

人体动作识别与评价综述

宋震, 张宇姝, 杨刚

(中央戏剧学院, 100710; 北京林业大学, 北京 100083)

摘要: 近年来, 对人体动作进行识别和评价正逐渐成为研究和应用热点。动作识别是指对给定的视频或三维动作序列中人的行为动作进行识别, 并判断其所属的行为类别。动作评价是指对特定人体动作(如体操、划船、舞蹈等的标准动作)的规范程度或完成质量进行评价, 从而辅助人们更好地完成标准动作。动作评价需要基于动作识别来进行。本文以人体动作识别和评价的处理流程为主线, 对近年来相关工作进行了全面而系统地梳理, 阐述了数据预处理、特征提取、动作分类、动作评价的标准等各个关键环节的研究进展及存在问题, 最后对当前研究的难点问题以及未来的发展趋势进行总结与展望。

关键词: 动作识别; 动作评价; 特征提取; 相似性度量

中图分类号: TP391.4 **文献标识码:** A **文章编号:** XXXXXX

Overview of Human Action Recognition and Evaluation

Song Zhen, Zhang Yu-shu, Yang Gang

(The Central Academy of Drama, Beijing 100710/102209; Beijing Forestry University, Beijing 100083, China)

Abstract: In recent years, the recognition and evaluation of human actions has gradually become a research and application hotspot. Action recognition refers to recognizing the actions of people in a given video or three-dimensional action sequence, and judging the behavior category it belongs to. Action evaluation refers to the evaluation of the degree of standardization or completion quality of specific human actions (such as gymnastics, rowing, dancing, etc.), so as to assist people to better complete standard actions. Action evaluation needs to be based on action recognition. This paper takes the processing flow of human action recognition and evaluation as the main line, comprehensively and systematically combs the related work in recent years, expounds the research progress and research progress of each key link such as data preprocessing, feature extraction, action classification, and action evaluation standards. There are problems, and finally the difficult problems of current research and future development trends are summarized and prospected.

Keywords: action recognition; action evaluation; feature extraction; similarity measurement

作者简介: 宋震(1976-), 男(汉族), 北京人, 中央戏剧学院教授, songzhen@zhongxi.cn; 张宇姝(1997-), 女(汉族), 湖北省十堰市人, 北京林业大学硕士研究生, 18810772678@163.com; 杨刚(1977-), 男(汉族), 山西省长治市人, 北京林业大学副教授, yanggang@bjfu.edu.cn.

项目资助: 北京高校卓越青年科学家计划项目(项目编号: BJJWZYJH01201910048035) 资助. Beijing Outstanding Young Scientist Program (BJJWZYJH01201910048035).

1 引言

近年来,对人体动作进行识别和评价正逐渐成为研究和应用热点。动作识别是指对给定的动作序列数据(视频或三维动作序列)进行分析,从中识别并判断出其包含的动作类别。自动人体动作识别在监控^[1-4],人机交互^[5],辅助技术^[6]和消费者行为分析^[7]等方面都有着广泛的应用。动作评价则是对某一标准动作的完成质量进行评价,其多应用于体操、划船、舞蹈等专业领域的动作评判和动作训练中。动作评价往往需要在动作识别的基础上进行,通过增加专家知识来对专业领域动作的规范性、流畅性、艺术性进行判断。

动作识别与动作评价之间有显著区别:动作识别可以看作一个多分类问题,主要在于定量地比较输入数据和标准参考的相似性,然后输出动作类型的标签;在动作评价中,则并非单纯地进行数据相似性判定,而是侧重于在专家知识的指导下,对专业领域动作的规范性、流畅性、艺术性甚至是人体肌肉发力程度的分析评价,衡量的是一种更深层次、更具专业性质的相似度。但动作识别和动作评价并不是完全独立的两项工作,两者在技术环节和技术方法上有着密不可分的关系,例如,二者都需要对动作序列进行时空对齐等预处理、都需要对动作进行特征提取等。而动作评价往往需要在动作识别中特征提取基础之上,进一步结合专家知识展开评价计算。

经过数十年的发展,动作识别的相关研究工作不仅数量丰富,而且已经取得了很好的效果。而动作评价相关研究还处于探索阶段,目前虽然已有一些关于动作评价的文章、系统,但他们所针对的主要还是简单、重复性的动作,比如高尔夫球^[8]、羽毛球^[9]、划船等体育动作;比较的也是定量的数据,如关节角度、速度、加速度等。而健美操^[10]、舞蹈^[11]、太极拳^[12]、京剧表演^[13]等动作的评价则具有更高的复杂性,对他们的评价不能仅限于简单的姿态参数,还要分析更高

