

# 计算机视觉与机器学习技术在三维人体动画中的应用综述

肖俊 庄越挺 吴飞

( 浙江大学计算机学院人工智能研究所 杭州 310027 )

( junx@cs.zju.edu.cn )

**摘 要** 近年来计算机视觉与机器学习技术广泛应用于三维人体动画领域中. 基于图像/视频的人体运动数据获取技术、数字角色和场景建模、交互式角色动画控制与运动生成等方面都大量应用了计算机视觉技术 ; 同时 , 机器学习理论在三维人体运动数据重用以及后期智能角色动画创作中扮演了越来越重要的角色 , 并得到一些很好的研究成果. 文中以三维人体动画创作流程为主线对这些方法的基本思想和关键技术进行了归纳与分析 , 并讨论了三维人体动画制作技术的发展趋势.

**关键词** 三维人体动画 ; 计算机视觉 ; 机器学习 ; 运动数据获取 ; 运动数据重用 ; 智能角色动画  
中图法分类号 TP391.41

## Computer Vision and Machine Learning in 3D Human Animation : a Survey

Xiao Jun Zhuang Yueting Wu Fei

( Institute of Artificial Intelligence , College of Computer Science and Technology , Zhejiang University , Hangzhou 310027 )

**Abstract** In recent years , computer vision and machine learning have been heavily applied in the research field of 3D human animation. Video/image based motion capture , digital character/scene modeling , interactive character controlling and motion synthesis techniques are all based on the computer vision theory. And machine learning has played an important role in the research of 3D motion data reuse and intelligent character animation and achieved some wonderful results. A survey on the computer vision and machine learning techniques used in 3D human animation is given in this paper , also with some discussion to the future work provided.

**Key words** 3D human animation ; computer vision ; machine learning ; motion capture ; motion data reuse ; intelligent character animation

近年来 , 随着数字媒体处理技术的迅速发展及计算机硬件成本的进一步降低 , 计算机动画技术及其应用得到有力推动与普及. 特别是在三维人体动画方面 , 运动捕获设备的广泛使用产生了大量真实感三维人体运动数据 , 并应用于虚拟现实、计算机游戏、动画生成、影视特技及运动仿真等多个领域.

三维人体动画技术总体上可分为 2 类 : 一是以传统计算机动画技术 , 特别是传统二维计算机动画

技术为基础发展而来的模型动画 ; 二是随着运动捕获系统的普及而产生的基于捕获数据的人体动画制作技术. 传统模型动画思想是以数学模型为基础来产生满足用户需求的动画结果 , 从本质上而言是一种基于模型的动画制作方法 , 包括关键帧动画技术、基于运动学知识的关节动画技术以及物理学/动力学方法等 ; 而基于运动捕获的动画制作技术则是以真实感三维运动数据为基础 , 采用数据驱动方式

产生三维人体动画,从本质上而言是一种数据驱动的动画制作手段,包括基于运动捕获数据的编辑、合成、重用等技术.商用人体运动捕获系统的普及使得获取真实感三维人体运动数据不再是制作真实感三维人体动画的瓶颈,可供重复利用的三维人体运动捕获数据库也已出现<sup>①</sup>,这使得数据驱动方式成为制作真实感三维人体动画的重要手段.

近年来,计算机视觉与机器学习技术在人体动画领域中的应用越来越广泛,包括基于图像/视频的人体运动数据获取技术、数字角色和场景建模、交互式角色动画控制与运动生成等方面都大量应用了计算机视觉技术.同时,机器学习理论在智能三维人体动画研究领域也被广泛应用.很多学者将研究方向转移到如下 2 个方面:1)考虑到现有商用运动捕获设备(如光学运动捕获系统)成本较高、操作过程专业性强等特点,他们致力于研究成本较低、操作简单的基于视觉方法的三维人体运动捕获技术<sup>[1]</sup>与角色动画控制技术;2)基于大量已有真实感三维人体运动数据,采用数据驱动或者机器学习方法实现三维运动数据重用.

三维人体动画创作流程分为三维运动数据获取、数据预处理、运动合成以及虚拟环境中角色行为动画生成等步骤,本文以这个流程为主线,对计算机视觉以及机器学习技术在三维人体动画创作中的应用进行了综述.

## 1 三维人体运动数据获取

采用运动捕获技术进行三维人体动画创作的本质是一种数据驱动的动画创作方式,具有数据获取容易、精度高、真实感强以及制作效率高等优点,以至于 Alias 的产品部经理 Mundell 曾说:“Motion capture is the future”.运动捕获技术的基本思想是采用特殊硬件设备实时捕获真实运动物体(演员、动物等)的三维运动数据,并将其映射到事先建立好的三维模型上,生成动画序列.

目前使用较为广泛的是光学运动捕获系统,代表性的如美国 MotionAnalysis 公司的 Hawk, Eagle 系统<sup>②</sup>以及英国 Vicon 公司的 Vicon MX 系统<sup>③</sup>.国内也出现了价格较为低廉的光学运动捕获系统,如北京迪生科技的 DMC 系统<sup>④</sup>以及大连东锐的 DVMC 系统<sup>⑤</sup>等.该类系统将被动反光标记贴在需要记录运动的物体上,利用高速摄像机拍摄具有高对比度的多视角同步视频,然后从视频中跟踪反光

标记点的二维位置,并采用机器视觉方法重建三维运动序列,如图 1 所示.就其本质而言,光学运动捕获技术也应归入基于视觉方法的三维人体运动捕获技术,只是这种系统更加依赖于高性能硬件设施来降低特征跟踪与三维重建的难度,而且这种设备非常复杂且价格昂贵.本节将重点介绍基于普通视频的三维人体运动捕获技术.



图 1 MotionAnalysis Eagle 光学运动捕获系统

### 1.1 基于视觉方法的三维人体运动捕获技术

早在 20 世纪七、八十年代,人们提出采用视觉方法从包含人体运动的视频/图像序列中提取运动参数,以达到对人体运动进行获取与分析的目的,该方法具有成本低、操作简单、应用范围广等特点,是计算机视觉与计算机动画的一个交叉研究领域,一直是学术界的研究热点.但是,从二维视频中准确地恢复三维人体运动数据是一个具有挑战性的课题<sup>[2]</sup>.基于视觉方法的三维人体运动捕获技术从技术手段上可分为基于特征跟踪、基于模型匹配以及基于侧影分析等几类.

