

人体运动捕获数据的重用技术综述

孙怀江 夏贵羽 张国庆 封磊

(南京理工大学计算机科学与工程学院, 南京, 210094)

摘要: 人体运动捕获数据作为一种新型的多媒体数据, 因其高度的逼真性, 已经在多个领域得到了广泛的应用。但运动捕获设备的高昂价格导致运动捕获数据使用的成本过高, 所以运动重用技术就成为解决这一问题的有效手段。可是运动捕获数据复杂的结构和特性给运动重用带来了很大的挑战, 尽管已经经过多年的研究, 但仍然有很多尚未解决的问题, 需要研究人员更多的关注和研究投入。本文针对运动重用过程中各个环节的重要技术, 从研究的意义、问题的难点、现有方法思路及使用的模型等多个角度进行了介绍, 并针对一些具有代表性的方法做了较为详细的描述。最后总结了现有运动捕获数据重用技术的研究进展, 展望了未来的发展趋势。本文旨在引起人们对该领域的深入思考, 并为以后的研究提供有价值的参考。

关键词: 人体运动; 计算机动画; 运动捕获; 运动重用; 机器学习

中图分类号: TP39 **文献标识码:** A

Survey on Reuse of Human Motion Capture Data

Sun Huaijiang, Xia Guiyu, Zhang Guoqing, Feng Lei

(School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, 210094, China)

Abstract: Human motion capture data, as a new type of multimedia data, is widely used in many areas because of its high fidelity, but the expensive motion capture equipment yields the high cost of the use of motion capture data. Therefore, the technologies of motion capture data reuse become the effective means to solve the problem. However, the complex structure and characteristics of motion capture data make the motion capture data reuse challenging. Even it has been researched for many years, there are still many problems to be solved and more attentions and research efforts are needed. In this paper, in terms of the important technologies used in the process of motion capture data reuse, we give introductions on the research significance, difficulties, strategy and used models of current methods and so on. And we give a detailed description on some representative methods. Finally, we conclude the research advances of motion capture data reuse and discuss the possible directions for future works. This aims to cause the deep thinking of this field and provides a valuable reference for the future research.

Key words: human motion; computer animation; motion capture; motion reuse; machine learning

引言

随着运动捕获技术的日趋成熟和虚拟现实技术的高速发展,运动捕获数据作为一种新兴的多媒体数据,被用于电影、游戏、医疗和体育等多个领域。近年来,多部运动捕获数据驱动角色的动画电影都取得了不错的票房成绩,运动数据高度的逼真性赋予动画角色生命力,给观众带来了超越现实的视听感受,其中就包括大家所熟知的电影《魔兽》和《超能陆战队》(见图1)。除了在电影制作上,运动捕获数据在电子游戏领域也取得了令人瞩目的成就。运动捕获数据的应用使得游戏人物更加真实,打斗动作更加流畅,带来了更好的游戏体验。此外,对运动员技术动作的捕获可以让教练更好地分析运动中存在的问题,从而做出相应的调整。

光学运动捕获系统,如 Vicon,是一类常用的运动捕获系统,演员的各个关节位置会被贴上一些反光的标签,捕获系统中的多个摄像机会跟踪这些反光标签在运动中的位置变化,经过处理之后被记录成某一时刻人体各个关节的空间位置或朝向,也就是运动中的一帧,如图2,一个序列的运动帧就构成了一段人体运动。

虽然运动捕获数据的出现给人们的生活带来了许多积极的影响,但一些尚未解决的问题却给运动捕获数据的推广带来了困难,而问题的根源在于运动捕获的成本太高。尽管市场上已经出现了一些廉价的运动捕获设备,如微软公司的 Kinect,但此类设备捕获的精度太低,导致捕获的运动数据只能用于一些对数据质量要求不高的应用,而高精度运动捕获设备的价格则高达数十万甚至上百万,只有少数的大公司和科研机构才能进行高精度的人体运动捕获。并且,从硬件技术上来说,运动捕获的高成本问题在短时间内无法解决,所以利用已有的运动捕获数据生成新的符合各种应用需求的运动数据,即运动重用,就成了最有效的解决方法。可是运动数据重用是一项十分具有挑战性的工作,因为人体运动是一种复杂的自然现象,并且包含一些重要的特性:

(1)黎曼流形结构。人体运动是一种典型的非线性数据。人体骨架的树形结构导致每个关节相对于其父关节的运动严格地处于一个球上,球的半径就是这段骨骼的长度。而球则是一个最简单的黎曼流形,所以人体运动从几何上具有黎曼流形结构^[1-3]。



(a) 《魔兽》
(a) 《Warcraft》



(b) 《超能陆战队》
(b) 《Big Hero 6》

图1 动画电影海报

Fig.1 Posters of two animated movies

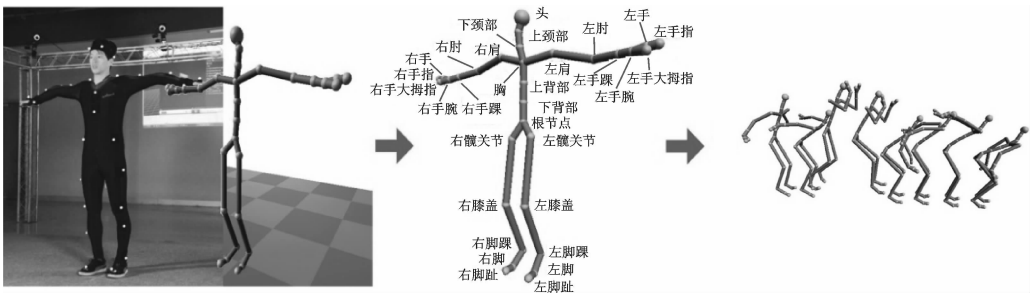


图2 光学运动捕获系统

Fig.2 Optical motion capture system

(2)低秩和稀疏性。这是人体运动的一个本质属性。在一个运动序列中,大多数运动都是可以由少数几个关键帧进行运动插值产生,运动插值可以理解为是流形上的线性变换,这与低秩的定义十分相符,那些关键帧的个数也可以理解成该运动序列的秩。而低秩和稀疏在本质上相同,从稀疏表示^[4-5]的角度来说,关键帧自然也就可以理解成最具代表性(系数不为0)的帧。

(3)时序连续性。人体运动是一种序列数据,每一帧都与前后帧有着密不可分的关系,在进行各种操作时都要考虑到是否会破坏运动的时序连续性。

(4)含噪性。除了捕获系统的电子设备会产生一些噪声之外,演员的表演也不可避免地会产生一些噪声。例如,运动过程中衣服的滑动会导致反光标记位置的突变,身体和衣服的相互遮挡会导致标记在镜头中消失,这些噪声都属于非高斯噪声,不易处理。

(5)运动学特性。人体运动自然具备一些运动学特性,例如平滑性、骨段长度的刚性和动作的合理性。

运动重用的最终目的是生成新的运动数据,即运动合成,但为了完成运动合成,还需要完成许多前期的工作,这也就衍生出了一系列运动重用过程中的相关技术,这些技术基本上贯穿了运动捕获数据重用的整个周期。如图3所示,建立一个可持续运作的运动捕获数据重用系统大约需要4个步骤:捕获后处理、数据存储、数据浏览以及最终的运动合成。

(1)捕获后处理。运动捕获是一个复杂的过程,除了演员表演外还涉及到各种机械和电子设备,任何一个环节都可能产生噪声,而噪声会严重影响运动数据的使用。因此,运动数据的捕获后处理不仅仅是运动数据重用的一个重要步骤,也决定了后续很多应用的质量。

(2)数据存储。运动数据通常是以120帧/秒的频率被记录,这为数据的存储和浏览带来了很大的挑战。如果仍按120帧/秒的频率进行存储不仅占用过大的空间,也为查询和浏览带来了不便。所以进行数据压缩能够有效地减少存储成本,方便数据浏览,但运动数据的特殊性导致其需要定制的压缩算法。

(3)浏览。与视频类似,如果浏览整个运动序列必然效率低下,用户只需要知道运动的主要内容就能够决定是否需要该运动。构建一个有效的浏览机制需要解决两个重要问题:(1)如何为每一个运动生成一个高质量的摘要;(2)如何根据用户提供的输入快速、准确地搜索到满足要求的运动。

(4)运动合成。运动合成是运动重用最终的目的。高效便捷的运动合成算法能够有效地提高运动重用的普及度,降低运动数据的使用成本。

这其中囊括了6项基本技术:运动恢复和去噪、运动分割、运动压缩、运动关键帧提取、运动检索和运动合成。但每种技术并不绝对地只适用于其中一个步骤,图3反映了它们之间的大致关系,如运动分割就是一个基础性的操作,在多种应用之前都有可能需要进行运动分割。近年来,由于机器学习方法的兴起,传统的信号处理和运动学方法在这几项技术上已经失去了优势,因为传统的方法并不能很好地对上述运动数据的一些重要特性进行建模,而合适的机器学习方法不仅能够对其进行建模,还能有效地利用这些特性完成特定任务。

