析, 包括系统定量评价和使用者定性评价; 第 4 节进行总结与展望.

2 基于运动捕捉数据的拉班舞谱自动生成

在本文中, 拉班舞谱的自动生成步骤包含 3 个部分: 运动捕捉数据分析和格式转换、运动分割和

元素动作分析.

2.1 运动捕捉数据分析和格式转换

在运动捕捉数据 BVH 文件中, 包含了对人体骨架模型的定义, 以及使用欧拉角记录的节点运动信息. 本文中应用的人体骨架模型包含 26 个节点.

在分析人体运动时, 用欧拉角形式描述的数据不能直观地表现节点间的相对位置, 所以需要将 BVH 格式的数据转换为三维坐标形式. 对于一个子节点 J 和其父节点 J_p , 欧拉角记录了 J_p 沿 X , Y 和 Z 轴的转动角度, 所以在运动后会产生新的角位移. 假设开始时 J 和 J_p 的相对位置是 (x_0, y_0, z_0) , 那么发生运动后的相对位置 (x_c, y_c, z_c) 可以由角位移矩阵 M 和 (x_0, y_0, z_0) 计算得到

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

对于一条节点链 $J, J_1, J_2, \dots, J_r$ (叶子节点 J 的父节点是节点 J_1 , J_r 是节点链的根节点), $M_1, M_2, \dots, M_r$ 是对应的角位移矩阵, 那么式 (1) 可以扩展为

$$\begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} = M_r \cdot \left(M_{r-1} \cdots \left(M_2 \cdot \left(M_1 \cdot \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} \right) \right) \right), \quad (2)$$

由此, 可以计算出节点链上的任意一个节点的三维坐标, 从而完成由欧拉角形式到三维坐标形式的数据转换. 更多关于欧拉角数据格式转换的介绍请参阅文献 [7].

2.2 运动分割

在三维人体运动捕捉数据中, 包含了一系列元素动作, 其中每个元素动作与一个拉班符号相对应. 为了生成拉班舞谱, 需要将运动捕捉数据切分成元素动作片段. 本文采用拉班方向切分和速度阈值分割两种方法分别切分人体重心的移动和四肢的运动, 然后利用拉班舞谱的节奏优化分割结果.

2.2.1 拉班方向切分

人体重心的移动体现了一个人整体的运动趋势, 与重心移动相关的拉班舞谱符号记录在支撑栏中. 考虑到重心的移动与四肢的运动相比相对平缓, 所以采用拉班方向切分的方法分割重心的运动.

在 BVH 文件中, 根节点位于人体骨架模型的中心, 所以将根节点的移动近似地看做重心的移动. 假设在 t_1 和 t_2 时刻, 根节点的三维坐标分别是 (x_1, y_1, z_1) 和 (x_2, y_2, z_2) , 则其运动向量可以表示为

$$\mathbf{m} = (\Delta x, \Delta y, \Delta z) = (x_2, y_2, z_2) - (x_1, y_1, z_1). \quad (3)$$

通过量化运动向量计算出重心移动的拉班方向, 然后每两帧数据进行一次方向判断, 当方向发生变化时分割重心的运动.

图 5 四二拍 (2/4) 的拉班舞谱谱面结构
Figure 5 2/4 beat rhythm of Labanotation

图 6 (网络版彩图) 骨架模型中人体平面及其法向量

Figure 6 (Color online) Body plane and vector that represents the front of human body

2.2.2 速度阈值分割

人体四肢的运动比重心灵活, 所以采用速度阈值分割方法在四肢发生停顿的位置做切分处理. 以手臂的挥动为例: 假设手臂先挥向左边, 然后挥向右边. 当手臂运动的方向发生改变时, 手臂会出现停顿并伴随一个局部最小速度, 该运动过程的切分点就在手臂速度最小的位置. 每当肢体的运动速度降低到阈值以下时, 意味着将要开始一个新的动作. 考虑到小于 0.1 s 的动作几乎不存在, 当切分出的元素动作时间过短时, 将其当做噪声处理. 对于速度阈值的选择, 考虑到不同身高的人肢体长度差别较大, 因此本文选择角速度代替肢体末端的线速度. 参考文献 [14] 中的经验数据, 本文设置的角速度阈值约为 0.18 rad/s.

2.2.3 优化分割结果

经过分割后, 连续的运动捕捉数据被切分成了一组元素动作片段, 这种分割结果与节奏无关. 但是, 拉班舞谱是有节奏的, 因此需要将分割结果与舞谱的节奏相匹配.