#### 1) 基于特征跟踪的人体运动捕获技术

这类方法的基本思想是采用特征跟踪方式(如点特征跟踪、光流场跟踪等)从图像序列/视频中提取二维人体运动参数,进而采用视觉原理恢复三维人体运动参数序列.

Bregler 等<sup>[3]</sup>提出一种差分运动估计方法,通过求解简单线性系统从单目视频流中恢复具有高自由度的人体运动序列.刘小明<sup>[4]</sup>提出一种基于单相机的特征跟踪和三维运动提取方法,如图 2 所示.该方法首先采用 Kalman 滤波器预测待跟踪特征的位置,然后采用基于子块的模式匹配方法完成特征匹

① <http://mocap.cs.cmu.edu>

② <http://www.motionanalysis.com>

③ <http://www.vicon.com>

④ <http://www.disontech.com.cn/products/DMC.htm>

⑤ <http://www.dorealsoft.com/product/?category=17>

配,通过预测—修正—预测过程完成二维特征跟踪;最后利用人体模型中各部位之间存在的固定比例关系以及透视投影理论恢复人体三维运动。朱强<sup>[5]</sup>提出基于色块的人体运动特征提取方法,并为此设计了一套紧身衣,采用 Kalman 滤波与特征块匹配算法对图像序列中的彩色块进行跟踪,并基于视觉原理从单目视频流中恢复三维人体运动参数。罗忠祥<sup>[6]</sup>开发了一个基于双相机与色块紧身衣的人体运动捕获系统,如图 3 所示。该系统基于 Kalman 预测和极线约束的双相机特征跟踪、属性量化的特征跟踪以及不完全运动特征的跟踪等算法,将特征分类和特征识别等关键思想引入特征跟踪中,较好地解决了人体运动提取中的自遮挡和非刚性运动等复杂问题。

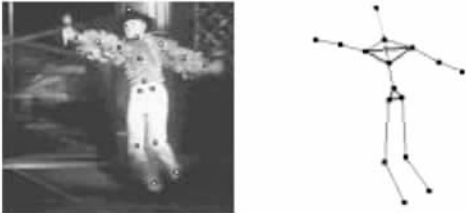


图 2 采用特征点跟踪方式获取三维人体运动

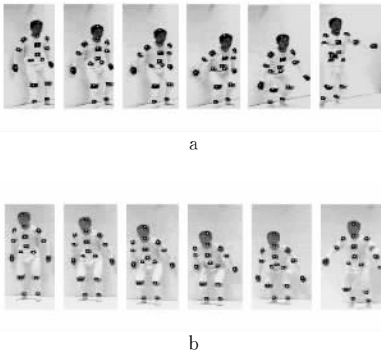


图 3 基于双相机与色块紧身衣的人体运动捕获

2) 基于模型匹配和侧影分析的人体运动捕获技术

这类方法的基本思想是给定一个人体姿态的搜索空间,在提取出来的图像特征(特征点、侧影等信息)与建立好的模型(大多是三维人体姿态模型)之间寻找合理的匹配方式,以达到三维人体运动姿态恢复的目的。通常这类问题的求解可以看作是一个优化过程。

Zhao 等<sup>[7]</sup>采用遗传算法在三维人体姿态空间中搜索一个最佳姿态来匹配二维图像中提取出来的特征点,但这种方法的准确性在很大程度上依赖于二维特征点跟踪的准确性,因而在背景比较复杂和

存在自遮挡现象的视频中较难获得满意效果。Chen 等<sup>[8]</sup>首先将三维人体姿态模型参数化,并将不同参数设置下得到的三维人体姿态绘制到二维平面,进而寻找与真实视频中获取的人体轮廓侧影最匹配的三维姿态。该算法的最大困难在于所采用的三维人体姿态模型是中性的且形状过于简单,绘制到二维平面上之后很难与真实视频流中提取的人体轮廓进行精确匹配,且三维模型到二维平面的频繁绘制导致时间效率低下。Micilotta 等<sup>[9]</sup>对辅助数据库进行预计算,运动提取时通过侧影和 Edge Map 之间的匹配实现三维人体姿态和二维图像特征之间的对比,从而获得事先计算好的三维人体姿态数据;但其中只使用了简单的上肢人体运动数据,且人体相对摄像机的角度是固定的。DiFranco 等<sup>[10]</sup>利用指定的三维关键帧信息、人体关节限制以及人体运动学知识来限制三维人体姿态与二维图像特征之间的匹配约束,从而最优化地对人体运动参数进行估计和恢复。Chen 等<sup>[11]</sup>提出一种基于侧影分析的自然视频人体运动跟踪算法,如图 4 所示。该算法首先通过侧影提取、局部侧影特征分析及欧氏距离变换等操作获取图像中的人体骨架结构,然后采取模拟退火算法在三维人体姿态空间中搜索最佳三维人体姿态数据作为恢复结果。其优点在于通过对侧影中人体骨架的提取,避免了传统侧影分析方法中采用中性人体模型所带来的匹配误差。

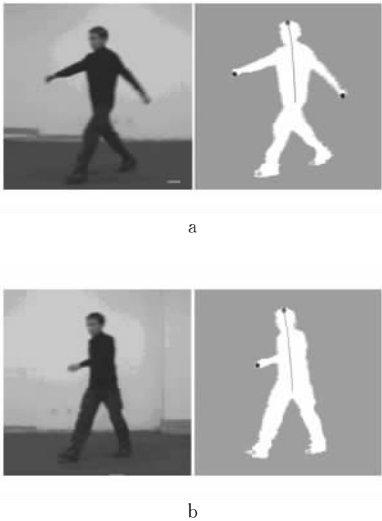


图 4 基于侧影分析的人体运动跟踪

1.2 基于统计学习的三维人体运动捕获技术

从没有明显特征或标记的视频中恢复三维人体姿态是一个缺乏足够约束条件的优化求解过程,克服这类问题的一种可行方法是基于学习方式获取

先验模型并对其进行约束. 统计学习理论在计算机视觉领域中有广泛深入的应用, 近年来也有学者将其引入基于视频的人体运动捕获技术中进行应用. 这类方法的基本思路是基于事先学习的先验知识对图像/视频中的人体运动参数进行估计, 但与机器学习理论存在的问题一样, 采用这类方法只能对预先学习过的特定姿态和运动类型取得较好恢复效果.