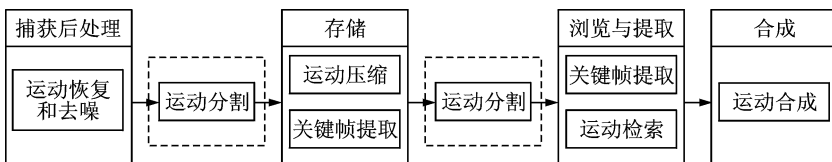


图3 运动捕获数据重用流程与相关技术

Fig. 3 Process of motion capture data reuse and related technologies

1 运动恢复与去噪

运动恢复和运动去噪在任务上略有差别,前者致力于恢复在捕获过程中丢失的关节位置,后者重点在于抗击噪声,但它们所使用的技术却是相似的。早期研究人员使用高斯低通滤波器和卡曼滤波器^[6-7]对运动数据进行去噪,此类方法单独处理运动数据的每一个自由度,但却破坏了运动数据的时空特性。而基于插值的方法^[8-10]能够保留运动数据的时空特性,但它们只能处理一些短时间的运动恢复问题。因为插值只是一种简单线性方法,当某一标记丢失时间变长时,其非线性特征就更加明显,线性方法的劣势也就显现出来。

鉴于传统方法的局限性,研究人员开始使用机器学习方法来进行人体运动的去噪和恢复。Lou 等^[11]首先从完整、无噪的运动数据中学习得到一组滤波基,然后用鲁棒的统计方法滤掉了运动中的噪声。这是一种典型的数据驱动的方法,该方法能够获取训练数据中的统计特性,但无法保证去噪后运动的细腻程度,因为滤波的过程有细节损失。稀疏表示是另一类重要的数据驱动的运动恢复和去噪方法,此类方法利用了运动数据稀疏特性,并基于这样一个假设:完整和无噪声的运动帧的稀疏表示系数与不完整或含噪的运动帧的稀疏表示系数相同或者相似。基于这样一个假设,Xiao 等^[12]将运动恢复问题转化成了稀疏表示问题。用完整的运动数据集做字典 \mathbf{D} 求一个不完整运动帧 $\tilde{\mathbf{f}}$ 的稀疏表示,对应的不完整的字典的表示为 $\tilde{\mathbf{D}}$,则运动恢复的目标函数为

$$\min_{\alpha} \|\tilde{\mathbf{f}} - \tilde{\mathbf{D}}\alpha\|_F^2 + \lambda \|\alpha\|_1 \quad (1)$$

式中: λ 为一个调谐参数,根据之前的假设,当得到 α 之后,可以认为 α 也是完整运动帧 \mathbf{f} 的稀疏表示,这样就可以用

$$\mathbf{f} = \mathbf{D}\alpha \quad (2)$$

来预测丢失的数据。在这样的一个理论框架下,字典 \mathbf{D} 决定了运动恢复的质量,字典太小则其包含的信息不够完备,字典太大则会导致求解问题(1)太过耗时,所以 Xiao 等维护了一个规模中等的字典,但根据恢复运动类型以及字典中原子使用的频率不断地更新字典中的内容,这样就能两者兼顾。

仅仅使用一个未经学习的字典并不能很好地获取和凝练藏于数据的有效信息,所以 Hou 等^[13]学习了一个由各个关节的运动轨迹构成的字典,然后用类似的方式进行运动恢复。但该方法存在两个问题:(1)轨迹的数量过于庞大,导致字典的学习时间太长;(2)连续丢失的轨迹长度有可能超过字典中原子的长度,这就导致该方法无法正常工作。为了获得更为细腻的去噪效果,Xiao 等^[14]将人体运动拆分成 5 个部分(四肢和躯干)分别构建字典并进行去噪,将得到的去噪后的各部分运动再重新组合在一起。该方法得到的运动细节更为丰富,但忽略了人体各部分之间的关系,有可能导致去噪后的运动肢体不协调。鉴于以上方法存在的问题,Xia 等^[15]在稀疏表示的框架下,综合利用运动数据的统计信息和运动学信息进行运动恢复。他们首先通过实验的方式验证了上文提到的假设的可靠性,如图 4 所示。

从图中可以看出,完整运动帧和不完整运动帧在同一个字典下的稀疏表示系数非常接近,即便如此,Xia 等^[15]还通过修改传统的字典学习方式使这种相似性取得一定程度上的理论保证,其字典学习的目标函数为

$$\begin{aligned} & \min_{\alpha} \sum_{i=1}^N \sum_{k=1}^k \frac{1}{2} \|\overline{\mathbf{A}^k} \mathbf{f}_i - \overline{\mathbf{A}^k} \mathbf{D} \mathbf{z}_i^k\|_2^2 \\ \text{s. t. } & \mathbf{z}_i^k = \arg \min_{\alpha} \|\mathbf{A}^k \mathbf{f}_i - \mathbf{A}^k \mathbf{D} \alpha\|_2^2 + \lambda \|\alpha\|_1 \quad i=1,2,\dots,N; k=1,2,\dots,K \\ & \|\mathbf{D}(:,j)\|_2 \leq 1 \quad j=1,2,\dots,n \end{aligned} \quad (3)$$

式中: \mathbf{f}_i 为完整的运动帧; \mathbf{A}^k 为退化算子,用于模拟各种关节丢失情况; $\overline{\mathbf{A}^k}$ 为 \mathbf{A}^k 的共轭算子。该目标函数与传统字典学习目标函数的主要差别在于用于更新字典的是不完整帧的稀疏表示系数,这就通过数学模型强化了不完整帧与完整帧之间稀疏表示系数的相似性,使得假设更加可靠。利用学习得到的字

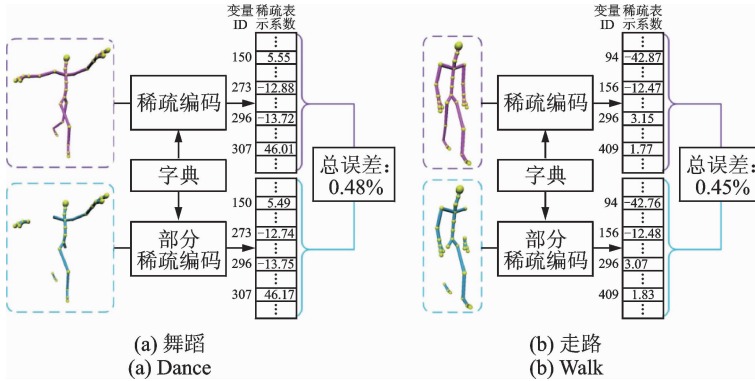


图4 完整运动帧和不完整运动帧的稀疏表示系数对比

Fig. 4 Comparison between sparse representations of complete and incomplete frames

典,计算不完整帧的稀疏表示系数之后,他们并没有直接使用式(2)的方式恢复运动,而是加上两个运动学约束——骨段长度约束和平滑约束再次进行优化。这使得恢复后的运动不仅满足了运动学特性,而且细节更加丰富。虽然以上的运动恢复和去噪算法取得了不错的效果,但它们同样存在一个数据驱动算法的普遍问题——out-of-sample问题。也就是当用于训练的数据中如果不包含待恢复或去噪样本的运动类型时,此类方法往往无法获得令人满意的效果。一些研究人员^[16-19]提出用低秩矩阵恢复^[20]进行运动恢复和去噪。根据低秩的定义可知,如果一个矩阵是低秩的,那该矩阵中的行或列可以由少数的一些行或列通过线性变换得到。当其中某些元素丢失或含有噪声时,就破坏了矩阵的低秩性,所以可以通过最小化矩阵秩的策略来填补某些丢失元素或抑制噪声,这也就是利用低秩矩阵恢复进行运动恢复和去噪的基本原理。Lai等^[17]利用人体运动的低秩特性首次提出基于低秩矩阵恢复的运动恢复模型,用 M 表示一个含有丢失关节数据的运动,其中 M 的每一列代表一帧,用 X 表示完整的运动,则

$$\begin{aligned} & \min_{\mathbf{X}} \text{rank}(\mathbf{X}) \\ & \text{s. t. } \mathbf{H} \circ \mathbf{X} = \mathbf{M} \end{aligned} \quad (4)$$

式中: \circ 代表矩阵中每个元素相乘; \mathbf{H} 为一个状态矩阵,表示 \mathbf{M} 中每一个元素的丢失况,如果 M_{ij} 丢失,则 $H_{ij}=0$,否则 $H_{ij}=1$,直接最小化矩阵的秩是一个NP-Hard问题,通常的做法是用 x 核范数代替矩阵的秩进行优化。随后Tan等^[19]改进了这一方法,他们用一个由各个关节运动轨迹构成的矩阵代替了由运动帧构成的矩阵 \mathbf{M} ,因为前者的秩比后者更小,而矩阵的秩越低,恢复的结果越可靠。Feng等^[16]则不仅考虑到了人体运动的低秩特性,同时还考虑到了运动学特性,他们在低秩恢复的目标函数里加上了平滑约束项,避免恢复后的运动出现抖动,影响数据的使用。