层面的专业“相似度”。

在广泛分析相关研究工作的基础上,本文对动作识别与动作评价的关键技术环节、主要技术方法进行了系统分析与梳理,得到了如图1所示的技术框架图。根据图1所示,动作识别和动作评价首先都需要进行数据输入、数据预处理和特征提取;之后,动作识别会基于所提取的特征进行动作分类;而动作评价则需要结合领域专家知识,在动作特征的基础上进一步提取出评价特征,以进行动作评价。

本文将基于图1所示的技术路线展开讨论,分别从数据对象、数据预处理、特征提取、动作识别的分类方法和动作评价研究这几个方面出发对相关研究工作进行全面地归纳;分析各种典型方法的研究进展、优缺点及其应用情况;并对当前研究的难点问题以及未来的发展趋势进行总结与展望。

2 数据预处理

2.1 数据类型

人体动作识别和动作评价中所使用的输入数据一般有两种:视频数据和3D骨骼数据。

基于视频数据的动作识别就是从相机拍摄的动作视频序列中识别出人体动作,解决这种基于视频的动作识别问题有两种思路:一是直接抽取并分类序列的时空特征的方法;二是提取骨骼信息(2D或3D骨骼信息)进行训练的方法。基于视频的动作识别和图像分类很相似,只是相比于图像分类,视频多了一个时序维度。由于视频数据的获取非常方便,因此受到研究和应用的青睐,相关工作层出不穷。但视频数据在拍摄时容易出现遮挡、抖动、视角变化等情况,这也为动作识别带来了困难。