Agarwal 等<sup>[12]</sup>基于一组已有人体侧影图像计算其局部形状上下文分布情况, 然后用相关向量机构造一个局部形状上下文回归函数用于估计三维人体姿态. 但该方法只被应用于简单的走路运动, 当扩展到其他运动类型时回归函数构造将变得非常复杂、不准确且难以控制. Elgammal 等<sup>[13]</sup>通过学习二维图像可视特征与三维人体姿态之间的映射关系建立起 activity manifold, 如图 5 所示. 当给定一个二维图像中的可视特征时, 首先将其投影到建立好的流形中找到对应的特征点, 进而采用插值方式恢复出对应三维人体姿态. 类似的研究思路还可以在 Shakhnarovich 等<sup>[14]</sup>的工作中找到.

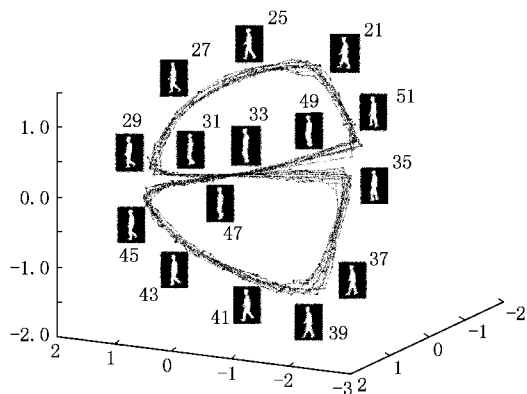


图5 人体运动侧影在底维流形空间中的统计分布

Dimitrijevic 等<sup>[15]</sup>采用已有运动捕获数据绘制一组二维人体侧影模板, 在训练中使用统计学习技术分析不同二维人体侧影部分之间的关联, 并将统计学习结果存储下来作为侧影估计的先验模型, 该方法能够对室内/室外拍摄的视频中的二维人体侧影进行跟踪与提取. Urtasun 等<sup>[16]</sup>采用主成分分析 (principle component analysis, PCA) 方法学习并得到已有三维人体运动捕获数据的运动模型, 以此作为强有力的先验知识对三维人体姿态估计的优化过程进行约束, 即采用人体运动模型构建一个基于梯度上升约束的优化函数用于估计前后 2 帧图像中三维人体姿态变化, 能够对单目/多目视频流中的走、跑、挥动球杆等三维人体姿态序列进行恢复. Ong

等<sup>[17]</sup>采用聚类算法对运动样例片断进行学习, 并将人体运动中的动力学参数表示为运动数据聚类子集之间的转换关系, 并以此为指导采用随机搜索算法从单目视频序列中估计得到三维人体姿态, 该方法能恢复与视点无关的单目视频中的走、跑、跳等三维运动数据.

在特征跟踪方面, 罗忠祥<sup>[6]</sup>提出一种无需大量训练样本的基于隐马尔可夫模型 (hidden markov model, HMM) 的特征跟踪方法. 该方法所需的训练样本由基于属性匹配和运动预测的跟踪算法自动得到, 采用 HMM 方法对其进行训练并以此产生运动参数预测, 用于指导后续视频帧中的运动参数跟踪与恢复, 而且随着跟踪的继续, 更多的跟踪结果也可以作为训练样本, 不断提高特征跟踪的精确度.

## 2 运动数据预处理

由于受捕获条件和误差影响, 或是为满足特定应用需求, 捕获得到的三维人体运动数据可能需要进行特定预处理后才能应用于三维人体动画创作. 运动数据预处理包括数据中缺失特征点重建、自然真实感三维人体运动数据评估、运动数据压缩、关键帧提取以及运动序列分割与识别等. 三维人体运动数据本质上也是一种多媒体数据, 因此在多媒体内容分析与理解中大量应用的机器学习技术在三维人体运动数据处理中同样可以得到深入应用.

Liu 等<sup>[18]</sup>通过 PCA 方法为训练集中的三维人体运动建立一组具有层次化结构的低维局部线性模型, 当处理存在丢失标记点的运动数据时, 事先建立好的分类器会选择合适的局部线性模型对丢失标记点进行重建. 此外, 他们还提出一种用于约减运动数据标记点的方法<sup>[9]</sup>, 工作流程如下: 1) 对训练集中的运动数据进行 PCA 运算, 选择能够代表运动特征的“主标记点”; 2) 对训练集中的运动数据进行概率 PCA, 将其分割为独立运动片断, 建立分段局部线性模型并提取对应特征; 3) 采用 Random Forest 分类器建立主标记点集合与分段局部线性模型之间的映射关系; 4) 在数据重建时, 对于只采用主标记点集合表示的运动数据, 采用分类器选择合适的局部线性模型进行重建. 实验结果表明: 该方法能够用于运动数据压缩, 而且在运动控制与动画制作中可以大大减少所需的数据量. 同样地, Arikan<sup>[20]</sup>采用 Bézier 曲线和聚类 PCA 方法来逼近较短运动数据片断以达到数据压缩目的.