在拉班舞谱中, 用小节线来表示节奏. 以四二拍 (记作 2/4) 的拉班舞谱为例, 如图 5 所示, 以四分音符为一拍, 每小节有两拍. 通常在记录拉班舞谱时, 以四分之一拍作为符号的最小长度. 基于拉班舞谱自身的节奏, 本文选择最临近的四分之一拍处、二分之一拍处、整拍处或者小节线处作为符号的起始点. 分割结果经过规整, 每个符号的起始点将不会出现在不符合节奏的位置, 从而使生成的舞谱符号更加规范和准确. 此外, 一段拉班舞谱记录的应该是一段动作的基本内容, 而非某个人的特定表现^[11]. 本文利用节奏来规整分割结果的做法, 强调了动作的内容, 降低了动作的个性化表现, 因此有利于提高生成舞谱的准确率.

2.3 元素动作分析

在拉班舞谱中, 一种符号对应一类元素动作. 对于一个元素动作, 为了判断其符号, 需要确定动作的时间、对应的身体部位和方向. 动作时间可以通过运动捕捉数据的帧数来确定, 对应的身体部位可以通过人体骨架模型来确定. 方向的判断需要有参考, 在拉班舞谱中, 通常用身体的前方作为参考来量化其他方向. 如图 6 所示, 在人体骨架模型中, 选择 Spine1, LeftUpLeg 和 RightUpLeg 这 3 个节点来表示人体所在的平面, 该平面法向量的水平分量就是身体的前方.

表 1 夹角 α 的值与动作水平方向量化结果的关系Table 1 Relationships between angle α and the horizontal direction of Labanotation

Value of angle α	Horizontal direction
$[-22.5^\circ, 22.5^\circ]$	Forward
$(22.5^\circ, 67.5^\circ]$	Left forward
$(67.5^\circ, 112.5^\circ]$	Left
$(112.5^\circ, 157.5^\circ]$	Left back
$(157.5^\circ, 180^\circ] \cup [-180^\circ, -157.5^\circ)$	Back
$[-157.5^\circ, -112.5^\circ)$	Right back
$[-112.5^\circ, -67.5^\circ)$	Right
$[-67.5^\circ, -22.5^\circ)$	Right forward

表 2 夹角 β 的值与动作竖直方向量化结果的关系Table 2 Relationships between angle β and the vertical direction of Labanotation

Absolute value of angle β	Vertical direction
$[0^\circ, 30^\circ]$	High
$(30^\circ, 150^\circ]$	Middle
$(150^\circ, 180^\circ]$	Low

在识别元素动作时, 根据拉班舞谱中对支撑动作与非支撑动作的不同定义, 本文采用两种分析方式. 通常, 上肢动作不会引起重心的移动, 属于非支撑动作; 而下肢动作一般会引起重心在水平、竖直方向上的移动, 属于支撑动作. 非支撑动作的识别侧重于分析动作结束时的姿态, 本文采用基于拉班空间分析的方法进行处理; 支撑动作的识别侧重于分析动作的过程, 本文采用基于时序数据的 ELM 进行处理.

2.3.1 上肢动作分析

通过运动捕捉数据确定动作的时间和对应的身体部位后, 上肢动作采用拉班空间分析的方法确定动作的方向. 拉班空间包括 9 种水平方向和 3 种竖直方向, 见 1.2 小节图 3.

对于手臂部位的元素动作, 在判断方向时, 先找到动作最终的姿态, 然后利用肩部节点坐标和手部节点坐标相减计算出手臂向量 $\mathbf{V}$, 再分别量化 $\mathbf{V}$ 的水平方向和竖直方向.

量化水平方向时, 需要计算向量 $\mathbf{V}$ 的水平分量与 Z 轴正方向的夹角 α . 在图 3 所示的空间划分中, 水平方向上的 Z 轴正方向表示人体前方, 划分空间的相邻两条实线的夹角是 45° , 图中虚线位于角平分线上, 因此当夹角 α 取值在 $[-22.5^\circ, 22.5^\circ]$ 之间时表示动作的水平方向为前方. α 的其他取值与各个水平方向的关系如表 1 所示.

量化竖直方向时, 需要计算向量 $\mathbf{V}$ 的竖直分量与 Y 轴正方向的夹角 β , β 的计算和量化关系, 如下所示:

$$\beta = \arccos\left(\frac{\mathbf{V}_y}{|\mathbf{V}|}\right), \quad (4)$$

其中 $\mathbf{V}_y$ 是向量 $\mathbf{V}$ 的竖直分量. 在图 3 所示的空间划分中, 竖直方向上的 Y 轴正方向表示人体上方, 相邻两条实线的夹角是 60° , 图中虚线位于角平分线上, 因此 β 与动作竖直方向的关系如表 2 所示.