随着光学和惯性动作捕捉设备的进步,人们可以直接捕捉人体动作的3D骨骼数据。3D骨骼数据具体来说就是骨骼动画数据,它

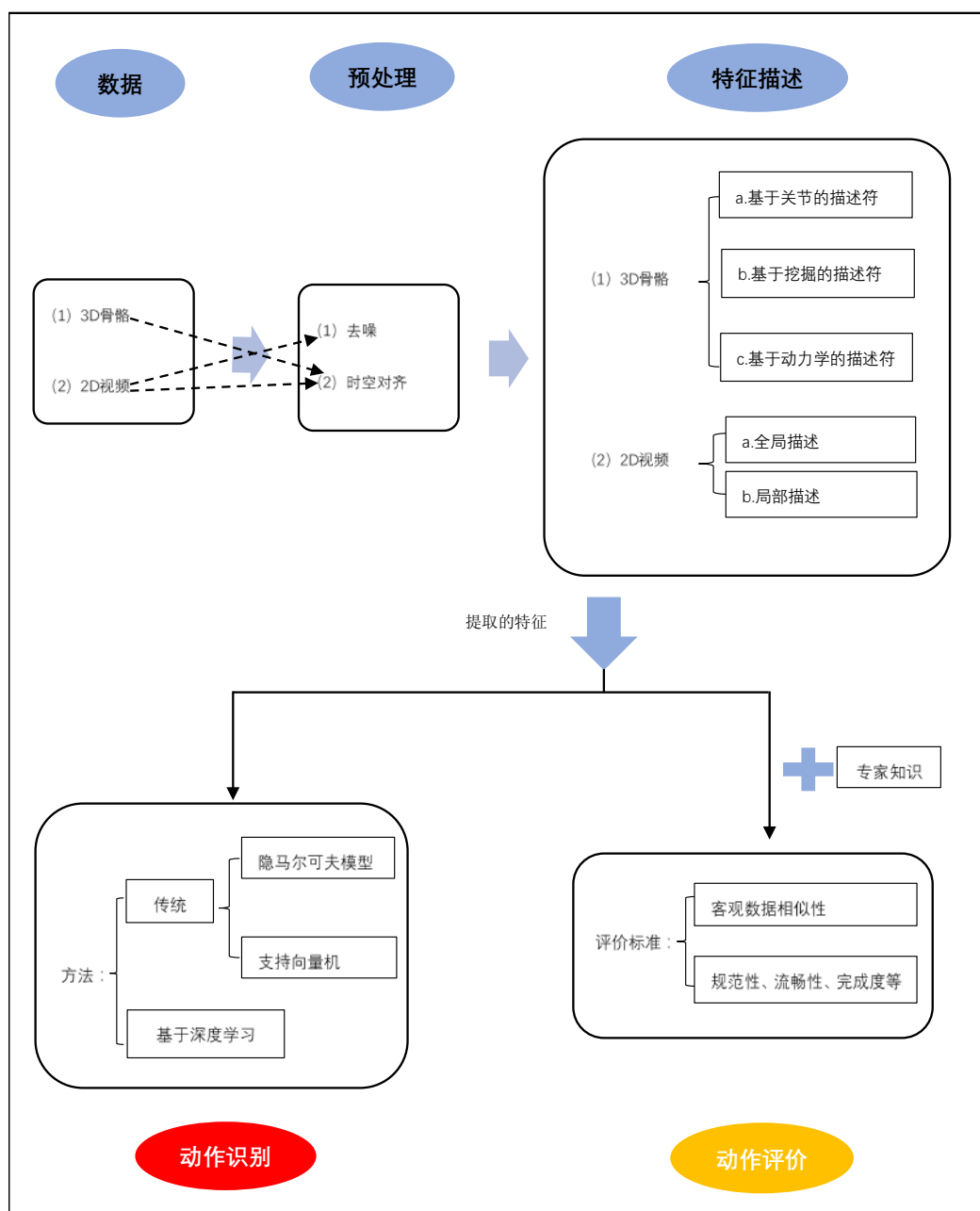


图 1 动作识别与动作评价的技术框架图

是以图形形式对人体所进行的描述。如图 2 是一个典型的人体骨骼示例图。模型骨骼数量一般都有一个限制,以使得既能全面覆盖所需骨骼信息,又能减少数据冗余。而且通常会选择模型的盆骨做为模型的根骨骼,基于根骨骼,可以递推出各个骨骼相对于根骨的转换矩阵。3D 骨骼数据可以更加直接的表示身体部位运动相关特征,如关节角度、速

度等,从而可以更方便准确地进行动作识别^[14]。

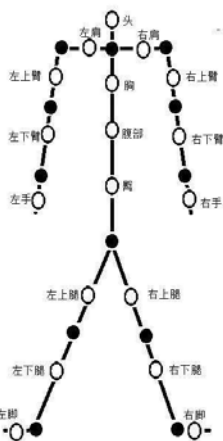


图 2 人体骨骼示意图

2.2 数据预处理的方法

为了使数据更加精确且便于进行后续工作,对其进行预处理是动作识别首先要完成的步骤,预处理包括去噪和时空对齐问题。

3D 骨骼数据基本不受采集环境的影响,噪声很少,因此无需去噪处理。而视频数据,由于测量时受到一些外在不确定因素的影响,原始数据会存在信息不稳定或者冗余的现象,所以必须进行处理。视频数据的去噪处理包括空洞修复、图像平滑等操作^[15]。Diebel 等人^[16]最先提出了一种基于马尔可夫随机场的深度升采样方法进行图像平滑处理。Park 等人^[17]通过改进原始马尔可夫随机场模型中的平滑约束项,并使用彩色图像的分割信息以及彩色图像的梯度信息作为新的约束项,解决了之前方法处理结果中的深度图像边缘过平滑的问题。

所谓“时空对齐问题”是指:在比较两个骨骼动作时,需要保证所比较的两套动作在时间上快慢基本一致,在空间上大小基本相同。具体而言,时间上,人完成动作的快慢是不同的,必须先对齐两个序列,使关键帧能对应起来;空间上,不同人的骨骼大小不同,这对某些参数的比较会造成干扰,比如关节角度、角速度等,所以要先进行骨骼标准化处理。

关于对齐时间序列,Nowozin 和 Shotton^[18]采用了一个固定的滑动窗口,引入了时间锚定动作的动作点概念^[19]。关于对

齐空间骨骼,Fotini 等^[14]在数据分析之前,应用骨骼标准化作为预处理步骤,旨在防止由于身体结构差异导致的错误分析。Asteriadis 等^[20]使用身体之间(由肩部和躯干 3D 位置提取)的旋转偏移实现空间动作对齐。

3 动作识别的特征描述方法

特征描述是动作识别和分析的前提,它是将原始特征转换为一组具有明显物理意义(如几何特征、纹理)或者统计意义的特征,以突出更具有代表性的特征,是一种提取有效信息的方法。从序列中将描述人体姿势的有效特征提取出来是保证准确识别动作的重要前提。

视频数据和 3D 骨骼数据差异比较大,对其进行特征描述的方式也是不同的。视频数据的特征描述可以分为局部描述和全局描述两类。3D 骨骼数据的特征描述可以分为三类:(1)基于关节的描述符;(2)基于挖掘的描述符;(3)基于动力学的描述符。

下面我们对这些特征描述方法分别进行介绍。

3.1 视频数据的特征描述

3.1.1 局部特征描述

局部特征描述是一种由下到上的描述方式,是从特征点周围提取出有用的几何区域,并生成一个标识性的向量来代表这个区域的特征^[21]。局部特征不容易受环境噪声、物体遮挡或者人体运动变化的影响,对缩放、平移和旋转等操作也具有较好的稳定性。