Ren 等<sup>[21]</sup>提出一种数据驱动方法用于评估三

维人体运动数据的真实感与自然性,其工作流程如下:1)采用统计学习方式建立随时间变化的人体运动特征(关节角度、速度等)之间的概率相关性模型,以揭示真实/自然人体运动数据的内在特点;2)给定新的人体运动捕获数据,用建立好的概率模型对其进行评估。在实验中,他们分别采用了混合高斯模型(mixtures of Gaussian, MoG)、HMM 和转换式线性动力学系统等方法。

在三维人体运动数据分割方面,Arikan 等<sup>[22]</sup>基于一组手工标注的训练样本构建支持向量机分类器用于人体运动序列自动标注,该方法同样也可用于人体运动序列的自动分割。Kahol 等<sup>[23]</sup>提出层次行为分割算法,使用一个动态层次数据结构来表示人体运动元数据,进而采用简单贝叶斯分类器用于人体运动分割,但是该方法在很大程度上依赖于训练样本和来自于专家的经验数据。Barbic 等<sup>[24]</sup>认为不同类型的人体运动数据应该具有不同的内在维度,从而基于 PCA 构建运动数据分割器,当运动数据子空间模型内在维度发生变化时进行分割;进一步地,基于不同类型运动数据具有不同高斯分布模型的假设,还提出基于概率主成分分析方法(probabilistic principle component analysis, PPCA)的运动数据分割方法。这 2 种算法具有良好效果,但是运算效率比较低。除此之外,Barbic 等还认为不同类型运动数据会形成独立的聚类,且每个聚类可以采用统一的高斯分布模型来表示,并基于此假设提出了基于 GoM 的运动数据分割算法。Xiao 等<sup>[25]</sup>提出一种基于非线性流形学习的三维人体运动数据自动分割算法,如图 6 所示。该算法首先采用非线性流形降维方法将高维空间中结构复杂的三维人体运动数据投影到低维流形上,进而采用聚类方法进行自动分割,有效地提高了运动数据分割的准确性、效率和自动化程度。

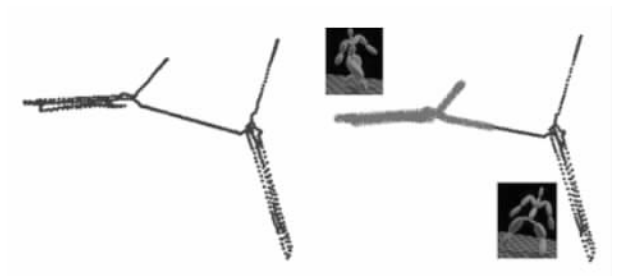


图 6 运动序列在三维流形空间中的表示及分割结果

3 运动数据重用技术

提高已有运动数据的重用性是三维人体动画创作中最受关心的问题之一,特别是运动捕获设备以

及人体运动数据库的出现,为运动数据重用技术提供了数据基础。运动数据重用技术本质上可以看作是一种数据驱动的运动生成技术,也正是由于这个特性,近年来各种子空间分析、统计学习、流形学习等机器学习技术被大量应用其中,用于对已有三维人体运动数据进行分析、学习并指导新运动数据的生成。

其中 PCA 技术应用较为广泛,如 Alexa 等<sup>[26]</sup>提出采用主成分来表示动画序列。Glardon 等<sup>[27]</sup>提出一个基于 PCA 和运动捕获数据的走路运动生成引擎:首先利用 PCA 将原始运动数据投影到低维子空间并提取可量化高层运动参数,如运动速度等;其次以这些量化高层运动参数为控制参数,通过内/外插值实时生成新的走路运动数据,并能够扩展到具有不同尺寸的角色模型上生成动画序列。Safonova 等<sup>[28]</sup>认为采用优化技术能够让用户轻松、直观地指定约束条件,从而生成满足用户要求的运动数据。但这类求解方法的原始运动数据所存在的搜索空间太大,因此他们提出一种基于 PCA 的方法将原始运动数据投影到低维空间中,以发现所需要运动数据的最本质特征并在低维空间中对其进行求解。同样地,Bowden<sup>[29]</sup>通过非线性 PCA 方法学习得到人体运动的产生式模型,可用于人体运动仿真与跟踪预测等。Mori 等<sup>[30]</sup>则利用独立成本分析(independent component analysis, ICA)得到松散姿态子空间来生成真实感人体运动数据。首先从原始高维人体运动数据中抽取能够反映人体行为特点的独立运动特征信号构建 ICA 子空间,然后通过 ICA 子空间中的插值操作生成新人体运动数据。与采用 PCA/ICA 方法构建子空间不同,Rose 等<sup>[31]</sup>首先设定具有多条坐标轴的运动“副词空间”,如图 7 所示。“副词空间”中每条坐标轴代表了不同风格的运动类型,然后采用

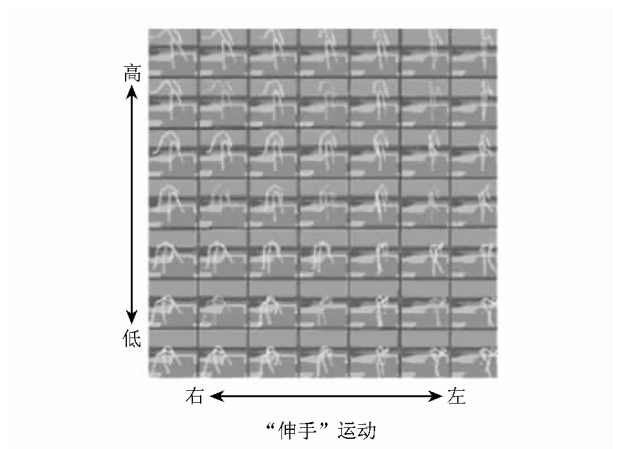


图 7 运动“副词空间”

手工标注方式为每条坐标轴标注相应的运动数据样本,进而采用径向基插值方式合成满足动画师要求的新运动数据,这种方法可以看作是一种基于监督学习机制的三维人体运动数据合成技术。

为了使运动数据的分析与生成过程更加可控与可视化,Shin 等<sup>[32]</sup>将原始运动数据通过多尺度缩放技术投影到低维空间并对其参数化与可视化,如图 8 所示。用户可以在低维空间中采取勾画草图、直接操纵三维人体姿态以及指定关键帧等方式对原始运动数据在时域和空域上进行编辑、插值、合成等操作。Mukai 等<sup>[33]</sup>将运动插值看作是在预定义参数空间中的统计预测问题,并基于泛克里格方法来估计运动非相似性与其在对应参数空间中采样点距离之间的关系。该算法能够基于统计方法对给定参数值的插值核进行优化,并将参数空间中指定的空间约束进行准确预测与生成。