在确定了动作的时间、对应的身体部位和方向后, 就可以得到对应的拉班舞谱符号.

2.3.2 下肢动作分析

下肢动作大多与重心的移动相关, 通常属于支撑动作. 对支撑动作的识别需要分析其运动过程, 本文采用基于时序数据的超限学习机络进行处理. ELM 是一种单隐层前馈神经网络 (SLFNS) [18], 它适用于处理和分类时序数据.

为了强调对过程的分析, 本文用表示动作过程的时序数据作为网络的输入. 根据图 6 所示的人体骨架模型, 腿部共有 5 个节点. 每个节点有 3 个坐标维度, 因此一帧腿部动作共有 15 个数据. 对于一个下肢元素动作数据段, 假设共有 N 帧, 那么表示该动作的矩阵即为 $15 \times N$ 的实数矩阵. 因为每个元素动作对应一种拉班基本符号 (详见图 3), 所以元素动作矩阵的类别标签共有 27 种. 在训练时, 输入动作的时序数据矩阵和与之对应的类别标签. 在本文中, 超限学习机的输入层为 15 个节点 (对应一帧数据), 隐藏层使用 300 个节点, 输出层为 27 个节点 (对应 27 种分类结果). 训练的目标是获得最小的训练误差和最小的参数矩阵权重范数. 对于 L 个训练样本 $\{\mathbf{x}_i, y_i\}$, 其中 $i = 1, \dots, L$, $\mathbf{x}_i$ 为 $15 \times N$ 的实数矩阵, y_i 为类别标签, 则第 n 个隐藏节点的输出为 $g(\mathbf{x}; \mathbf{w}_n, b_n) = g(\mathbf{x} \cdot \mathbf{w}_n + b_n)$, 其中 $g(\cdot)$ 为非线性激活函数——多元二次函数 (multiquadric function), $\mathbf{w}_n$ 是输入层与该隐层节点间的权重向量, b_n 为偏置. 那么, 从输入层到隐藏层的映射矩阵为

$$\text{Mapping}(\mathbf{x}) = [g(\mathbf{x}; \mathbf{w}_1, b_1), \dots, g(\mathbf{x}; \mathbf{w}_{300}, b_{300})], \quad (5)$$

其中, 激活函数为

$$g(\mathbf{x}; \mathbf{w}_n, b_n) = \sqrt{1 + (\mathbf{x} \cdot \mathbf{w}_n + b_n)^2}. \quad (6)$$

令 u_{nm} 为第 n 个隐藏节点与第 m 个输出节点间的权重, 则输出节点与隐藏节点的关系为

$$k_m(\mathbf{x}) = \sum_{i=1}^{300} u_{im} g(\mathbf{x}; \mathbf{w}_i, b_i), \quad (7)$$

那么, 神经网络从输入层到输出层的映射矩阵可表示为

$$\mathbf{K}(\mathbf{x}) = \text{Mapping}(\mathbf{x})\mathbf{U}. \quad (8)$$

在 ELM 网络中, 权重向量 $\mathbf{w}_n$ 和偏置 b_n 被随机确定后, 隐层的输出矩阵也将被唯一确定, 那么单隐层神经网络 ELM 的训练可以转化为一个线性系统的求解问题. 详细的训练过程参见文献 [19, 20].

完成训练后, 对于一段未知的动作数据 $\mathbf{x}$, 超限学习机通过式 (9) 判断动作 $\mathbf{x}$ 的类别标签:

$$\text{label}(\mathbf{x}) = \arg_j \max k_j(\mathbf{x}), \quad (9)$$

其中 $j = 1, 2, \dots, 27$, 对应 27 个输出节点; $k_j(\mathbf{x})$ 表示动作 $\mathbf{x}$ 属于第 j 类的概率. 通过类别判断, 可以得出动作的符号种类, 即动作方向. 再通过运动捕捉数据确定动作的时间和所属身体部分, 就可以判断出符号的长度和在舞谱中的位置, 进而完成该符号的绘制.