2 运动分割

运动分割是一项基本的操作,在多种基于运动捕获数据的应用中都有可能需要进行运动分割。运动分割虽然只是一个序列分割的问题,但因为人体运动数据的复杂结构,导致运动分割远不止想象的那么简单。在早期的方法^[21-22]中通常先提取运动序列的一些物理特征,如速度、加速度和关节夹角等,然后对提取的运动特征进行处理,根据这些特征的分布变化,如局部极值,提取一段运动序列的分割点。此类方法方便、快捷易于实现,但仅仅利用了运动数据的低层次的特性,使得分割结果缺乏语义含义。为了获取运动数据的高层次信息,Barbic等^[23]提出了3种方案,即分别利用主成分分析(Principal component analysis, PCA),概率主成分分析(Probabilistic principal component analysis, PPCA)和高斯混合模型(Gaussian mixture model, GMM)进行运动分割。具体地说,基于PCA的方法只保留前 ρ

个主成分,然后分别利用这前 ρ 个主成分重构每一帧,当重构误差增加过快的时候说明出现了一个分割点,该方法利用原始数据的统计分布代替了原先的物理特征的分布进行分割,提高了利用信息的层次。Barbic 等首先用 PPCA 对运动数据的概率分布进行估计,然后仍然采用与之前类似的策略,在概率分布变化最快的位置进行分割。基于 GMM 的方法假设不同的运动片段来自于不同的高斯分布,所以运动分割问题就转化成了在不破坏时序连续性的前提下为运动序列寻找最优的高斯分布问题。此类方法也十分简单、高效,但只适用于短运动的分割,当运动过长时,其分割效率将大大降低。Gong 等^[24]将运动分割转化成了一个一维搜索问题,并为此提出了一种核化的时序分割算法(Kernelized temporal cut, KTC)^[25]。该方法使用一个核函数计算帧与帧之间的内积,并设计了一个以帧与帧之间关系为基础的目标函数,不同的分割点会影响目标函数的值。然后用在一个预先设定的窗口内,通过一维线性搜索的方式寻找使得目标函数最小的帧,即为分割点。接着以该分割点为起点,开始新一轮的搜索,直到完成所有的分割。该方法最大的问题在于窗口的大小难以确定,窗口过小,容易导致过分割,窗口过大会严重降低分割的速度,并且有可能漏掉一些分割点。因为该方法是一种依赖于局部极值的方法,窗口较大时就会有多个局部极值落在该窗口内,而根据算法策略只会选取其中一个使得目标函数最小的极值,其他的则会被忽略。在核 k-means、谱聚类以及他们之前工作^[26]的基础上,Zhou 等^[27]针对人体运动数据提出了一种叫分层对齐聚类分析(Hierarchical aligned cluster analysis, HACA)的时序聚类方法。该方法可以首先将运动数据分成多个运动片段,每个片段属于其中一个聚类,然后在一个更低的层次再将各个运动片段拆分成更小的运动单元以实现更细腻的分割。该方法需要计算一个核矩阵,矩阵的规模由待分割运动序列的长度决定,所以 HACA 在分割之前使用了一个时序约减算法以消除时序上的冗余并减少核矩阵的规模。但该时序约减算法过分依赖于数据的无噪性,即使数据只含有少量噪声,该时序约减算法就无法有效地消除时序冗余,导致核矩阵的计算过于耗时,也增加了分割的计算量。Elhamifar 等^[28]构建了一个源数据集和目标数据集之间的不相似性矩阵,根据这个矩阵用目标数据集中的样本表示源数据集的样本,然后根据这些表示选择源数据集中的部分样本作为其中一个类别,从而达到聚类的目的。Elhamifar 等将此方法用于运动捕获数据的分割作为验证。该方法的可扩展性很强,源数据集与目标数据集不需要是同一类型的数据,只需要构建它们之间的不相似性即可,而且算法的框架也不仅限于聚类,也可适用其他类型的应用,如变量选择等。但针对于运动数据的分割,该方法缺少时序聚类的控制,导致分割的结果缺乏时序连续性。

上述的多种方法在一定条件下都能取得不错的分割结果,但它们并没有考虑到运动捕获数据的一些特性,例如黎曼流形结构,也没有利用运动数据关节与关节之间的相关性,而只是把一帧看成一个整体来处理。并且,目前尚没有一种鲁棒的运动捕获数据的分割方法来处理运动捕获过程中产生的多种噪声,所以这些都将是运动分割领域新的研究方向。

3 运动压缩

高效实时的人体运动捕获数据压缩方法能够有效地降低运动数据的存储空间,提高网络传输速度,为很多基于运动捕获数据的在线应用提供更好的用户体验。人体运动数据是一种序列的空间数据,由时域和空域两个部分组成,所以运动数据时空域的冗余是运动压缩可行性的基础。此外,相似运动片段间的信息冗余也可以利用进行运动压缩的重要因素。下面本文将从消除冗余的角度^[29]对运动压缩及其所应用的技术进行介绍。

3.1 基于消除时空域数据冗余的方法

此类方法又可以细分为基于曲线拟合的方法、基于数据降维的方法、基于信号处理的方法、基于四元数空间稀疏表达的方法和基于关键帧提取的方法。其中基于关键帧提取的方法将在下一节做详细介绍。

(1) 基于曲线拟合的方法。Khan 等^[30]将人体运动位移及旋转数据作为 N 维欧氏空间的曲线,采

用二次 Bézier 曲线最小二乘进行拟合。Arikan 等^[31]则首先将运动数据进行分割,然后采用等间隔分割策略利用三次 Bézier 曲线对分割后的片段进行拟合。Han 等^[32]提出利用多项式曲线拟合对人体运动数据进行精简表示,从而使得人体运动数据可存储且应用在功耗要求较苛刻的移动设备上。

(2) 基于数据降维的方法。Liu 等^[33]首先基于运动分割方法对运动进行分割,将运动数据分割成短小的片段。然后利用 PCA 对其进行降维后,将每个运动片段的几个关键帧投影到 PCA 得到的少数几个主元上。在该方法中,提取关键帧消除了时域冗余,PCA 降维消除了空间冗余。文献^[34]提出分析人体运动各局部节点误差与目标误差矩阵的关系,引入拉格朗日乘子来平衡压缩比与误差,然后以一段运动序列为单位进行 PCA 的降维,最后采用熵编码对降维后的数据进行编码压缩。主测地线分析(Principal geodesic analysis,PGA)是 PCA 在黎曼流形上的扩展,Tournier 等^[35]运用 PGA 对人体运动进行降维,并利用逆向运动学(Inverse kinematics,IK)对 PGA 的参数进行匹配和优化,最后只需要保留末端效应器及根节点轨迹的相关参数,就能有效地对数据进行还原。

(3) 基于信号处理的方法。基于信号处理的方法,如小波技术,通常将各自由度作为相互独立的信号进行处理。Beaudoin 等^[36]定义了一个变形矩阵来度量压缩前后的误差变化,然后针对每个自由度,选择使得变形矩阵最小的优化小波系数。通过这种优化,能很大程度上提高压缩质量,并在解码速度上优于 Arikan 等^[31]的方法。Li 等^[37]舍弃了一些低频分量对应的小波系数,在对数据进行压缩的同时,并未造成很大的重构误差。针对文献^[36]方法计算复杂的缺点,Lee 等^[38]在其之前工作^[39-40]的基础上设计了一个基于多分辨率小波的压缩方法,根据自适应的误差估计矩阵来平衡需保留的小波系数与视觉效果,以达到最优的压缩误差比。Chattopadhyay 等^[41]提出了一个在移动设备上压缩和传输人体动画参数数据的算法——BAP-Indexing。该算法利用人体骨骼的层次结构智能地索引运动数据,根据运动矩阵的浮点数分布情况将数据放入不同的桶中,使得每个浮点数都可用 1 个索引来表示,从而有效地压缩了数据,该算法利用 3 个预先定义的质量控制参数就能有效地控制压缩质量与压缩比的平衡。

(4) 基于四元数空间稀疏表达的方法。Zhu 等^[42]提出在四元数空间中直接进行人体运动的稀疏分解的方法(Quaternion space sparse decomposition, QSSD),它将人体的旋转数据分解成字典和稀疏表示系数 2 个部分。这样,欧拉空间中的线性组合及权重操作就分别转变成了四元数空间的乘及幂操作。在压缩之前,首先提取参考姿态、根节点信息及所有节点的旋转信息,并利用 Tournier 等^[35]提出的方法对根节点及旋转信息进行压缩,然后将压缩后的节点旋转信息利用 QSSD 方法进行分解,最后利用算术编码进行编码压缩。