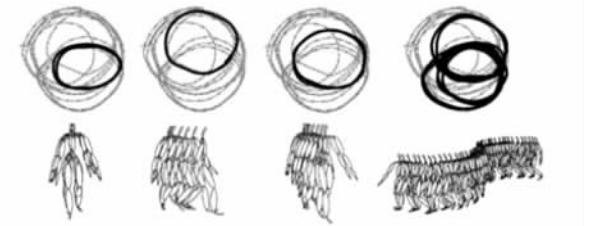


图 8 在低维空间中采取勾画草图方式生成运动数据

在三维人体动画中,人体骨架运动受到一定的运动学约束,通常这种约束知识都是由专家给出的;而 Ong 等<sup>[34]</sup>则提出一种层次化聚类模型用于对已有运动数据进行学习,它不仅能够通过学习得到各个关节的运动学约束知识,同时还能揭示不同关节运动角度之间的内在联系,从而建立起整个角色骨架的运动学约束模型用于指导运动生成。Li 等<sup>[35]</sup>提出运动纹理概念,运动纹理实际上是一个具有两层结构的统计模型,位于底层的是 Motion Texton,即可以采用相同线性动力学系统表示的一组运动数据片段,然后采用 HMM 学习得到不同 Motion Texton 之间的转换矩阵,最后根据学习得到的运动纹理模型就可以生成形式各异真实感三维人体运动数据。Lim 等<sup>[36]</sup>还提出一种基于遗传算法的交互式运动进化方法用于动画创作:系统首先根据用户要求生成一段候选运动数据,并让用户根据其喜好进行标注,然后根据用户反馈,基于遗传算法生成下一代运动数据。此外,在基于标注的运动合成方法中,Arikan 等<sup>[22]</sup>采用支持向量机对人体运动数据片段进行自动分类和标注,大大地提高了运动合成效率。

风格化三维人体运动数据生成是当前人体数据重用中的研究热点,它在个性化动画制作及群体动画制作中有着重要应用。近年来,采用机器学习技术进行风格化三维人体运动数据的编辑与生成受到较多学者关注并取得较多科研成果。实际上,前文提到的“副词空间”<sup>[31]</sup>技术以及基于 PCA 的走路运动合成引擎<sup>[27]</sup>都能够用于风格化运动数据生成。

在这方面的研究中,较为典型的是 Brand 等<sup>[37]</sup>提出的通过构建风格机器从人体运动数据训练集中学习各种运动模式,并利用训练好的模型产生不同风格的人体运动数据,但这种方法在生成风格化人体运动数据时不能对“风格”进行量化处理。Torresani 等<sup>[38]</sup>通过学习方式来建立风格参数化模型,他们认为真实感的风格化人体运动数据可以通过不同运动序列之间的时空插值方式来实现,提出了一种能够学习不同运动数据之间插值权重函数的模型。Hsu 等<sup>[39]</sup>提出“运动风格迁移”概念,将输入运动数据的风格进行改变,但同时保留原有运动数据的内容(即运动的形式)。该算法首先对运动数据进行对齐,然后采用线性时不变模型学习与表达具有不同风格的同类型运动数据之间的差异,最后采用训练好的模型改变输入运动数据的风格特征。Grochow 等<sup>[40]</sup>在对输入人体姿态进行训练学习的基础上提出一个逆向动力学系统,能够得到针对不同风格的逆向运动学算子,但这里的风格定义稍有不同,他们的研究认为每一个具体运动数据就是一种风格。

肖俊等<sup>[41]</sup>提出一个自动框架对风格化人体运动数据进行自动、实时和定量的生成与编辑,如图 9 所示。PCA 用来将人体运动数据映射到子空间,在最大程度保持原始数据特征的条件下降低了计算复

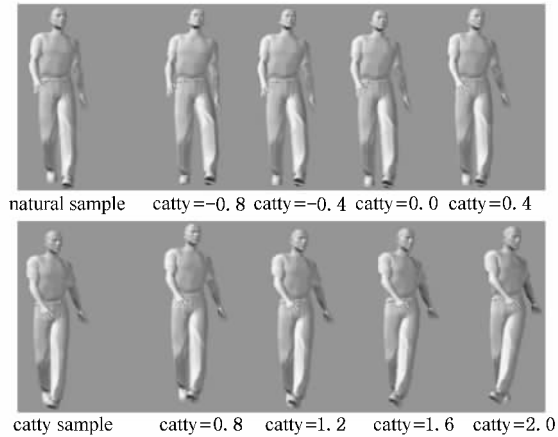


图 9 风格化运动生成结果

杂度. 风格化运动生成与编辑算法在 PCA 子空间中得到应用并产生出新的风格化人体运动数据. 为了解决实际人体运动存在多风格融合的问题, 肖俊等还提出一种新颖的方法用于多风格人体运动数据的生成与编辑. 同样地, Urtasun 等<sup>[42]</sup>将 PCA 方法进行了扩展, 能够生成特定角色具有不同风格的三维人体运动数据.

人体运动的风格参数非常抽象, 很难通过用户进行手工设定, Liu 等<sup>[43]</sup>提出非线性逆向优化算法, 能够从已有风格化人体运动数据集中自己学习估计最优化的风格参数. Wang 等<sup>[44]</sup>也提出一种能够实时生成多种风格化人体姿态的合成算法, 采用自组织混合网络学习已有人体运动数据的风格模型, 并通过风格变量控制风格模型产生风格各异的三维人体姿态. Silva 等<sup>[45]</sup>提出一种用于生成“情感姿态”的方法, 首先对已有三维人体姿态提取空间特征, 采用统计学习技术建立起姿态空间特征与情感语意之间的映射模型, 从而利用映射模型生成情感各异的三维人体姿态.

## 4 行为动画生成

计算机动画, 特别是三维人体动画研究中的一个重要内容就是实现自动化、智能化的人体行为动画生成. 有的学者通过构建具有自主决策能力的自治体来实现这个目的, 也有的学者构建行为模型来指导虚拟角色行为动画生成<sup>[46]</sup>.

Conde 等<sup>[47]</sup>使用增强学习理论对虚拟角色所在的虚拟环境进行学习并分析虚拟场景的层次结构, 在此基础上构建虚拟角色在场景中漫游的真实感行为驱动引擎. Noser 等<sup>[48]</sup>和 Kuffner 等<sup>[49]</sup>基于合成视觉、记忆及高层推理学习机制, 为虚拟角色建立了多通道、高层次的行为决策与驱动模型, 实现了障碍物场景中虚拟角色的自主漫游行为动画生成.