局部特征的提取一般分为局部特征区域的检测和对局部特征区域描述两部分。局部特征区域检测一般针对由运动突变引起的时空兴趣点,因为在运动突变时产生的点包含了对人体动作进行分析的大部分信息。Laptev^[22]提出的 3Dharris 算子在 Moravec 角点算子的基础之上进行了严格的数学建模和改进,它能够检测运动目标在局部的时

空维度里同时产生显著变化的点, 将二维 Harris 角点检测扩展到了时空域。

而局部特征区域描述, 它的核心问题是可区分性和不变性(鲁棒性), 其描述子应该对背景的杂乱程度, 尺度和方向变化等均不敏感。局部特征区域描述中最常用的方法是 SIFT 描述符, 它利用图像局部的梯度方向, 为每个关键点分配一个或多个方向参数, 通过局部梯度的分布或者目标边缘方向来对局部目标的外观和形状进行特征描述。Wang^[23]比较了各种局部描述算子, 并发现在大多数情况下整合了梯度和光流信息的描述算子其效果最好。

3.1.2 全局特征描述

全局特征描述就是把识别目标当作一个整体进行描述^[24], 全面涵盖了人体信息, 代表着图像中的高层特征或语意。但由于容易受到识别目标定位的准确性以及背景剔除等图像预处理的影响, 全局特征描述也具有一定的局限性, 例如对噪声、遮挡以及摄像机视角变化等因素十分敏感。

Bobick 等人^[25]以两种方式把监测视频里运动目标的轮廓变化信息聚合成二维图像, 采用运动历史图(Motion History Image, MHI)来反映不同姿势在运动过程中存在的先后顺序。TASWEER 等^[26]利用归一化距离和关节流的特征来处理与某些动作最相关的注意关节, 并设计了一种新的基于骨骼的动作识别的注意关节图卷积神经网络, 在三个公共基准上达到了最先进的性能。

3.2 3D 骨骼数据的特征描述

3.2.1 基于关节的描述符

基于关节的描述符旨在建立身体关节位置之间的相关性。最简单的关联身体关节位置的方法就是考虑所有 3D 关节之间两两成对的距离。Masood 等人^[27]在文献中, 通过(1)当前帧中所有可能的关节对之间的距离; (2)当前帧和前一帧中的关节之间的距

离; (3)当前帧和中性姿势(通过平均所有动作序列的初始骨架计算)的关节之间的距离来表示身体姿势。每个单独的特征值通过 K-means 聚类为 5 个组中的一个, 并用二进制向量来表示每个聚类索引。但这种描述符缺少时间信息, 对动作的描述不够精确。

3.2.2 基于挖掘的描述符

基于挖掘的描述符指根据身体部位对动作的参与情况来区分动作类别。Wang 等人^[28]将人体分成五个身体部位, 采用空间域中的对比挖掘算法^[29]来检测身体部位的各种姿势, 这些身体部位的集合可以形成一个数据字典, 通过采用词袋模型(bag-of-words)方法, 动作序列被表示为检测到的子集的直方图, 并且采用 1 对 1 交叉核 SVM(1-vs-1 intersection kernel SVMs)来对序列进行分类。虽然每个人执行相同的动作会有自己的特性, 但是通常动作涉及的关节子集是类似的, 因此检测参与动作的关节子集可以帮助区分不同的动作类别。

3.2.3 基于动力学的描述符

基于动力学的描述符着重于将动作表示为关节三维轨迹的集合, 可以更加清晰直观地描述骨骼的特征。Zanfira 等人^[30]将动作表示为随着时间的推移身体关节位置的连续和可微函数, 可以在当前时间步的周围设置一个窗口, 在窗口中通过二阶泰勒变换, 在局部近似地逼近这个表示人体动作的函数。这样就可以通过当前的关节位置和人体关节的速度和加速度等微分特性来表征局部三维人体姿态。

4 动作识别的分类方法

通过以上步骤提取出特征之后, 就可以基于特征来构建分类器以进行动作的分类。分类算法是动作识别过程中最后也是最关键的一部分, 它根据特征向量进行训练, 从而给每一个被测对象进行不同类别的标记。

分类算法有很多,在这里分成两种:传统的方法和基于深度学习的方法进行介绍。传统的方法包括隐马尔可夫模型、支持向量机等;而基于深度学习的方法是目下使用最广、性能最先进的方法。