3.2 基于消除片段域数据冗余的方法

运动数据中存储着大量相似的运动片段,而这些类型相同的运动片段之间存在着大量的冗余信息。以走路运动为例,在一个标准的走路运动的基础上,可以衍生出许多不同风格的走路运动,比如高兴地走、垂头丧气的走和蹒跚地走等等。如果可以有效地提取不同风格走路运动与标准走路运动之间的关系,那么只需要存储标准的走路运动,再根据提取的关系重构出不同风格的走路运动,此举可以有效地减少对冗余信息的存储,达到运动压缩的目的。Lin 等^[43]定义了主片段、重复片段和唯一片段的概念。主片段是运动库中具有代表性的一些动作,重复片段是与此类代表性动作相似的重复动作,而唯一性片段是指没有其他相似片段的孤本。该方法首先进行运动分割,然后对这些片段进行聚类。聚类后,用 PCA 对每个运动片段进行降维。随后,各聚类中的运动被连接起来,并运用 Kovar 等^[44]的运动检索方法进行重复运动片段的分析,就可得到相似姿态路径,对这些相似路径进行分析、切分,就能得到主片段以及与其相似的重复片段。这样,在每个聚类中,只需要存储主片段的原始值以及重复片段与主片段的差值就能还原数据,但对唯一片段需要单独保存。Lin 等^[43]算法集成了多种优秀的运动分析及编码算法,取得了很高的压缩比,但由于该算法依赖 Match web^[44]进行运动片段的分析提取,其较高的时间复杂度使得算法难以扩展到大型运动数据库。

Hou 等^[45]将人体运动切分成统一长度的片段集,利用运动片段之间的关联性,采用低秩逼近的方

法将这些片段投射到一对正交的矩阵上,从而抛弃大部分冗余的元素。

4 关键帧提取

关键帧提取在存储、传输和浏览的过程中都扮演着很重要的角色,所以在运动数据相关的研究领域一直都是一个热门的研究课题。关键帧提取本质上就是一个时序重采样问题,所以最简单的关键帧提取方法就是均匀采样,即间隔相同的时间段提取一个运动帧作为关键帧。但问题是运动的节奏并不固定,所以当运动过快或者过慢时就会出现过采样或者欠采样问题。过采样导致信息冗余,欠采样导致信息丢失。所以基于内容的自适应关键帧提取技术得到了广泛的应用。现有的关键帧提取技术按照实现方式可以分为以下 5 类。

(1) 基于曲线简化的关键帧提取。曲线简化^[46-51]是一种常用的关键帧提取技术。一个运动帧可以看成高维姿态空间中的一个点,由于运动的时空连续性,连续的运动帧就构成了运动曲线,所以关键帧提取就转化成了如何在曲线上找到一些能最好地描述该曲线的点,其基本原理如图 5 所示。Halit 等^[50]使用多尺度高斯滤波器识别运动曲线中特别突出的帧,并使用 centre-surround 算子^[52-53]计算每一帧的显著性,最后根据显著性曲线的局部极值来提取关键帧。Wei 等^[51]等提出了一种基于最小平方距离曲线的关键帧提取算法。该方法根据每个运动帧与同一个标准帧的距离,将一个运动序列表示成一个距离曲线,然后通过局部极值搜索的方式提取关键帧。此类方法通常需要人工指定关键帧的数目,但实际上关键帧的数量很难人为预测。此外,曲线简化只利用了运动序列的局部信息,而忽略了全局信息。

(2) 基于聚类的关键帧提取。此类方法^[54-57]的策略很简单,即使用诸如 k-means 之类的聚类方法将一个序列中相似的帧聚成一个类,然后选取该类中最具代表性的一帧作为关键帧。但它们忽略了一个重要的问题——运动的周期性。时序信息也是人体运动的一个重要组成部分,但显然基于聚类的关键帧提取方法基本上都没有考虑并利用运动的时序信息。当一个运动序列中存在周期性运动时,比如走路运动,那么多个走路周期就会被聚在一个类中,并且只提取一组关键帧,所以并不能从提取的关键帧中了解整个运动的概况,比如运动的长度。

(3) 基于矩阵分解的关键帧提取。一些研究人员提出将一个运动序列可以表示成矩阵,然后通过奇异值分解^[58]或低阶离散余弦变化^[59]之类的技术将该矩阵分解成一个关键帧矩阵和一个权重矩阵。Huang 等^[60]将关键帧提取转化成了一个带约束的矩阵分解问题,并通过最小二乘优化高效地解决了这个问题。此类方法通常以最小重构误差为优化目标,所以可以保证提取关键帧的信息完整性。但在使用的过程中通常需要针对不同的输入设定不同的阈值,这为此类方法的推广带来困难。

(4) 基于优化的关键帧提取。近年来,出现了一些基于遗传算法^[61-62]的关键帧提取方法。遗传算法很适合复杂的离散优化问题,它能在不需要任何初始化的情况下找到全局最优解。此类方法通常将重构误差作为适应度函数,用 0-1 序列表示一个运动序列中每帧是否被选为关键帧,可以无障碍地进行变异、交叉等操作。但其缺陷也很明显,就是过慢的收敛速度导致了过大时间开销,严重制约了此类方法的实际应用。即使 Liu 等^[61]试图通过概率单纯形算法加速遗传算法的收敛速度,但效果依然不明显。

(5) 基于稀疏表示的关键帧提取。利用人体运动的稀疏特性,Xia 等^[63]提出采用基于稀疏表示的框架进行关键帧提取。该方法首先利用测地线指数核^[64]将运动数据投影到高维的希尔伯特空间中。测

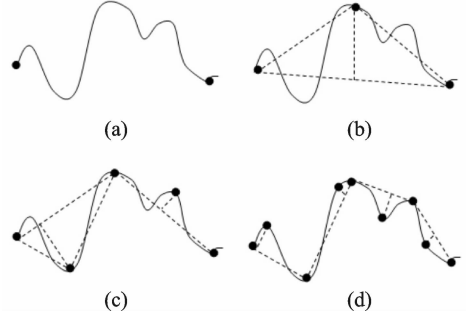


图 5 文献[46]中曲线简化的基本步骤
Fig. 5 Main steps of curve simplification^[46]

地线指数核由测地线距离演化而来,而相比于欧式距离,测点线距离能够更好的描述球面上两点之间的距离,如图 6(a)所示,所以测地线指数核的引入能够有效地解决运动数据的流形结构问题。然后在希尔伯特空间将一个运动序列表示成一个矩阵,矩阵的每一列就是一个运动帧,以该矩阵作为字典对其自身进行表示,通过 group lasso^[65]使得表示系数为“行稀疏”,那么在重构系数矩阵中,稀疏行对应的运动帧并没有用来表示运动序列中的任何帧,所以不稀疏的行所对应的帧则为关键帧。除此之外,该方法还通过三角约束阻断了相距非常远的运动帧之间的联系,有效地解决了周期性运动关键帧分布不合理的问题。其目标函数为

$$\min_{\mathbf{W}} \sum_{i=1}^d \|\phi(\mathbf{X}_i) - \phi(\mathbf{X}_i)\mathbf{W}_i\|_{\mathbb{F}}^2 + \lambda \|\mathbf{W}\|_{p,2} \quad (5)$$

式中: $\mathbf{W}=[{}^w\mathbf{W}_1, {}^w\mathbf{W}_2, \dots, {}^w\mathbf{W}_d]$ \mathbf{W}_i 为第*i*关节带有三角约束的重构系数; \mathbf{X}_i 为在一个关节中所有帧的第*i*个关节的转向; $\phi(\cdot)$ 为基于测地线指数核的隐式映射; λ 为一个调谐参数。式(5)以及关键帧提取的过程可以用图 6(b)形象地表示。

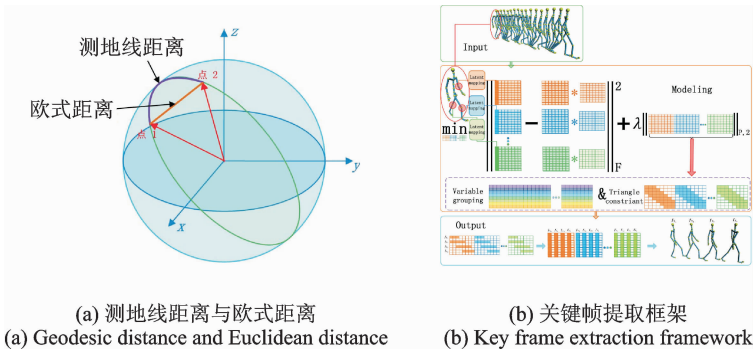


图 6 测地线距离与欧式距离示意图^[63]和文献[63]的关键帧提取框架

Fig. 6 Geodesic distance and Euclidean distance^[63] and the key frame extraction framework in Ref. [63]