Dinerstein 等<sup>[50]</sup>将虚拟角色的行为选择存储为一组“状态-行为”对, 用于表示虚拟角色在某种情形下的行为选择方式, 进而采用神经网络对事先建立好的“状态-行为”对进行离线训练, 建立起不同“状态”和“行为”之间的映射关系函数, 以指导虚拟角色的行为决策. 由于是离线训练, 因此这种方法表现出了较好的实时性, 能够为虚拟场景中多个角色的自主行为动画生成提供帮助, 场景中的角色能够根据外部环境自主选择合适的行为方式, 如图 10 所示. 还有一些学者对角色的行为决策进行预判, 并预

先计算相应的行为数据用于实时动画生成, Lee 等<sup>[51]</sup>从已有拳击运动数据中学习得到一组控制规则, 用于对虚拟角色的拳击行为进行预计算, 从而达到实时制作拳击行为动画的目的.



图 10 自主行为动画示例

Dinerstein 等<sup>[52]</sup>还采用学习手段为虚拟角色建立行为与感知模型, 其中包括使用 K 近邻算法和基于实例的学习技术在内的方法, 以使得智能体能够快速、鲁棒地学习与不同用户之间的交互行为, 能够为交互式虚拟环境和游戏中的虚拟角色提供动态行为规划与生成. 在 Badler 等<sup>[53]</sup>和 Lee 等<sup>[54]</sup>的工作中, 机器学习技术被用于为虚拟角色提供记忆模型, 使之能够记住用户早些时候提供的信息和发出的指令. Liu 等<sup>[55]</sup>提出一种自组织结构用于学习虚拟场景结构以及到达场景中某个目标点的行为方式等, 用于实现自主动画生成. Mataric 等<sup>[56]</sup>提出虚拟人模仿学习概念, 即采用机器学习理论赋予虚拟角色一定的自主学习能力, 通过训练使之能够模拟用户演示的肢体行为.

在仿真训练中, Fernlund 等<sup>[57]</sup>利用基于上下文的推理和遗传规划算法来学习士兵在战斗中的技战术行为规划与产生模型, 并以此在战斗仿真训练中自动产生合理的技战术动作.

除了采用机器学习方式进行自动行为动画生成外, 也有学者采用视觉交互方式来实时地合成行为动画序列. Lee 等<sup>[58]</sup>提出一种基于视觉的交互式运动序列生成技术, 由身着紧身衣的演员进行表演, 并将跟踪得到的侧影图像输入系统作为控制信号, 进而从已有三维人体运动数据库中检索相似运动数据进行合成. 同样地, Chai 等<sup>[59]</sup>将反光标记点贴在演员的主要关节处, 对其进行跟踪和三维重建后作为控制信号输入系统, 用于从运动数据库中检索相似运动帧进行行为生成. Ren 等<sup>[60]</sup>采用多个相机从不同角度拍摄演员运动序列以提取其侧影特征, 用来

索引运动数据库中的三维人体姿态,能够实时地合成三维人体运动序列。

## 5 总 结

一般来说,计算机动画属于计算机图形学的研究范畴,但是从技术手段上来看,三维人体动画涉及计算机视觉、计算机图形学、人工智能、机器学习甚至认知科学等领域,已经成为一个多学科技术交叉的研究领域。早在 1998 年,就有学者对计算机图形学和计算机视觉技术的融合趋势进行了深刻的分析和展望<sup>[61]</sup>。在人体动画研究领域,这种融合得到了尤其直观的体现,不仅仅是基于图像/视频的人体运动数据获取技术涉及计算机视觉方法,包括基于图像/视频的数字角色和场景建模、交互式角色动画控制与生成等方面都大量应用了计算机视觉技术;同时,机器学习理论在三维人体动画创作中也得到了广泛深入的应用,并取得了一些很好的研究成果。同样地,不仅是在三维人体动画制作领域,在整个图形学领域的研究中机器学习理论也正在得到广泛深入的应用<sup>[62]</sup>。正是基于这样的—个事实,本文对计算机视觉与机器学习技术在三维人体动画制作中的应用进行了综述,并以三维人体动画制作流程为主线进行了归纳和分类。

从本文的归纳可以看到:计算机视觉和机器学习理论正在非常迅速地融入到人体动画研究领域并已经取得了较好的结果,这也是将来人体动画研究领域的一个重要趋势。从人体动画制作本身的需求来看,方便快捷的素材获取机制以及智能高效的后期制作技术已然是 2 个最为重要的研究课题,因此未来几年可能取得突破的研究方向有:1)在动画制作素材获取方面,包括基于视觉原理的更加高效、鲁棒、方便的人体运动和表情获取技术,基于图像/视频/手绘草图的角色及场景建模技术以及基于视觉交互方式的行为动画制作技术及硬件平台等;2)在后期动画制作方面,包括更具一般性的基于机器学习的三维人体运动/表情数据重用技术以及更具推广性的基于学习机制的智能人体动画生成技术等。但是我们也要看到,机器学习理论是通过已有样例的统计、分析来挖掘隐藏其中的高层产生式模型,其本质可以看作是一种数据驱动的工作机制,这使得它解决问题的一般性和可推广性受到限制。而基于心理学和认知科学的研究也有可能产生突破,发掘出更具一般性的人体行为、表情乃至情感的产生