表 3 基于规则、模板匹配、HMM 与本文的方法对四肢元素动作的识别准确率

Table 3 Comparison of the approach based on rules^[12], template^[13], HMM^[21] and our method

Accuracy (%)	Rules ^[12]	Template ^[13]	HMM ^[21]	Ours
Left arm	80.25	71.03	—	83.69
Right arm	82.50	73.42	—	83.17
Left leg	64.23	85.83	87.09	88.72
Right leg	60.71	83.90	86.62	86.24
Weighted average	68.37	80.90	—	86.20

3 实验分析

3.1 数据获取

考虑到目前公开的 BVH 格式的运动捕捉数据比较少, 因此本文构建了一个运动捕捉数据集用于实验验证. 实验中, 采集数据的运动捕捉设备是成熟的商用产品 OptiTrack 系统³⁾, 其精度可以满足实验要求. 基于已有的工作^[12, 13, 17], 从 2012 年至今, 采集的运动捕捉数据种类包括蒙古族、藏族、维吾尔族、彝族、土家族、朝鲜族、傣族、苗族、傈僳族、京族、汉族等多个民族的舞蹈, 戏曲身段、武术梅花拳、太极拳等. 为了便于追溯每段数据的文化背景和内涵, 在进行运动捕捉时为所有数据建立了对应的“动态数字资源元数据信息表”, 表中详细记录了动作名称、类别、所属民族、流传地域、表演人员籍贯和师承等信息. 现已采集舞蹈、戏曲和武术运动捕捉段落超过 200 段, 每段的时间长度以 1~4 min 居多; 采集元素动作实验数据超过 60 段, 时间长度以 2~6 min 居多. 经过运动分割, 运动捕捉数据段被切分成了与拉班基本舞谱符号对应的元素动作. 已经提取并分类的人体四肢元素动作中, 手臂非支撑动作和腿部支撑动作覆盖了全部 27 种拉班舞谱基本符号, 元素动作总数超过 20000 个.

3.2 实验结果和评价

实验包括比较不同方法识别四肢元素动作的准确率, 以及基于整段运动捕捉数据的舞谱自动生成结果分析, 这部分属于客观性评价. 此外, 还邀请了一位拉班舞谱专家记录运动捕捉数据舞谱并将自动生成结果与之对比, 并邀请一组学生体验使用本文的方法生成舞谱并进行反馈, 这部分属于主观性评价.

3.2.1 客观性评价

首先, 在已分类的四肢元素动作数据库上, 将本文的方法与基于规则的方法^[12]、基于模板匹配的方法^[13]和基于 HMM 分析下肢动作的方法^[21]进行动作识别准确率的对比. 由于拉班舞谱通常更注重对下肢支撑动作的分析, 因此本数据库中下肢元素动作的数量高于上肢. 库中共有四肢元素动作约 20000 个, 包含左臂动作 3320 个、右臂动作 3220 个、左腿动作 7100 个、右腿动作 7260 个. 每个元素动作数据段的帧率是 150 帧每秒, 本文将所有元素动作统一降采样为 20 帧. 方法^[12, 13]和本文方法以骨架模型节点的三维坐标作为每种动作识别方法的输入, 方法^[21]以其提出的“关节归一化向量”特征作为输入. 元素动作的识别结果如表 3 所示.

3) OptiTrack documentation center. <http://wiki.optitrack.com/>.

图 7 (网络版彩图) 基于鼓子秧歌运动捕捉数据生成的拉班舞谱 (部分有修改)

Figure 7 (Color online) Generated Labanotation based on the motion capture data of drum Yangko dance (partially modified)

基于规则的方法适用于分析每帧运动捕捉数据表达的人体姿态, 但缺少对运动过程的分析, 因此该方法对侧重姿态分析的上肢非支撑动作识别效果较好. 所以表格中, 基于规则方法对上肢元素动作的识别准确率明显高于下肢. 基于模板匹配的方法适用于分析连续的多帧数据, 即一段运动过程, 因此对侧重过程分析的下肢支撑动作识别效果较好. 所以表格中, 模板匹配方法对下肢元素动作的识别准确率高于上肢. 基于 HMM 的方法通过链式隐 Markov 模型建模运动数据, 然后利用 Gauss 混合模型模拟状态, 所以该方法侧重对时序数据的过程分析, 适用于对下肢支撑动作的识别. 但链式结构的描述能力在某些情况下逊色于网状的 ELM 神经网络, 因此在表格中其对左下肢元素动作地识别准确率明显低于本文, 仅右下肢略高于本文. 本文深入研究了拉班舞谱的记谱规则, 根据支撑与非支撑动作的不同性质, 采用两种方法分别进行处理, 即利用空间划分分析上肢非支撑动作的姿态以及利用时序 ELM 网络分析下肢支撑动作的过程, 因此与其他几种方法相比获得了更高的加权平均元素动作识别准确率, 提高了生成舞谱的可靠性.

然后, 本文对一个长段的运动捕捉数据的舞谱生成结果进行分析. 该运动捕捉数据段的信息如下:

- (1) 捕捉地点、时间. 中国, 北京, 文化部民族民间文艺发展中心, 2016 年 10 月 11 日;
- (2) 捕捉设备. OptiTrack 运动捕捉系统;
- (3) 动作类型. 山东鼓子秧歌传统套路片段;
- (4) 运动捕捉数据. BVH 格式, 帧率 150 帧每秒, 总帧数 6344 帧.