5 运动检索

对运动数据的浏览和检索是运动数据使用过程中一个不可缺少的环节。随着数据量的不断增加,为用户提供一个高效快捷的浏览和检索方式对运动捕获以及计算机动画产业来说非常重要。与单个运动帧不同,一段运动捕获数据是一种随着时间变化的高维矢量,这大大增加了处理的难度,因此一些简单的检索方法往往难以奏效,且效率低下。针对这一问题,研究人员近年做了大量的研究,开发了许多杰出的运动检索方法^[42,44,66-69],这也使得基于内容的运动检索成为一个重要的研究课题。解决运动检索问题的关键在于设计一种合适的相似性度量准则。早期的方法通常将一个运动片段看成高维向量,但由于运动的长度不等,所以导致样本的维度不一致。为了进行统一的相似性度量,研究人员通常使用时间动态弯曲算法(Dynamic time warping, DTW),根据帧与帧之间的距离,建立运动间的整体时间对应关系,从而达到统一运动长度的目的。然后再根据提取的特征,计算运动片段之间的相似度,选取相似度最高的运动片段作为检索结果。但 DTW 算法的复杂度高,占用存储空间大,不适合大规模数据库的检索。Kovar 等^[44]在 DTW 的基础上提出了一种更高效的检索方法。该方法先按给定阈值检索出相似运动,再将检索出的相似运动作为新样本继续检索,经过多轮迭代得到最终符合用户需求的结果。Jeff 等^[70]提出了一种自适应的特征提取方法可以有效地获取与待检索样本密切相关特征,该方法能够根据检索的差异性自适应地选取特征子集,但为了满足实时性的要求,需要对运动片段进行编码和对检索输入进行人工编辑。

将运动片段作为高维向量进行处理需要大量的计算开销,所以近年来,不少研究人员将人体运动转换为离散表示,从而有效地避免了庞大的计算开销。这种做法不仅能有效地节约时间成本,还能使得检索的过程和结果更接近于人们的直观理解,这是因为包括字符串^[71]、状态^[70]和文本文档^[42]在内的离散表示具有更高的语义层次,使得计算运动间的相似度时能够获得逻辑上正确的结果。Muller 等^[72]在原数值特征的基础上,计算除了实际中具有几何意义的特征,如双手间的距离、膝关节的瞬时速度和大小臂间的夹角等,再通过设定阈值,将它们离散化成二值向量,这种离散表示为提高了特征的语义层次和检索出逻辑相似的运动奠定了良好的基础。虽然对运动的离散化处理导致部分细节丢失,但特征的高层语义有助于提高检索结果的逻辑正确性,而大部分的检索需求都是逻辑上的正确性,而并非运动细节的绝对相似。Zhu 等^[42]使用隐狄利克雷分配(Latent dirichlet allocation, LDA)来挖掘人体运动中的主题,并将运动表示成这些主题上的概率分布,该方法选用一些特征来作为运动词汇表,并将人体运动用几何特征描述后转化为“词”,在忽略词序的前提下,用“词”的切换频率和运动幅度近似替代一个运动文本中该“词”出现的频率。相比之前的方法,主题特征在语义层次上有了较大的提升。但随后, Lan 等^[73]指出, Zhu 等^[42]的方法并没有合理地建立文本的概念,运动字典的扩展性不够,在需要表示新的运动时还需要重新构建字典,并且,以几何特征为词汇的运动文档无法可视化,不能给人一种直观的感受。所以 Lan 等^[73]提出了一种新的通用运动表示方式,如图 7 所示。他们将所有运动的关键帧进行聚类,从每一个类别中选取一个运动帧作为“词”,如图 7 中的第 1 行,然后统计每一个运动中这些“词”出现的频率,构建运动数据的类文本表示方式,最后通过计算这类文本表示间的相似度给出检索结果。该方法的主要贡献在于提出了一种自动的可扩展的运动表示方式,并且可对运动字典进行可视化,进一步提升了运动表示的语义层次。2015 年, Wang 等^[69]则提出了一种基于语义缩略图的运动检索方法。该方法首先对运动数据进行聚类 and 特征匹配,然后对用于训练的待检索运动进行检索,更新数据库中每个运动被正确检索的频率。在实际应用时,根据数据库中每个运动被正确检索的频率,对最新得出的检索结果进行排序以提高检索的正确率。最后使用一个基于可视化数据分析的关键帧提取算法为每个运动制作语义缩略图。

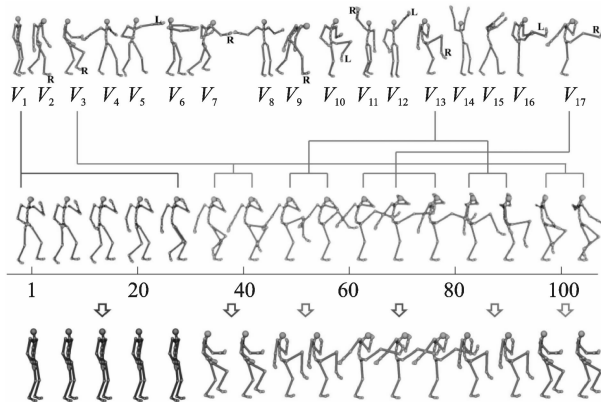


图 7 运动词汇表及文档的示意图^[73]

Fig. 7 Illustration of motion vocabulary and document^[73]

6 运动合成

运动合成是运动数据重用的重点和关键技术,也是运动重用过程中最困难的环节。运动捕获数据维度高、信息量大且结构复杂,具有时空连续性以及黎曼流形结构,这些都给运动合成带来了挑战。合成后的运动数据最终要用来被人观赏,而人眼又十分善于发现运动中不协调的地方,这又对运动合成提

出了更高层次的要求,所以一直以来运动合成都是一项十分热门的研究课题,随着研究的深入也出现了一些优秀的运动合成算法。其中,数据驱动的运动合成方法能够保持运动数据的时空特性,因而得到了广泛的研究和应用。

(1)运动混合。运动混合^[74-77]是一种简洁高效的运动合成模型。此类方法首先对同类型的运动片段进行预处理,包括利用 DTW 算法对其进行时序对齐,然后通过线性变换使得每个运动帧都有相似的空间坐标,即坐标对齐。经过时序对齐和坐标对齐的运动在结构上得到了统一,对这些结构统一的运动进行有权重的插值,再对插值后的运动进行约束重建就能得到十分逼真的新运动。但此类方法的数据组织方式过于简单,无法挖掘出数据中的内在规律,且用户无法与系统进行实时交互,很难控制运动合成的结果使其满足用户的需求。

(2)基于图搜索和运动过渡的运动合成。此类方法^[78-80]将运动图作为一种表示运动数据之间关系的数据结构,节点表示运动帧,边表示不同运动片段之间的运动过渡。根据用户设定的不同阶段的运动状态,在运动图中搜索最优路径,并合成出符合要求的运动。其优势在于对建图运动数据的要求比较宽泛,并且能产生长度不同、过渡自然的运动,但合成运动的表达能力依赖于原运动数据,且容易造成根节点的位置计算不准确,导致脚步的滑动和朝向抖动。

(3)参数化的运动合成。参数化的运动模型^[44,81-82]能够利用运动的一些物理特性有效地解决运动图存在的问题。Kwon 等^[82]将运动的类型、速度、加速度和落脚点等以参数的形式引入合成模型,在合成的过程中进行控制,可以解决一些诸如脚步滑动、朝向抖动之类的问题。Heck 等^[81]将运动图的节点构造为连续参数空间,为原本非常有限的拼接组合方式带来了细粒度的控制,例如该方法可以通过对参数的调节合成出更为丰富和细腻的出拳运动,如图 8(a)所示。这些方法很大程度上提高了运动合成过程的可控性,但这些物理参数的语义层次太低,内容需要预先人为指定,无法自动适应运动类型的变化。

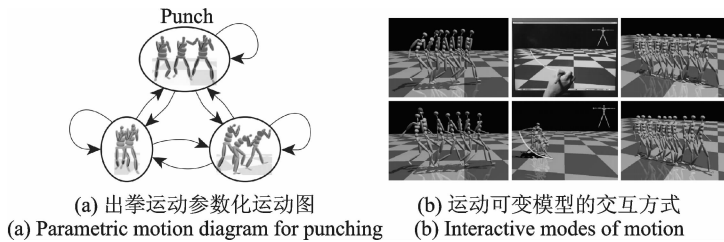


图 8 出拳运动参数化的运动图^[81]和运动可变模型的交互方式^[83]

Fig. 8 Parametric motion diagram for punching^[81] and interactive modes of motion^[83]