4.1 传统方法

隐马尔可夫模型(HMMs)是历史比较长、应用非常广泛的一种方法。HMMs 是一种与时序有关的,基于转移概率和传输概率的随机模型,系统当前所处状态的概率只与前一个时刻的状态有关,与其它历史状态条件无关。在 HMMs 的识别过程中,首先提取出特征向量序列,然后通过学习算法进行模型参数训练,最后对未知的运动序列进行识别分类。Yamato 等人^[31]首次用 HMMs 进行人体动作姿态的识别,之后又出现了 HMMs 的各种改进模型并成为人体动作识别的主流方法。Duong 等人^[32]提出了 S-HSMM(Switching Hidden-semi Markov Model)模型,S-HSMM 的识别性能高于 HHMMs,但是算法复杂度较高。

支持向量机(Support Vector Machine, SVM)是机器学习领域最常用的一种分类方法,它是一类按监督学习方式对数据进行二元分类的广义线性分类器,其决策边界是对学习样本求解的最大边距超平面^[33,34]。使用 SVM 进行姿态识别的文献很多,1995 年,Corinna Cortes 和 Vapnik 提出了软边距的非线性 SVM 并将其应用于手写字符识别问题,这份研究为 SVM 在各领域的应用提供了参考。Manzi 等人^[35]利用 X-means 算法提取关键位姿特征,经过自组织特征映射网络优化训练后采用多类 SVM 进行分类识别。该方法识别准确率高,但是不能识别数据集以外的未知动作。

4.2 基于深度学习的方法

近年来,动作识别的算法研究仍然在逐步地发展,尤其是基于深度学习的动作识别算法,使得动作识别的准确率大大提高。

卷积神经网络是深度学习的代表算法

之一,在动作识别领域也有广泛的应用。Mohamed 等人^[36]利用 RGB-D 相机提供的不同类型的视觉数据,分别采用 SVM 和 CNN 两种方法进行了对比。SVM 使用的是 3D 基础骨骼特征,CNN 使用的是 2D 深度图像。最后在 FLIC 数据集和 LSP 数据集上的实验结果表明,这两种方法性能差异不大且都具有较高的精度,在深度图像上应用时,基于 CNN 的方法更胜一筹。

而 Simonyan 等^[37]提出的双流网络(Two-Stream Network)采用两个分支的网络架构,分别捕捉视频的空间和时间信息。空域利用 RGB 图像作为输入提取外观特征,时域利用光流信息作为输入提取时序特征,并通过多任务训练的方法对两个行为识别数据集进行分类,去除过拟合,进而获得更好的效果。这是目前的基准之一,许多网络结构也是在此基础上进行的后续探索。双流结构的优点在于精度高,但它的速度比较慢。

综上所述,不同的动作识别算法由于自身算法结构的差异性以及所采用的特征描述不同,因此适用范围也有差别,不存在适用所有的分类问题的完美通用的算法,因此要使人体的动作识别效果达到相对较高的水平,根据不同的特征描述和适用范围选取合适的算法具有重要意义。

5 动作评价研究现状

动作评价是近些年才逐步发展起来的研究领域,目前尚没有研究者对其给出明确的概念定义和理论阐述。根据动作评价的目的和主要处理过程,本文认为:动作评价是指将输入的“学习者”数据在动作识别之后,与相应的“专家”数据进行对比,综合定量相似性和专家知识,对专业领域动作的规范性、流畅性、艺术性分析评价,并给“学习者”以反馈。

5.1 动作评价的特征描述

动作评价的整体流程和步骤与动作识

别有一些共通之处,比如对输入数据的预处理,对序列的特征描述等。但是针对动作评价的特征描述,是决定评价结果是否和人类的感受一致以及是否符合专家知识非常关键的一步,而且它和动作识别中的特征描述的侧重点也不太相同,因此,动作评价的特征描述是今后需要研究的重点,下面对已经出现的动作评价研究中关于特征描述的一些想法进行思考讨论。

动作评价的工作都会试图引入更独特的特征描述方法,目的是寻找一种更具各领域专业性的数学表达来表示身体关节特征。第一,对于不同的专业动作,每个身体关节在动作中起到的作用是不同的,那么就可以根据专家知识给各关节分配权重,这会给动作评价打下更好的基础;第二,可以将动作评价与大数据结合起来,这样一来,对每一个动作细节如何进行评价都有据可依,大大增加了动作评价的可靠性;第三,对于某些纠正性的研究,可以自定义规则并给所有可能的错误制定标签,评价动作执行的正确性。