这段鼓子秧歌的表演者是山东省济南市济阳县文化馆老馆长, 在他录制完成回程后, 因意外事故不幸离世, 所以这段数据弥足珍贵. 为了更好地展示这段动作, 本文将多路视频数据、运动捕捉数据, 以及生成的舞谱数据制作成可以同步演示的视频, 以便于直观地理解本文提出的动态艺术数字化方法. 生成的拉班舞谱共 4 页, 如图 7 所示. 为了将生成的舞谱与原始动作进行对比, 在运动捕捉数据中以大约 5 s (约 700~800 帧) 为间隔, 选取了 9 处动作将其与对应的舞谱符号进行分析, 如图 8 所示. 与选取动作对应的 9 帧同步演示视频如图 9 所示.

图 8 中的舞谱符号与人体动作分析如下, 先分析腿部支撑动作或非支撑腿的姿态, 然后分析手臂姿态. 本文在分析动作和姿态时, 先说水平方向再说竖直方向, 例如: 前中、左前高、原位低等, “前”、“左前”、“原位”表示水平方向, “中”、“高”、“低”表示竖直方向. 水平和竖直方向的名称以及划分见图 3. ①为初始动作. 人体呈自然站立状态, 重心处于中位, 因此用原位中表示人体双腿的支撑动作. 双臂位于身体两侧, 自然下垂, 因此用原位低表示双臂姿态. ②中右脚原地抬起, 人体重心处于中位, 用原位中表示右腿支撑动作. 左臂姿态为左前低, 右臂姿态为右前高. ③中左腿向前迈步使重心水平前移, 用前中表示左腿支撑动作. 左臂姿态左前高, 右臂姿态前中. ④中右腿向后迈步使重心水平后移, 用

图 8 (网络版彩图) 生成的舞谱符号与对应的原始动作比较

Figure 8 (Color online) Comparison of original human motion and the corresponding generated Labanotation

图 9 (网络版彩图) 用视频数据 (三路)、运动捕捉数据和生成的拉班舞谱合成的山东鼓子秧歌传统套路片段视频截图。截图共 9 张, 每张左侧为三路不同角度拍摄的实景视频, 中间为运动捕捉数据, 右侧为对应的拉班舞谱

Figure 9 (Color online) Video screenshots of traditional routine clips of Shandong drum Yangko that synthesized with video data (three channels), motion capture data and generated Labanotation. There are nine screenshots, each of which is a live video shot from three different angles on the left, with motion capture data in the middle and corresponding Labanotation on the right

后中表示右腿支撑动作. 左臂姿态前低, 右臂姿态右前中. ⑤ 中左腿向前迈步使重心水平前移, 用前中表示左腿支撑动作. 左臂姿态左前高, 右臂姿态右前高. ⑥ 中双腿的支撑栏有一个瞬间均没有符号, 表示存在跳跃. 左腿跳跃后支撑动作为原位中, 右腿没有接触地面, 其姿态为原位低. 左臂姿态左高, 右臂姿态右高. ⑦ 中双腿的支撑栏有一个瞬间均没有符号, 表示存在跳跃. 右腿跳跃后支撑动作为原位低, 左腿没有接触地面, 其姿态为原位低. 左臂姿态左高, 右臂姿态原位高. ⑧ 中人体双腿弯曲站立, 重心处于低位, 用原位低表示双腿支撑动作. 左臂姿态前低, 右臂姿态前中. ⑨ 中人体双腿分开站立, 重心处于中位, 左腿支撑动作为左中, 右腿支撑动作为右中. 左臂姿态左高, 右臂姿态右高. 分析结果证明, 9 处舞谱符号均可以表达出对应的人体动作, 因此, 生成的拉班舞谱可以反应该段动作的主要内容, 即生成的舞谱基本正确.

在图 9 中, 多路视频、运动捕捉、拉班舞谱 3 种记录方式生动形象地表现出了这段山东鼓子秧歌传统套路的动作.

3.2.2 主观性评价

考虑到拉班舞谱的人工记录具有一定的主观性, 即使是两位拉班舞谱专家对同一段动作进行记录也可能不会完全相同, 所以仅用客观性的分析不足以全面地衡量生成的舞谱结果. 为此, 本文邀请了著名的拉班舞谱专家罗秉钰老师对多段基础动作的运动捕捉数据进行记录并将记录结果与本文生成的舞谱结果进行对比, 另外还邀请了一组具有拉班舞谱相关研究经验的学生对本文的结果进行评价和反馈. 罗老师是原空政文工团一级编导, 译有《拉班记谱法》^[22], 对拉班舞谱有近 40 年的记录和研究经历, 绘谱经验丰富. 另一组中包括 9 位学生, 每人接触和学习拉班舞谱的时间是 6 个月至 3 年不等, 其中 8 人已经发表与拉班舞谱有关的文献, 6 人参加过北京航空航天大学开设的关于拉班舞谱的课程, 因此每位参与评价的学生都对拉班舞谱有比较深入的了解.