机制,这将极大地推动现有三维人体动画创作技术的进步。

## 参 考 文 献

- [1] Moeslund T B, Hilton A, Kruger V. A survey of advances in vision-based human motion capture and analysis[J]. *Computer Vision and Image Understanding*, 104(2):90-126
- [2] Gleicher M, Ferrier N. Evaluating video-based motion capture [C]//*Proceedings of Computer Animation*, Geneva, 2002:75-78
- [3] Bregler C, Malik J. Video motion capture [R]. Berkeley: University of Berkeley, CSD-97-973, 1997
- [4] Liu Xiaoming. Study on video based human animation techniques [D]. Hangzhou: Zhejiang University, 1999 (in Chinese)  
(刘小明. 基于视频的人体动画技术的研究 [D]. 杭州:浙江大学, 1999)
- [5] Zhu Qiang. Human motion extraction and regenerating from video [D]. Hangzhou: Zhejiang University, 2002 (in Chinese)  
(朱强. 视频流中人体运动信息的提取和再生技术 [D]. 杭州:浙江大学, 2002)
- [6] Zhongxiang Luo. Video motion extraction and motion synthesis [D]. Hangzhou: Zhejiang University, 2002 (in Chinese)  
(罗忠祥. 视频流中人体运动提取与运动合成 [D]. 杭州:浙江大学, 2002)
- [7] Zhao J H, Li L. Human motion reconstruction from monocular images using genetic algorithms[J]. *Computer Animation and Virtual Worlds*, 2004, 15(3/4):407-414
- [8] Chen Y S, Lee J, Parent R, *et al.* Markerless monocular motion capture using image features and physical constraints [C]//*Proceedings of the Computer Graphics International*, Stony Brook, 2005:36-43
- [9] Micilotta A S, Ong E J, Bowden R. Real-time upper body 3D pose estimation from a single uncalibrated camera [C]//*Proceedings of EUROGRAPHICS 2005*, Dublin, 2005:41-44
- [10] DiFranco D E, Cham T, Rehg J M. Reconstruction of 3-D figure motion from 2-D correspondences [C]//*Proceedings of 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Prague, 2001:307-314
- [11] Chen C, Zhuang Y Ting, Xiao J. Towards robust 3D reconstruction of human motion from monocular video [C]//*Proceedings of the 16th International Conference on Artificial Reality and Telexistence*, Hangzhou, 2006:594-603
- [12] Agarwal A, Triggs B. Recovering 3D human pose from monocular images[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(1):44-58
- [13] Elgammal A, Lee C S. Inferring 3D body pose from silhouettes using activity manifold learning [C]//*Proceedings of 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington D C, 2004, 2:681-688
- [14] Shakhnarovich G, Viola P, Darrell T. Fast pose estimation with parameter sensitive hashing [C]//*Proceedings of the 9th IEEE International Conference on Computer Vision*, Nice, 2003, 2:750-757



- [15] Dimitrijevic M, Lepetit V, Fua P. Human body pose detection using Bayesian spatio-temporal templates[J]. *Computer Vision and Image Understanding*, 2006, 104(2):127-139
- [16] Urtasun R, Fleet David J, Fua P. Temporal motion models for monocular and multiview 3D human body tracking[J]. *Computer Vision and Image Understanding*, 2006, 104(2):157-177
- [17] Ong E J, Micilotta A S, Bowden R, *et al.* Viewpoint invariant exemplar-based 3D human tracking[J]. *Computer Vision and Image Understanding*, 2006, 104(2):178-189
- [18] Liu G D, McMillan L. Estimation of missing markers in human motion capture[J]. *The Visual Computer*, 2006, 22(9/11):721-728
- [19] Liu G, Zhang J D, Wang W, *et al.* Human motion estimation from a reduced marker set[C]//*Proceedings of the 2006 symposium on Interactive 3D Graphics and Games*, Redwood, 2006:35-42
- [20] Arikan O. Compression of motion capture databases[J]. *ACM Transactions on Graphics*, 2006, 25(3):890-897
- [21] Ren L, Patrick A, Efros A A, *et al.* A data-driven approach to quantifying natural human motion[J]. *ACM Transactions on Graphics*, 2005, 24(3):1090-1097
- [22] Arikan O, Forsyth D A, O'Brien J. Motion synthesis from annotations[J]. *ACM Transactions on Graphics*, 2003, 22(3):402-408
- [23] Kahol K, Tripathi P, Panchanathan S. Automated gesture segmentation from dance sequences[C]//*Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition*, Seoul, 2004:883-888
- [24] Barbic J, Safonova A, Pan J Y, *et al.* Segmenting motion capture data into distinct behaviors[C]//*Proceedings of Graphics Interface*, London, Ontario, 2004:185-194
- [25] Xiao J, Zhuang Y T, Wu F. Getting distinct movements from motion capture data[C]//*Proceedings of Computer Animation and Social Agents*, Geneva, 2006:33-42
- [26] Alexa M, Müller W. Representing animations by principal components[J]. *Computer Graphics Forum*, 2000, 19(3):411-426
- [27] Glardon P, Boulic R, Thalmann D. PCA-based walking engine using motion capture data[C]//*Proceedings of the Computer Graphics International*, Crete, 2004:292-298
- [28] Safonova A, Hodgins J K, Pollard N. Synthesizing physically realistic human motion in low-dimensional, behavior-specific spaces[J]. *ACM Transactions on Graphics*, 2004, 23(3):514-521
- [29] Bowden R. Learning statistical models of human motion[C]//*Proceedings of IEEE Workshop on Human Modeling, Analysis, and Synthesis*, Hilton Head Island, 2000:10-17
- [30] Mori H, Hoshino J. ICA-based interpolation of human motion[C]//*Proceedings of IEEE International Symposium on Computational Intelligence in Robotics and Automation*, Kobe, 2003, 1:453-458
- [31] Rose C, Cohen M F, Bodenheimer B. Verbs and adverbs: multidimensional motion interpolation[J]. *IEEE Computer Graphics and Applications*, 1998, 18(5):32-40
- [32] Shin H J, Lee J. Motion synthesis and editing in low-dimensional spaces[J]. *Computer Animation and Virtual Worlds*, 2006, 17(3/4):219-227
- [33] Mukai T, Kuriyama S. Geostatistical motion interpolation[J]. *ACM Transactions on Graphics*, 2005, 24(3):1062-1070
- [34] Ong E J, Hilton A. Learnt inverse kinematics for animation synthesis[J]. *Graphical Models*, 2006, 68(5):472-483
- [35] Li Y, Wang T S, Shum H Y. Motion texture: a two-level statistical model for character motion synthesis[J]. *ACM Transactions on Graphics*, 2002, 21(3):465-472
- [36] Lim I S, Thalmann D. Pro-actively interactive evolution for computer animation[C]//*Proceedings of Eurographics Workshop on Animation and Simulation*, Milan, 1999:45-52
- [37] Brand M E, Hertzmann A. Style machines[C]//*Proceedings of the 27th annual Conference on Computer Graphics and Interactive Techniques*, New Orleans, 2000:183-192
- [38] Torresani L, Hackney P, Bregler C. Learning to synthesize motion styles[OL]. [2007-12-17] <http://movement.stanford.edu/learning-motion-styles/>
- [39] Hsu E, Pulli K, Popovi J. Style translation for human motion[J]. *ACM Transactions on Graphics*, 2005, 24(3):1082-1089
- [40] Grochow K, Martin Steven L, Hertzmann A, *et al.* Style-based inverse kinematics[J]. *ACM Transactions on Graphics*, 2004, 23(3):522-531
- [41] Xiao Jun, Zhuang Yueting, Chen Cheng, *et al.* Automatic synthesis and editing of motion styles[C]//*Proceedings of the 3rd Conference on Intelligent CAD and Digital Entertainment*, Ji'nan, 2006:249 (in Chinese)  
(肖俊, 庄越挺, 陈成, 等. 风格化人体运动数据的自动生成与编辑[C]//*第三届智能CAD与数字娱乐学术会议论文集*. 济南, 2006:249)
- [42] Urtasun R, Glardon P, Boulic R, *et al.* Style-based motion synthesis[J]. *Computer Graphics Forum*, 2004, 23(4):799-812
- [43] Liu C K, Hertzmann A, Popovic Z. Learning physics-based motion style with nonlinear inverse optimization[J]. *ACM Transactions on Graphics*, 2005, 24(3):1071-1081
- [44] Wang Y, Liu Z Q, Zhou L Z. Key-styling: learning motion style for real-time synthesis of 3D animation[J]. *Computer Animation and Virtual Worlds*, 2006, 17(3/4):229-237
- [45] De Silva P R, Bianchi-Berthouze N. Modeling human affective postures: an information theoretic characterization of posture features[J]. *Computer Animation and Virtual Worlds*, 2004, 15(3/4):269-276
- [46] Millar R J, Hanna J R P, Kealy S M. A review of behavioral animation[J]. *Computer and Graphics*, 1999, 23(1):127-143