针对上述问题,一些研究人员^[83-90]根据运动各维度之间高度相关这一特性,提出使用降维方法自动获取运动的内在自由度,并将其作为参数参与运动合成。在降维后的空间对其直接修改后,重构回原始空间以合成新的运动。例如王宇杰等^[85]将原始运动投影到低维流形上,然后学习低维样本点与原始运动的对应关系,形成低维参数空间到原始空间的逆映射,就能够通过修改低维变量完成原始空间的运动合成。Min 等^[83]利用 PCA 对运动数据进行降维后,获取时域和空域两类参数,然后根据用户输出的合成条件,优化调整这两类参数,重构回原始空间,达到交互式合成的目的,且交互方式并不单一(见图 8(b))。为了尽可能多地保留原始数据的信息或是能够在低维(两到三维)空间显示降维后样本的位置,此类方法通常使用诸如 PCA 和多元尺度(Multi-dimensional scaling, MDS)之类的降维方法,这使得获取的低维参数与所有的人体关节都有关,所以这类方法获取的参数同样缺乏语义含义,不易被用户理解和使用。

因此蓝荣祎等^[91]利用稀疏主成分分析对同一类型的一组运动进行降维,使得降维后的主成分只依赖于少数几个人体关节,从而有效地为低维参数赋予了语义含义。但用此方法进行降维后,某些关节出

现在在多个低维参数里,这就使得调整不同的参数时,会对某些关节造成交叉影响。为解决此问题,夏贵羽等^[92]提出了模板化的人体运动合成模型。该模型进一步强化了低维运动参数语义含义,并通过 Exclusive group lasso^[65] 有效地抑制了运动参数之间交叉影响的问题,其目标函数为

$$(A, B) = \arg \min_{A, B} \sum_{n=1}^N \|x_n - x_n B A^T\|^2 + \lambda \sum_{j=1}^k \|\beta_{1,j}\|^2 + \sum_{j=1}^k \lambda_{1,j} \|\beta_{1,jk^2}\|_{1,2} + \sum_{i=1}^s \lambda_{2,i} \|\beta_{2,ig^2}\|_{1,2} \quad (6)$$

式中: x_n 为对齐后的第 n 个运动; $A = [\alpha_1, \dots, \alpha_K] \in R^{d \times K}$, 且满足 $A^T A = I_{K \times K}$, $B = [\beta_{1,1}, \dots, \beta_{1,K}] \in R^{d \times K}$, $\beta_{1,j}$ 是 B 的列向量,也是待求的载荷向量,对其进行 L_1/L_2 惩罚以产生“成组稀疏”的效果,使其只依赖于很少的关节,从而获取该运动的一个自由度,它所对应的低维参数则作为该运动模板的一个接口。 $\beta_{2,i}$ 为元素重排后的载荷矩阵(记为 $B' = [\beta_{2,1}, \dots, \beta_{2,s}]^T \in R^{S \times TK}$) 的行向量,其中 S 表示关节数, $TK = T \times K$ 表示帧数与参数个数的乘积,对其进行 Exclusive group lasso 惩罚,本质上是让所有运动参数“竞争”获取每一个关节的“控制权”, g^1 和 g^2 分别对应 Group lasso 和 Exclusive group lasso 两种不同的分组方式。该方法可以将每一种类型的运动制作成运动模板,用户只要控制几个具有语义含义的运动参数就能利用运动模板完成运动合成,大大降低了运动合成技术的使用门槛,对运动捕获数据的使用以及动画合成具有积极的推动作用。图 9(a)是踢腿运动模板的示意图。此方法原理上是一种局部的运动混合方法,稀疏主成分分析能够根据运动的自由度将同类型的运动拆分成几个部分,而每个运动参数也就对应着该部分进行混合的权重。因此该方法可以通过局部控制,合成出远比运动混合更丰富的运动,而且,该方法通过一个多元高斯混合模型有效地解决了运动混合方法会出现的“畸形”运动的问题。

(4) 基于深度学习的运动合成。自适应提取参数的运动合成方法存在缺陷,即需要将训练数据进行统一处理,这其中包含了一系列操作,如运动分割、运动分类、时序对齐和坐标对齐,处理过程复杂且耗时。并且,其合成后的运动有可能会出现脚步滑动朝向抖动等问题。随着深度学习理论的快速发展, Holden 等^[93] 提出了一个基于深度学习框架的运动合成方法,该方法对训练数据的格式要求比较宽泛,不需要上述的各种操作,任何类型和长短的运动捕获数据都可以参与训练。其网络结构和方法框架如图 9(b)所示,经过此框架学习得到的运动流形由一个卷积自编码器的隐藏单元表示,能够根据用户提供的高层参数,合成出多种类型的复杂运动,并能通过对隐单元空间的约束解决脚步滑动和朝向抖动等问题。

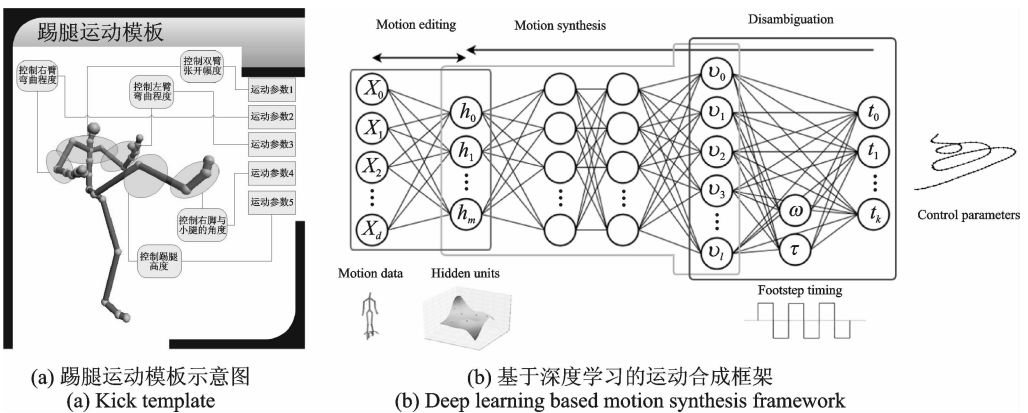


图 9 踢腿运动模板示意图^[92]和基于深度学习的运动合成框架^[93]
 Fig. 9 Kick template^[92] and deep learning based motion synthesis framework^[93]

7 结束语

随着以运动数据驱动的多部动画电影的热映,运动捕获技术得到了空前的关注,但大部分注意力还是集中在运动的捕获和数据的使用上。虽然运动捕获数据相关的研究已经持续了很多年,但早期的研究主要集中于图形学领域,仅采用一些运动学和统计学技术处理运动数据,针对性不够。近年来,机器

学习研究的热潮也覆盖到了运动捕获数据领域。机器学习技术能够有效地挖掘运动捕获数据中蕴藏规律,但大多数方法只是将通用技术简单地应用到了运动数据上,这就丢弃了早期研究积累的经验,导致对运动数据的研究不够深入,所以未来需要更多在机器学习框架下针对运动数据本身特性的、结合运动学和统计学优势的定制型方法的出现。

参考文献:

- [1] Devanne M, Wannous H, Berretti S, et al. 3-D human action recognition by shape analysis of motion trajectories on Riemannian manifold[J]. *IEEE Transactions on Systems Man & Cybernetics*, 2014, 45(7):1023-1029.
- [2] Hauberg S, Pedersen K S. Spatial measures between human poses for classification and understanding[M]// *Articulated Motion and Deformable Objects*. Berlin, Heidelberg:Springer, 2012:26-36.
- [3] Tuzel O, Porikli F, Meer P. Pedestrian detection via classification on Riemannian manifolds[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2008, 30(10):1713-1727.
- [4] Mairal J, Bach F, Ponce J, et al. Online dictionary learning for sparse coding[C]// *International Conference on Machine Learning*. Montreal, Quebec, Canada:[s. n.], 2009:689-696.
- [5] Schölkopf B, Platt J, Hofmann T. Efficient sparse coding algorithms[C]// *Advances in Neural Information Processing Systems 19, Proceedings of the Twentieth Conference on Neural Information Processing Systems*. Vancouver, British Columbia, Canada: DBLP, 2006:801-808.
- [6] Uller-Ulhaas K D. Robust optical user motion tracking using a Kalman filter[C]//*Symposium on Virtual Reality Software & Technology*. [S. l.]:ACM,2003:45-48.
- [7] Welch G, Vicci L, Keller K, et al. The HiBall Tracker: High-performance wide-area tracking for virtual and augmented environments[C]// *ACM Symposium on Virtual Reality Software and Technology*. [S. l.]: DBLP, 1999:1-10.
- [8] Guo S, Roberg J. A high-level control mechanism for human locomotion based on parametric frame space interpolation[C]// *Eurographics Workshop on Computer Animation and Simulation*. New York: Springer-Verlag, 1996:95-107.
- [9] Liu G, Mcmillan L. Estimation of missing markers in human motion capture[J]. *The Visual Computer*, 2006, 22(9): 721-728.
- [10] Rose C, Cohen M F, Bodenheimer B. Verbs and adverbs: Multidimensional motion interpolation[J]. *IEEE Computer Graphics & Applications*, 1999, 18(5):32-40.
- [11] Hui L, Chai J. Example-based human motion denoising[J]. *IEEE Transactions on Visualization & Computer Graphics*, 2010, 16(5):870-879.
- [12] Xiao J, Feng Y, Hu W. Predicting missing markers in human motion capture using l1-sparse representation[J]. *Computer Animation and Virtual Worlds*, 2011, 22(2/3):221-228.
- [13] Hou J, Chau L P, He Y, et al. Human motion capture data recovery via trajectory-based sparse representation[C]// *IEEE International Conference on Image Processing*. Melbourne, Australia: IEEE, 2013:709-713.
- [14] Xiao J, Feng Y, Ji M, et al. Sparse motion bases selection for human motion denoising[J]. *Signal Processing*, 2014, 110: 108-122.
- [15] Xia G, Sun H, Zhang G, et al. Human motion recovery jointly utilizing statistical and kinematic information[J]. *Information Sciences*, 2016, 339:189-205.
- [16] Feng Y, Xiao J, Zhuang Y, et al. Exploiting temporal stability and low-rank structure for motion capture data refinement [J]. *Information Sciences*, 2014, 277(2):777-793.
- [17] Lai R Y Q, Yuen P C, Lee K K W. Motion capture data completion and denoising by singular value thresholding[C]// *Proc Eurographics Association*. Llandudno, UK; [s. n.], 2011:1-4.
- [18] Liu X, Cheung Y M, Peng S J, et al. Automatic motion capture data denoising via filtered subspace clustering and low rank matrix approximation[J]. *Signal Processing*, 2014, 105(12):350-362.
- [19] Tan C H, Hou J, Chau L P. Human motion capture data recovery using trajectory-based matrix completion[J]. *Electronics Letters*, 2013, 49(12):752-754.
- [20] Candès E, Recht B. Exact matrix completion via convex optimization[J]. *Foundations of Computational Mathematics*, 2009, 9(6):717-772.
- [21] Ller M, Der T, Clausen M. Efficient content-based retrieval of motion capture data[J]. *ACM Transactions on Graphics*, 2005, 24(2):677-685.
- [22] Kwon T, Shin S Y. Motion modeling for on-line locomotion synthesis[C]// *ACM Siggraph/Eurographics Symposium on Computer Animation (SCA)*. Los Angeles, CA, USA: ACM, 2005:29-38.
- [23] Becher J C, Keeling J W, McIntosh N, et al. Segmenting motion capture data into distinct behaviors[J]. *Proceedings-Graphics Interface*, 2004, 2004(5):185-194.

- [24] Gong D, Medioni G, Zhao X. Structured time series analysis for human action segmentation and recognition[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2014, 36(7):1414-1427.
- [25] Gong D, Medioni G, Zhu S, et al. Kernelized temporal cut for online temporal segmentation and recognition[J]. *Lecture Notes in Computer Science*, 2012, 7574(1):229-243.
- [26] Zhou F, Torre F, Hodgins J K. Aligned cluster analysis for temporal segmentation of human motion[C]// *IEEE International Conference on Automatic Face & Gesture Recognition*. Amsterdam, Netherlands: IEEE, 2008:1083-1084.
- [27] Zhou F, Torre F D L, Hodgins J K. Hierarchical aligned cluster analysis for temporal clustering of human motion[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2013, 35(3):582-96.
- [28] Elhamifar E, Sapiro G, Sastry S. Dissimilarity-based sparse subset selection [J]. *IEEE Transactions on Software Engineering*, 2014, 38(11):1.
- [29] 王鹏杰, 潘志庚, 李威. 人体运动捕获数据压缩技术研究进展[J]. *计算机辅助设计与图形学学报*, 2016, 28(7):1037-1046.
- Wang Pengjie, Pan Zhigeng, Li Wei. Recent advances in compression of human motion capture data[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2016, 28(7):1037-1046.
- [30] Khan M A. An efficient algorithm for compression of motion capture signal using multidimensional quadratic Bézier curve break-and-fit method[J]. *Multidimensional Systems and Signal Processing*, 2016, 27(1):1-23.
- [31] Arikan O. Compression of motion capture databases[J]. *ACM Transactions on Graphics*, 2006, 25(3):890-897.
- [32] Han Y. Computer animation in mobile phones using a motion capture database compressed by polynomial curve-fitting techniques[J]. *IEEE Transactions on Consumer Electronics*, 2008, 54(3):1008-1016.
- [33] Liu G, Mcmillan L. Segment-based human motion compression[C]// *ACM Siggraph/Eurographics Symposium on Computer Animation(SCA)*. Vienna, Austria: ACM, 2006:127-135.
- [34] Váša L, Brunnett G. Rate-distortion optimized compression of motion capture data[J]. *Computer Graphics Forum*, 2014, 33(2):283-292.
- [35] Tournier M, Wu X, Courty N, et al. Motion compression using principal geodesics analysis[J]. *Computer Graphics Forum*, 2009, 28(2):355-364.
- [36] Beaudoin P, Poulin P, Panne M V D. Adapting wavelet compression to human motion capture clips[C]// *Graphics Interface 2007 Conference*. Montreal, Canada: DBLP, 2007:313-318.
- [37] Li S, Okuda M, Takahashi S I. Compression of human motion animation using the reduction of interjoint correlation[J]. *Journal on Image & Video Processing*, 2008, 2008(2):1-15.
- [38] Lee C H, Lasenby J. An efficient wavelet-based framework for articulated human motion compression[C]// *International Symposium on Advances in Visual Computing*. Las Vegas, NV, USA: Springer-Verlag, 2008:75-86.
- [39] Lee C H, Lasenby J. 3D human motion compression using wavelet decomposition[C]// *ACM Siggraph Research Posters*. [S. l.]: ACM, 2006:104.
- [40] Lee C, Lasenby J. A compact representation for articulated human motion[C]// *ACM Siggraph 2007 Posters*. San Diego, California, USA: ACM, 2007:96.
- [41] Chattopadhyay S, Bhandarkar S M, Li K. Human motion capture data compression by model-based indexing: A power aware approach[J]. *IEEE Transactions on Visualization & Computer Graphics*, 2007, 13(1):5-14.
- [42] Zhu M, Sun H, Lan R, et al. Human motion retrieval using topic model[J]. *Computer Animation & Virtual Worlds*, 2012, 23(5):469-476.
- [43] Lin I C, Peng J Y, Lin C C, et al. Adaptive motion data representation with repeated motion analysis[J]. *IEEE Transactions on Visualization & Computer Graphics*, 2011, 17(4):527-538.
- [44] Kovar L, Gleicher M. Automated extraction and parameterization of motions in large data sets[J]. *ACM Transactions on Graphics*, 2004, 23(3):559-568.
- [45] Hou J, Chau L P, He Y, et al. Low-rank based compact representation of motion capture data[C]// *IEEE International Conference on Image Processing*. Paris: IEEE, 2014:1480-1484.
- [46] Lim I S, Thalmann D. Key-posture extraction out of human motion data [C]// *Proceedings of the 23rd Annual International Conference of Engineering in Medicine and Biology Society*. [S. l.]: IEEE, 2001:1167-1169.
- [47] Togawa H, Okuda M. Position-based key-frame selection for human motion animation[C]// *International Conference on Parallel and Distributed Systems Proceedings*. [S. l.]: IEEE, 2005:182-185.
- [48] Li N S, Okuda M, Takahashi S. Embedded key-frame extraction for cg animation by frame decimation[C]// *2005 IEEE International Conference on Multimedia and Expo*. Amsterdam, Netherlands: IEEE, 2005:1404-1407.
- [49] Xiao J, Zhuang Y, Yang T, et al. An efficient key-frame extraction from motion capture data[M]// *Advances in Computer Graphics*. Berlin, Heidelberg: Springer, 2006:494-501.
- [50] Halit C, Capin T. Multiscale motion saliency for key-frame extraction from motion capture sequences[J]. *Computer*

Animation and Virtual Worlds, 2011, 22(1):3-14.