本文将专家记录的 6 段基础动作动捕数据的舞谱与自动生成的舞谱进行了对比. 6 段基础动作中包含 3 类动作: 腿部支撑动作、腿部姿态动作和手臂姿态动作. 舞谱结果对比如图 10 所示. 每类动作的舞谱符号分别以图 10(a), (c), (e) 为例进行解析如下. 在腿部支撑动作中 (图 10(a)), 该组动作为在自然状态下人体向前走: 第 1 个符号表示人体右腿向前迈步, 牵引人体向前移动, 对应拉班舞谱右支撑栏符号前中; 第 2 个符号表示人体左腿向前迈步, 牵引人体向前移动, 对应拉班舞谱左支撑栏符号前中; 第 3 个符号与 1 相同. 在腿部姿态动作中 (图 10(c)), 该组动作为右腿的两个姿态: 第 1 个符号表示右腿姿态水平方向为前方, 竖直方向为低位; 第 2 个符号表示右腿水平方向向右, 竖直方向为低位. 在手臂姿态动作中 (图 10(e)), 该组动作为右臂的 3 个姿态: 第 1 个符号表示右臂姿态水平方向为前方, 竖直方向为中位; 第 2 个符号表示右臂水平方向向右, 竖直方向中位; 第 3 个符号表示右臂水平方向为后方, 竖直方向中位. 在图 10 中, 6 段生成舞谱中的动作符号均与专家绘谱一致, 且能够正确表达对应的肢体动作; 但动作符号在舞谱中的位置与专家绘谱不一致, 符号在舞谱中的位置代表了时间、节奏, 生成舞谱与专家绘谱的节奏不同是因为计算机会把每个动作的起始、结束位置定位的相当精准 (动捕数据的帧率为 150 帧每秒, 所以符号的时间精度为 $1/150$ s), 而人眼则会把节奏差不多的动作按照用时相同来处理, 而人体不可能分毫不差地按照节奏做动作, 所以对于符号的时间定位生成舞谱与专家绘谱不一致. 综上, 可以主观判定 6 段生成舞谱记录的动作与专家记谱十分接近, 生成舞谱结果正确.

另一组参评学生在使用了本文方法实现的舞谱生成系统后, 每人完成了一份调查问卷, 问题以及答案的统计如附录 A 所示. 对于问题 1, 使用者对生成舞谱的准确率评价为 68%~93%. 换言之, 生成的大多数舞谱符号是正确的, 但是对于某些动作系统还处理不好. 其原因是: 动作的分析方法主要针对

图 10 6 段基础动作捕数据的专家记谱与生成舞谱对比

Figure 10 Comparison of expert records and generated Labanotation of six kinds of basic motion. (a) Go forward; (b) go right forward; (c) forward low, right low; (d) forward low, origin low; (e) forward, right, backward; (f) backward, left, forward

的是 27 类拉班基本动作, 对复杂的旋转、抖动等动作处理能力不足。所以, 对于主要由基本动作构成的运动序列, 系统可以很好的生成舞谱, 而对于包含复杂动作的运动序列, 系统对其处理能力会降低。对于问题 2, 绝大多数参评者认为, 用生成的舞谱作为参考可以帮助他(她)们更好地完成拉班舞谱的记录任务, 并且在问题 3 的回答中得到了验证: 对于参评者独立完成的舞谱, 其准确率为 70%~90%, 而利用生成舞谱作为参考后准确率有所提高, 为 75%~93%。对于问题 4, 文本会在保持已有优点的同时, 将提出的缺点作为系统未来的改进和研究方向。

4 结论与展望

拉班舞谱是一种科学的、具有逻辑性的人体动作记录体系, 将其作为民族动态艺术的记录方式有助于保护和传承我国的民族民间舞蹈。本文基于三维人体运动捕捉数据提出了一种自动生成拉班舞谱的方法, 其中运动分割和元素动作分析是方法中的两个主要部分。在运动分割中, 利用拉班舞谱的节奏信息来规整分割结果的思想使生成的舞谱符号更加规则和准确。在动作分析中, 本文提出了用不同的方法分别处理性质不同的上肢和下肢动作。原因是在拉班舞谱中, 从动作分析的角度, 上、下肢动作的分析方法完全不同。然而在已有方法中, 如基于规则的方法或基于模板匹配的方法, 均使用一种思路