Julia Richter 等^[38]提出了一种自动和动态加权的方法,基于动作参与程度,赋予不同骨骼关节不同的重要性,再加上基于动能的描述符采样,将自动分段和识别的动作实例馈送到动作评估组件,该组件将它们与相应的标准参考进行比较,判定它们的相似性。然后利用模糊逻辑提供语义反馈,便于用户更准确地执行操作。吕默等^[10]为了辅助裁判对健美操难度动作打分,设计了基于大数据和动作识别算法的健美操辅助评审系统。文中通过扩充动作数据建立了评分比对标准数据库。现有的数据库中记录的数据多为简单的动作,并不能满足健美操中高难度动作的识别和评价,故他们在 MSR Action3D 数据集的基础上,采集大量高水平运动员的标准动作,再结合国际权威标准制作出了比对数据库和评分比对体系。Fotini Patrona 等^[14]关于治疗性相关锻炼的研究也有一定参考价值。他们自己先定义特定于运动的规则,并确定适当的阈值。正确的运动练习动作带有类别标签 C,其余的类别标签 UB, FO, BK, WP 和 NBK 对应于自己规定的运动错误,这样就可以分析病人的

动作执行情况,给出评价和指导。

5.2 动作评价的评价标准

引言中提到过,动作评价的标准和动作识别的相似性是不能混为一谈的。动作识别判定相似性是通过定量数据,如关节角度、速度、加速度等,对比待识别动作和标准参考,从而将待识别动作分类。而动作评价则不仅要对比定量相似性,还要对比更高层次的规范性、流畅性、艺术性等,侧重于对各种专业领域的动作,结合专家知识的指导,提出独特的特征描述方法,再进行分析与评价,衡量的是一种专业相似性。因此,提出独特的评价标准是动作评价非常关键的一部分。

首先是定量数据(关节角度、速度、加速度等)的对比,它也是动作评价中不可或缺的一部分。陈学梅等^[8]所研制出的高尔夫挥杆评价系统,可以将训练者进行挥杆动作时的关节角度与标准挥杆动作进行对比,并给出关键动作处训练者的姿态与标准姿态的差异,辅助球员的练习。李奎^[9]的工作定义了标准羽毛球挥拍动作,并用非定长稠密轨迹算法对这些动作进行表征,然后以这些标准动作为基准,通过计算待分析动作与标准动作之间的切比雪夫距离来衡量他们之间的相似度,再运用评分公式对羽毛球挥拍动作进行分析评估。

然后是更深层面动作流畅性、艺术性等的比较。但结合了专家知识的动作评价研究目前非常稀少,很大程度是因为许多领域的专家知识是专家常年积累出的感性感受,是一种经验式的知识,而且“使用基于动作的特征面临的挑战之一,就是语义相似的动作可能不一定在数值上相似”^[39]。王台瑞等人在京剧教学中,基于 3D 动捕得到的数据分析了专业表演者与学习者动作的异同。京剧表演的内容、形式,演员的表情、招式都是按照一定的程式来呈现的,而非由演员随意发挥,如何从客观定量数据中评价演员对艺术性的表现,是一件很困难的工作。研究有 9 个受试者,其中有科班学生、戏曲学校学生(非科班)、有扎实的舞蹈基础的学生

以及其他普通学生,实验监测了每个人的速度、加速度、角速度,并让京剧专家分别对他们的表现评分。结果专家评分高的学生,骨骼数据与专家之间的相似性并不一定高,所以要把客观数据与艺术性等的评价很好地结合起来有着巨大的挑战性,是今后研究的重点内容。

6 总结与展望

6.1 研究难点

近年来,人体动作的识别和评价研究已经取得了非常重要的成果,展现出了良好的发展前景。但下述几个方面仍是今后研究的难点问题:

(1)特征选择的困难。特征向量的选择非常关键,直接影响到所采用的识别、评价方法以及识别、评价的性能。如果选取过多的特征,特征向量维数就会过大,相应地就会增加计算的复杂度,相反,如果选取的特征过少,又可能无法获取较高的识别准确率。因此必须在选取合适的特征和获取较高的识别、评价准确率之间进行折中权衡。

(2)专家知识应用的难度。进行动作评价时,特征描述必须基于各领域的专家知识。但专家知识专业性非常强,而且一般与计算机领域没有过多的联系,要想找到能够较好表征专业性动作的特征描述方法还是非常具有挑战性的。

6.2 未来发展趋势

根据本文的广泛调研,之前的大部分研究侧重于自动人体动作识别,并且已经取得了很不错的成果,而动作识别和动作评价虽然是两种不相同的工作,但动作评价有很多部分是基于动作识别的,他们之间有着千丝万缕的联系,关于未来发展趋势的展望如下:

1. 动作识别方面

(1)运用深度学习,一般性动作识别已经可以达到相当高的程度,彭伟等人^[40]首次基于神经架构搜索自动生成图卷积结构,

在 NTU RGB+D 和 Kinetics 数据集上准确性已经高达 95.7%,刷新了人体动作识别准确率。

(2)但对于专业性更强的复杂动作识别,其实当前的模型还是有着“力不从心”的问题:

(a)对一些密集、快速的运动,如健美操里的各种空翻,不一定能准确识别出每一次的动作;

(b)空间语义信息存在细微差别,如在健美操或者京剧表演中腿部姿态的些微不同可能代表着不同的语义,但目前并不能很好地识别出它们之间的不同。因此高层视觉的理解表示是一个巨大的难题。

2. 动作评价方面

(1)目前虽然有一些关于动作评价的文章,以及高尔夫、健美操、舞蹈、太极等方面的评价系统,但他们所分析评价的基本还是简单、规范、标准的动作。不过他们提出的一些对特征进行描述的思想也可以运用到之后的研究中,比如给各个关节分配不同的权重、考虑动作质量与运动学和动力学的联系等等。

(2)复杂动作评价还面临着如下挑战:

(a)如何提出更独特的特征描述方法来对专业动作进行评价,这将是今后研究的一个重点;

(b)怎么将人类的感知和量化的数据通过一定的规则进行转换,将专家知识更好地融入评价标准中,深层地理解动作的语义以更好地进行评价。

集前人的工作精华,有助于更好地实现动作识别和动作评价。未来研究可以着眼于深入研究动作的高层语义理解,从而实现真正的“智能”识别与评价。

参考文献:

- [1] S. Kwak, B. Han, J. Han. Scenario-based video event recognition by constraint flow[C]. Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Colorado Springs, 2011, 3345 - 3352.

- [2] U. Gaur, Y. Zhu, B. Song, A. Roy-Chowdhury, A string of feature graphs model for recognition of complex activities in natural videos[C]. Proceedings of International Conference on Computer Vision (ICCV), IEEE, Barcelona, Spain, 2011, 2595 - 2602.
- [3] S. Park, J. Aggarwal. Recognition of two-person interactions using a hierarchical Bayesian network[R]. First ACM SIGMM International Workshop on Video surveillance, ACM, Berkeley, California, 2003, 65 - 76.
- [4] I. Junejo, E. Dexter, I. Laptev, P. Pérez. View-independent action recognition from temporal self-similarities[J]. IEEE Trans. Pattern Anal. Mach. Intell, 2011, 33(1): 172-185.
- [5] Z. Duric, W. Gray, R. Heishman, F. Li, A. Rosenfeld, M. Schoelles, C. Schunn, H. Wechsler. Integrating perceptual and cognitive modeling for adaptive and intelligent human-computer interaction Proc[J]. IEEE, 2002(90), 1272-1289.
- [6] Y.-J. Chang, S.-F. Chen, J.-D. Huang. A Kinect-based system for physical rehabilitation: a pilot study for young adults with motor disabilities[J]. Res. Dev. Disabil, 2011, 32(6): 2566-2570.
- [7] H. Moon, R. Sharma, N. Jung. Method and system for measuring shopper response to products based on behavior and facial expression[P]. US Patent 8,219,438, July 10, 2012.
- [8] 陈学梅. 基于人体三维姿态的动作评价系统[D]. 浙江大学, 2018.
- [9] 李奎. 羽毛球运动员挥拍动作的捕捉、识别与分析[D]. 电子科技大学, 2017.
- [10] 吕默, 万连城. 基于大数据和动作识别算法的体育竞赛辅助评审系统设计[J/OL]. 电子设计工程, 2019.
- [11] I-Kao Chiang, Ian Spiro, Seungkyu Lee, Alyssa Lees, Jingchen Liu, Chris Bregler, Yanxi Liu. Dancing with Turks[R]. ACM Multimedia, 2015.
- [12] Jesse Scott, Robert Collins, Christopher Funk, Yanxi Liu. 4D Model-based Spatiotemporal Alignment of Scripted Taiji Quan Sequences[J]. ICCV Workshops 2017: 795-804.
- [13] 王台瑞. 以三维拮取探索戏曲动作教与学之异同[J]. 艺术教育研究, 2018, 35, 69-92.
- [14] Fotini Patrona, Anargyros Chatzitofis, Dimitrios Zarpalas, Petros Daras. Motion analysis: Action detection, recognition and evaluation based on motion capture data[J]. Pattern Recognition, 2018, 76, 612 - 622.
- [15] 田元, 李方迪. 基于深度信息的人体姿态识别研究综述[J]. 计算机工程与应用, 2020, 56(04): 1-8.
- [16] Diebel J, Thrun S. An application of Markov random fields to range sensing[C]. Advances in Neural Information Processing Systems, 2006: 291-298.
- [17] Park J, Kim H, Tai Y W, et al. High quality depth map upsampling for 3D-TOF cameras[C]. 2011 International Conference on Computer Vision, 2011: 1623-1630.
- [18] S. Nowozin, J. Shotton. Action Points: A Representation for Low-Latency Online Human Action Recognition[C]. Technical Report MSR-TR-2012-68, Microsoft Research Cambridge, UK, 2012.
- [19] A. Chan-Hon-Tong, C. Achard, L. Lucat. Simultaneous segmentation and classification of human actions in video streams using deeply optimized Hough transform[J]. Pattern