- [ 47 ] Conde Y , Thalmann D . Learnable behavioural model for autonomous virtual agents : low-level learning[ C ]//Proceedings of the 5th International Joint Conference on Autonomous Agents and Multiagent Systems , Hakodate , 2006 : 89-96
- [ 48 ] Noser H , Renault O , Thalmann D . Navigation for digital actors based on synthetic vision , memory , and learning[ J ] . Computer & Graphics , 1995 , 19( 1 ) : 7-19
- [ 49 ] Kuffner J J , Latombe J C . Fast synthetic vision , memory and learning models for virtual humans [ C ] //Proceedings of the Computer Animation , Geneva , 1999 : 118-127
- [ 50 ] Dinerstein J , Egbert P , de Garis H , *et al.* . Fast and learnable behavioral and cognitive modeling for virtual character animation [ J ] . Computer Animation and Virtual Worlds , 2004 , 15( 2 ) : 95-108
- [ 51 ] Lee J , Lee K H . Precomputing avatar behavior from human motion data[ J ] . Graphical Models , 2006 , 68( 2 ) : 158-174
- [ 52 ] Dinerstein J , Egbert P K . Fast multi-level adaptation for interactive autonomous characters [ J ] . ACM Transactions on Graphics , 2005 , 24( 2 ) : 262-288
- [ 53 ] Badler N I , Chi D M , Chopra S . Virtual human animation based on movement observation and cognitive behavior models [ C ]//Proceedings of the Computer Animation , Geneva , 1999 : 128-137
- [ 54 ] Lee J R , Williams A B . Behavior development through task oriented discourse [ J ] . Computer Animation and Virtual Worlds , 2004 , 15( 3/4 ) : 327-337
- [ 55 ] Liu J M , Qin H . Behavioral self-organization in lifelike synthetic agents [ J ] . Autonomous Agents and Multi-Agent Systems , 2002 , 5( 4 ) : 397-428
- [ 56 ] Mataric M J . Getting humanoids to move and imitate[ J ] . IEEE Intelligent Systems , 2000 , 15( 4 ) : 18-24
- [ 57 ] Fernlund H K G , Gonzalez A J , Georgiopoulos Michael , *et al.* . Learning tactical human behavior through observation of human performance[ J ] . IEEE Transactions on Systems , Man , and Cybernetics , 2006 , 36( 1 ) : 128-140
- [ 58 ] Lee J , Chai J X , Reitsma P , *et al.* . Interactive control of avatars animated with human motion data [ J ] . ACM Transactions on Graphics , 2002 , 21( 3 ) : 491-500
- [ 59 ] Chai J X , Hodgins J K . Performance animation from low-dimensional control signals[ J ] . ACM Transactions on Graphics , 2005 , 24( 3 ) : 686-696
- [ 60 ] Ren L , Shakhnarovich G , Hodgins J , *et al.* . Learning silhouette features for control of human motion[ J ] . ACM Transactions on Graphics , 2005 , 24( 4 ) : 1303-1331
- [ 61 ] Lengyel J . The convergence of graphics and vision [ J ] . Computer , 1998 , 31( 7 ) : 46-53
- [ 62 ] Dinerstein J , Egbert P K , Cline D . Enhancing computer graphics through machine learning : a survey [ J ] . The Visual Computer , 2007 , 23( 1 ) : 25-43

## CAD 学报“ 动态场景实时绘制 ”专辑征文通知

真实感图形的实时绘制一直是图形学研究的热点和主流。《计算机辅助设计与图形学学报》拟于 2008 年下半年出版“ 动态场景实时绘制 ”专辑。本专辑的主题包括：

- ① 大规模、复杂、动态场景的实时绘制算法；
- ② 实时阴影算法；
- ③ 全局光照算法。

征文截稿日期 2008 年 4 月 30 日，热诚欢迎相关专家、学者和研究生踊跃投稿。

请通过本刊远程稿件处理系统投稿，网址 <http://jcad.ict.ac.cn> 或 <http://www.jcad.cn>。

稿件上请注明“ 动态场景实时绘制专辑征文 ”字样。