来处理该问题, 因此与已有方法相比, 本文在理论思想上具有优越性, 而且在实验中获得了准确度更高的结果. 对于人体四肢元素动作的识别, 本文的平均准确率超过了 85%, 基本满足实际应用的需求. 此外, 自动生成舞谱的方法解决了人工记录效率较低的问题, 用自动生成代替人工记录能提高拉班舞谱的绘制效率, 进而为使用舞谱、视频和运动捕捉数据 3 种方式共同记录和传承民族民间动态艺术——动作类非物质文化遗产做出了重要贡献.

致谢 本文感谢罗秉钰专家在拉班舞谱方面的耐心指导, 感谢李松专家在民间文化与技术结合方面的巨大帮助, 感谢文化部民族民间文艺发展中心在设备和数据上的大力支持.

参考文献

- 1 Guo H. Research on automatic generation of Labanotation based on human motion capture data. Dissertation for Master Degree. Beijing: Beijing Jiaotong University, 2015 [郭浩. 基于人体运动捕捉数据自动生成拉班舞谱的研究. 硕士学位论文. 北京: 北京交通大学, 2015]
- 2 Hutchinson A. Labanotation. *J Am Folklore*, 1955, 68: 89
- 3 Guest A H. Labanotation or Kinetography Laban: the System of Analyzing and Recording Movement. 3rd ed. New York: Theatre Arts Books, 1970
- 4 Xiang Z R, Zhi J Y, Xu B C, et al. Motion capture technology and its application research review. *Comput Appl Res*, 2013, 30: 2241–2245 [向泽锐, 支锦亦, 徐伯初, 等. 运动捕捉技术及其应用研究综述. 计算机应用研究, 2013, 30: 2241–2245]
- 5 Johansson G. Visual perception of biological motion and a model for its analysis. *Percept Psychophys*, 1973, 14: 201–211
- 6 Villegas R, Yang J M, Ceylan D, et al. Neural kinematic networks for unsupervised motion retargetting. 2018. ArXiv:1804.05653
- 7 Li T. Research on human motion editing technology based on motion capture data. Dissertation for Ph.D. Degree. Wuhan: Huazhong University of Science and Technology, 2008 [李婷. 基于运动捕获数据的人体运动编辑技术研究. 博士学位论文. 武汉: 华中科技大学, 2008]
- 8 Liang Q H. Research on key technology of motion capture in digital dynamic art. Dissertation for Ph.D. Degree. Beijing: Beijing Jiaotong University, 2016 [梁清华. 动态艺术数字化中的运动捕捉关键技术研究. 博士学位论文. 北京: 北京交通大学, 2016]
- 9 Hachimura K, Nakamura M. Method of generating coded description of human body motion from motion-captured data. In: *Proceedings of the 10th IEEE International Workshop on Robot and Human Interactive Communication*, 2001. 122–127
- 10 Chen H, Qian G, James J. An autonomous dance scoring system using marker-based motion capture. In: *Proceedings of the 7th Workshop on Multimedia Signal Processing*, 2005
- 11 Choensawat W, Nakamura M, Hachimura K. GenLaban: a tool for generating Labanotation from motion capture data. *Multimed Tools Appl*, 2015, 74: 10823–10846
- 12 Guo H, Miao Z J, Zhu F Y, et al. Automatic labanotation generation based on human motion capture data. In: *Proceedings of Chinese Conference on Pattern Recognition*, 2014. 426–435
- 13 Zhou Z M, Miao Z J, Wang J J. A system for automatic generation of labanotation from motion capture data. In: *Proceedings of the 13th International Conference on Signal Processing (ICSP)*, 2016. 1031–1034
- 14 Zhou Z M. Research on automatic generation of labanotation based on dynamic programming. Dissertation for Master Degree. Beijing: Beijing Jiaotong University, 2017 [周子鸣. 基于动态规划的拉班舞谱自动生成研究. 硕士学位论文. 北京: 北京交通大学, 2017]
- 15 Yu T, Shen X J, Li Q L, et al. Motion retrieval based on movement notation language. *Comp Anim Virtual Worlds*, 2005, 16: 273–282
- 16 Shen X J, Li Q L, Yu T, et al. Mocap data editing via movement notations. In: *Proceedings of International Conference on Computer Aided Design and Computer Graphics*, 2006. 463–470