- Recognit, 2014, 47 (12):3807 - 3818 .
- [20] S. Asteriadis , A. Chatzitofis , D. Zarpalas , D.S. Alexiadis , P. Daras , Estimating human motion from multiple kinect sensors[C]. Proceedings of the 6th International Conference on Computer Vision/Computer Graphics Collaboration Techniques and Applications, ACM, 2013, 3 .
- [21] 唐灿, 唐亮贵, 刘波. 图像特征检测与匹配方法研究综述[J]. 南京信息工程大学学报(自然科学版). 2020, 03: 261-273.
- [22] Laptev I. On space-time interest points[J]. International Journal of Computer Vision, 2005, 64 (2/3): 107-123.
- [23] Wang, H., M. M. Ullah, et al.. Evaluation of local spatio-temporal features for action recognition[C]. British Machine Vision Conference, 2009.
- [24] Ojala T, Pietikäinen M, Mäenpää T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24 (7): 971-987.
- [25] Bobick A F, Davis J W. The recognition of human movement using temporal templates[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23 (3): 257-267.
- [26] TASWEER AHMAD , HUIYUN MAO, LUOJUN LIN , GUOZHI TANG. Action Recognition Using Attention-Joints Graph Convolutional Neural Networks[J]. IEEE ACCESS, 2019, 8:305-313.
- [27] S.Z. Masood, C. Ellis, M.F. Tappen, J.J. LaViola, R. Sukthankar. Exploring the trade-off between accuracy and observational latency in action recognition[J]. Int. J. Comput. Vis., 2013, 101 (3): 420-436.
- [28] C. Wang, Y. Wang, A.L. Yuille. An approach to pose-based action recognition[C]. Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Portland, Oregon, 2013: 915 - 922.
- [29] G. Dong, J. Li. Efficient mining of emerging patterns: discovering trends and differences[C]. Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, San Diego, CA, USA, 1999:43 - 52.
- [30] M. Zanfir, M. Leordeanu, C. Sminchisescu. The moving pose: an efficient 3d kinematics descriptor for low-latency action recognition and detection[C]. Proceedings of International Conference on Computer Vision (ICCV), IEEE, Sydney, Australia, 2013: 2752 - 2759.
- [31] Yamato J, Ohya J, Ishii K. Recognizing human action in time sequential images using Hidden Markov model[C]. In Proc CVPR, IEEE, 1992:379-385.
- [32] Duong T V, Bui H H, Phung D Q, et al. Activity recognition and abnormality detection with the switching hidden semi-Markov model[C]. In Proc CVPR. San Diego: IEEE Computer Society Press, 2005:838-845.
- [33] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016: 121-139, 298-300.
- [34] 李航. 统计学习方法[M]. 北京: 清华大学出版社, 2012: 95-135.
- [35] Manzi A, Cavallo F, Dario P. A 3D human posture approach for activity recognition based on depth camera[C]. European Conference on Computer

Vision.Cham: Springer, 2016: 432-447.

[36] Mohamed El, Ismail C, Wassim B, et al. Posture recognition using an RGB-D camera: exploring 3D body modeling and deep learning approaches[C]. 2018 IEEE Life Sciences Conference, 2018: 69-72.

[37] SIMONYAN K , ZISSERMAN A. Two - Stream Convolutional Networks for Action recognition in videos[J]. arXiv : 1406.2199.

[38] Julia Richter, Christian Wiede, Ulrich Heinkel and Gangolf Hirtz. Motion Evaluation of Therapy Exercises by Means of Skeleton Normalisation, Incremental Dynamic Time Warping and Machine Learning: A Comparison of a Rule-Based and a Machine-Learning-Based Approach[C]. Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 14th International Conference on Computer Vision Theory and Applications), 2019, 497-504.

[39] A. Yao, J. Gall, G. Fanelli, L. J. Van Gool. Does human action recognition benefit from pose estimation[C]. Proceedings of the British Machine Vision Conference (BMVC), vol. 3, BMVA Press, Dundee, UK, 2011(67):1-11.

[40] Wei Peng, Xiaopeng Hong, Haoyu Chen, Guoying Zhao. Learning Graph Convolutional Network for Skeleton-Based Human Action Recognition by Neural Searching[C]. AAAI, New York, USA , 2020: 2669-2676.