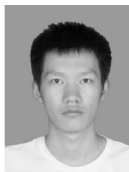
- [51] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1998, 20(11):1254-1259.
- [52] Lee C H, Varshney A, Jacobs D W. Mesh saliency[J]. ACM Transactions on Graphics, 2005, 24(3):659-666.
- [53] Wei X, Liu R, Zhang Q. Key-frame extraction of human motion capture data based on least-square distance curve[J]. Journal of Convergence Information Technology, 2012, 7(12):11-19.
- [54] Min J P, Shin S Y. Example-based motion cloning[J]. Computer Animation and Virtual Worlds, 2004, 15(3/4):245-257.
- [55] 朱登明, 王兆其. 基于运动序列分割的运动捕获数据关键帧提取[J]. 计算机辅助设计与图形学学报, 2008, 20(6):787-792.
- Zhu Dengming, Wang Zhaoqi. Extraction of keyframe from motion capture data based on motion sequence segmentation[J]. Journal of Computer-Aided Design & Computer Graphics, 2008, 20(6):787-792.
- [56] Qiang Z, Yu S P, Zhou D S, et al. An efficient method of key-frame extraction based on a cluster algorithm[J]. Journal of Human Kinetics, 2013, 39(1):5-13.
- [57] Liu F, Zhuang Y, Wu F, et al. 3D motion retrieval with motion index tree[J]. Computer Vision & Image Understanding, 2003, 92(2):265-284.
- [58] Gong Y, Liu X. Video summarization using singular value decomposition[C]// IEEE Conference on Computer Vision & Pattern Recognition. Madison, Wisconsin; IEEE, 2003:157-168.
- [59] Cooper M, Foote J. Summarizing video using non-negative similarity matrix factorization[C]// IEEE Workshop on Multimedia Signal Processing. [S. l.]; IEEE, 2002:25-28.
- [60] Huang K S, Chang C F, Hsu Y Y, et al. Key probe: A technique for animation key-frame extraction[J]. The Visual Computer, 2005, 21(8):532-541.
- [61] Liu X M, Hao A M, Zhao D. Optimization-based key frame extraction for motion capture animation[J]. The Visual Computer, 2013, 29(1):85-95.
- [62] Lee T Y, Lin C H, Wang Y S, et al. Animation key-frame extraction and simplification using deformation analysis[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2008, 18(4):478-486.
- [63] Xia G, Sun H, Niu X, et al. Key-frame extraction for human motion capture data based on joint kernel sparse representation [J]. IEEE Transactions on Industrial Electronics, 2017, 64(2):1589-1599.
- [64] Jayasumana S, Hartley R, Salzmann M, et al. Combining multiple manifold-valued descriptors for improved object recognition[C]// International Conference on Digital Image Computing: Techniques and Applications. Hobart, Australia; [s. n.], 2013:1-6.
- [65] Chen X, Yuan X, Yan S, et al. Towards multi-semantic image annotation with graph regularized exclusive group Lasso [C]// International Conference on Multimedia. Scottsdale, AZ, USA; ACM, 2011:263-272.
- [66] Liu F, Zhuang Y, Wu F, et al. 3D motion retrieval with motion index tree[J]. Computer Vision & Image Understanding, 2003, 92(2):265-284.
- [67] Forbes K, Fiume E. An efficient search algorithm for motion data using weighted PCA[C]// Proceedings of the ACM Siggraph/Eurographics Symposium on Computer Animation. Los Angeles, CA, USA; ACM, 2005:67-76.
- [68] Chiu C Y, Chao S P, Wu M Y, et al. Content-based retrieval for human motion data[J]. Journal of Visual Communication & Image Representation, 2004, 15(3):446-466.
- [69] Wang X, Chen L, Jing J, et al. Human motion capture data retrieval based on semantic thumbnail[J]. Multimedia Tools & Applications, 2015, 75(19):1-18.
- [70] Wu S, Wang Z, Xia S. Indexing and retrieval of human motion data by a hierarchical tree[C]// ACM Symposium on Virtual Reality Software and Technology. Kyoto, Japan; ACM, 2009:207-214.
- [71] Deng Z, Gu Q, Li Q. Perceptually consistent example-based human motion retrieval[C]// Symposium on Interactive 3D Graphics and Games. Boston, Massachusetts, USA; ACM, 2009:191-198.
- [72] Ller M, Der T, Clausen M. Efficient content-based retrieval of motion capture data[J]. ACM Transactions on Graphics, 2005, 24(2):677-685.
- [73] Lan R, Sun H, Zhu M. Text-like motion representation for human motion retrieval[M]// Intelligent Science and Intelligent Data Engineering. Berlin, Heidelberg; Springer, 2013:72-81.
- [74] Kovar L, Gleicher M. Flexible automatic motion blending with registration curves[C]// ACM Siggraph/eurographics Symposium on Computer Animation. San Diego, California, USA; ACM, 2003:214-224.
- [75] Witkin A, Popovic Z. Motion warping[C]// Conference on Computer Graphics and Interactive Techniques. [S. l.]; ACM, 1995:105-108.
- [76] Wang J, Bodenheimer B. Synthesis and evaluation of linear motion transitions. [J]. ACM Transactions on Graphics, 2008, 27(1):329-339.

- [77] Oshita M. Interactive motion synthesis with optimal blending[J]. *Computer Animation and Virtual Worlds*, 2014, 25(3/4): 311-319.
- [78] Kovar L, Gleicher M, Pighin F. Motion graphs[J]. *ACM Transactions on Graphics*, 2002, 21(3):473-482.
- [79] Arikan O, Forsyth D A. Interactive motion generation from examples[J]. *ACM Transactions on Graphics*, 2002, 21(3): 483-490.
- [80] Lee J, Chai J, Reitsma P S A, et al. Interactive control of avatars animated with human motion data[J]. *ACM Transactions on Graphics*, 2002, 21(3):491-500.
- [81] Heck R, Gleicher M. Parametric motion graphs[J]. *ACM Symposium on Interacted Graphics*, 2007, 2007:129-136.
- [82] Kwon T, Shin S Y. Motion modeling for on-line locomotion synthesis[C]// *ACM Siggraph/eurographics Symposium on Computer Animation (SCA)*. Los Angeles, CA, USA: ACM, 2005:29-38.
- [83] Min J, Chen Y L, Chai J. Interactive generation of human animation with deformable motion models[J]. *ACM Transactions on Graphics*, 2009, 29(1):89-97.
- [84] Shin H J, Lee J. Motion synthesis and editing in low-dimensional spaces: Research articles[J]. *Computer Animation & Virtual Worlds*, 2006, 17(3/4):219-227.
- [85] 王宇杰, 肖俊, 魏宝刚. 基于非线性流形学习的3维人体运动合成[J]. *中国图象图形学报*, 2010, 15(6):936-943.
Wang Yujie, Xiao Jun, Wei Baogang. 3D human motion synthesis based on nonlinear manifold learning[J]. *Journal of Image & Graphics*, 2010, 15(6):936-943.
- [86] Liu H, He F, Cai X, et al. Human motion synthesis using window-based local principal component analysis[C]// *International Conference on Computer-Aided Design and Computer Graphics*. Jinan, Shandong, China:IEEE, 2011:282-287.
- [87] Safonova A, Hodgins J K, Pollard N S. Synthesizing physically realistic human motion in low-dimensional, behavior-specific spaces[J]. *ACM Transactions on Graphics*, 2004, 23(3):514-521.
- [88] 李淳芃, 王兆其, 夏时洪. 人体运动的函数数据分析与合成[J]. *软件学报*, 2009, 20(6):1664-1672.
Li Chunpeng, Wang Zhaoqi, Xia Shihong, et al. Motion synthesis for virtual human using functional data analysis[J]. *Journal of Software*, 2009, 20(6):1664-1672.
- [89] 刘更代, 徐明亮, 张明敏. 基于独立时空特征空间的人体运动合成[J]. *计算机学报*, 2011, 34(3):464-472.
Liu Gengdai, Xu Mingliang, Zhang Mingmin. Human motion synthesis based on independent spatio-temporal feature space [J]. *Chinese Journal of Computers*, 2011, 34(3):464-472.
- [90] 蓝荣祎, 孙怀江. 基于逆运动学和重构式 ICA 的人体运动风格分析与合成[J]. *自动化学报*, 2014, 40(6):1135-1147.
Lan Rongyi, Sun Huaijiang. Style analysis and human locomotion synthesis based on inverse kinematics and reconstructive ICA[J]. *Acta Automatica Sinica*, 2014, 40(6):1135-1147.
- [91] 蓝荣祎, 孙怀江. 人体运动的稀疏语义参数化模型与交互式合成[J]. *计算机辅助设计与图形学学报*, 2013, 25(3): 341-349.
Lan Rongyi, Sun Huaijiang. A sparse semantic parametric model for interactive motion synthesis[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2013, 25(3):341-349.
- [92] 夏贵羽, 孙怀江. 模板化的人体运动合成[J]. *自动化学报*, 2015, 41(4):758-771.
Xia Guiyu, Sun Huaijiang. Templated human motion synthesis[J]. *Acta Automatica Sinica*, 2015, 41(4):758-771.
- [93] Holden D, Saito J, Komura T. A deep learning framework for character motion synthesis and editing[J]. *ACM Transactions on Graphics*, 2016, 35(4):1-11.

作者简介:



孙怀江(1968-),男,教授,研究方向:神经网络与机器学习、计算机视觉和人体运动捕获数据分析与合成,E-mail: sunhuaijiang@njust.edu.cn.



夏贵羽(1989-),男,博士研究生,研究方向:模式识别与机器学习、人体运动捕获数据分析与重用。



张国庆(1986-),男,博士研究生,研究方向:模式识别与机器学习、图像处理与计算机视觉,E-mail: xiayang14551@163.com.



封磊(1987-),男,博士研究生,研究方向:模式识别与机器学习、计算机视觉和压缩感知。