- 17 Wang J J, Miao Z J, Guo H, et al. Using automatic generation of Labanotation to protect folk dance. *J Electron Imag*, 2017, 26: 011028
- 18 Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: theory and applications. *Neurocomputing*, 2006, 70: 489–501
- 19 Huang G B, Zhou H M, Ding X J, et al. Extreme learning machine for regression and multiclass classification. *IEEE Trans Syst Man Cybern B*, 2012, 42: 513–529
- 20 Huang Z Y, Yu Y L, Gu J, et al. An efficient method for traffic sign recognition based on extreme learning machine. *IEEE Trans Cybern*, 2017, 47: 920–933
- 21 Li M, Miao Z. Automatic Labanotation generation from motion-captured data based on hidden Markov models. In: *Proceedings of the 4th Asia Conference on Pattern Recognition*, 2017
- 22 Guest A H. *Laban Recording Method: Action Analysis and Recording System*. Beijing: China Foreign Translation and Publishing Co., Ltd., 2013 [Guest A H. 拉班记谱法: 动作分析与记录系统. 北京: 中国对外翻译出版有限公司, 2013]

附录 A 调查问卷中的问题和回答统计

问题 1: 根据你的经验, 系统生成的拉班舞谱的准确率大概是多少?

回答 1: 68%~93%.

问题 2: 对于一个拉班舞谱记录任务, 你会选择用本系统生成的拉班舞谱作为参考吗? 如果会, 对你的帮助有哪些?

回答 2: 8 人选择会. “帮助” 归纳如下: 有辅助作用, 可以为记录舞谱提供思路, 缩短记录拉班舞谱的时间; 对于有歧义性的动作可以提供参考.

问题 3: (看完一段运动捕捉数据完成舞谱记录后) 根据你的经验, 记录的拉班舞谱准确率大概是多少? 在生成系统的辅助下, 记录的准确率大概是多少?

回答 3: 70%~90% (自己记录); 75%~93% (系统辅助).

问题 4: 反馈意见.

回答 4: (1) 生成的舞谱能够反映整体性的动作特点、节奏, 因此可以为记录任务提供思路、减少工作量, 效率提升约 20%~50%. (2) 系统只能处理一个人的运动数据, 对于具有交互性的双人动作无法处理; 对于简单的节奏分明的动作处理的较好, 对于复杂的旋转动作处理的不好.

Automatic generation of Labanotation for national dynamic art digitalization

Jiaji WANG*, Zhenjiang MIAO, Xueyan ZHANG & Qiang ZHANG

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

* Corresponding author. E-mail: 12112069@bjtu.edu.cn

Abstract Labanotation is a symbol system for recording human motions. It is a powerful tool for protecting, inheriting and disseminating folk dances and other action arts. At present, Labanotation data are mainly recorded manually by professionals, and the working efficiency of the system is low. This paper proposes a method that automatically generates Labanotation from three-dimensional human-motion capture data. The method identifies the elemental movements corresponding to basic Laban symbols. The method contains two components: motion segmentation and unit movement analysis. The motion segmentation divides the motion capture data into segments that are easily identified by the speed threshold method. To standardize the Laban symbols and improve their accuracy, the motion segments are then matched to the Labanotation rhythms. In the unit movement analysis, Labanotation divides the movements into support and non-support ones. As the two movement types have different characteristics, they are identified by different methods. In experiments, the proposed method successfully transformed the movements represented by the motion-capture data into digitized Labanotation. The automatic Labanotation generation is far more efficient than manual recording. Therefore, the method can quickly record ethnic dances on the verge of loss, helping to protect and preserve the intangible cultural heritage of action arts.

Keywords motion capture, Labanotation, motion segmentation, unit movement analysis

Jiaji WANG received his B.E. degree from Beijing Jiaotong University in 2012, and is currently pursuing his Ph.D. degree. His current research interests are pattern recognition, image processing, multi-view computer vision, multi-view human motion capture, and recognition based on motion-capture data.

Zhenjiang MIAO received his B.E. degree from Tsinghua University in 1987, Beijing, China, and his M.E. and Ph.D. degrees from Northern Jiaotong University, Beijing, in 1990 and 1994, respectively. He joined Beijing Jiaotong University, Beijing, in 2004. He is currently a professor at Beijing Jiaotong University. His current research interests are image and video processing, multimedia processing, and intelligent human-machine interactions.

Xueyan ZHANG received her B.E. degree from Tianjin Normal University in 2012, and is currently pursuing her M.S. degree at Beijing Jiaotong University. Her current research interests are pattern recognition, image processing and human-computer interactions.

Qiang ZHANG received his B.E. degree from Beijing Jiaotong University in 2010, and is currently pursuing his Ph.D. degree. His current research interests are pattern recognition, image processing and multi-view computer vision.