计算机科学与探索
Journal of Frontiers of Computer Science and Technology
1673-9418/2022/16(05)-0991-17
doi: 10.3778/j.issn.1673-9418.2110022

人体动作识别与评价——区别、联系及研究进展

杨　刚[1],张宇姝[1],宋　震[2+]
1. 北京林业大学 信息学院,北京 100083
2. 中央戏剧学院 传统戏剧数字化高精尖研究中心,北京 100710
+ 通信作者 E-mail: songzhen@zhongxi.cn

摘　要:人体动作识别与动作评价是近年来的热点研究问题。两者在数据类型、数据处理、特征描述等方面有许多相通之处。近年来,随着应用需求的显著增长,出现了大量有关动作识别与评价的研究工作,但两者间的区别与联系,以及它们的理论方法和技术路线还未见系统的分析与总结。从应用目的与技术特点等方面出发,探讨了两者的联系,给出了两者较为明确的概念界定。在此基础上,从数据处理流程的角度出发,将动作识别与动作评价归纳到一个统一的技术框架中;依据此框架,对动作识别与评价所涉及到的各个重要环节,包括数据类型、预处理、特征描述、分类方法、评价方法等的研究进展和存在的问题进行了系统阐述。其中,在分类方法环节,将当前动作识别的分类方法划分为基于统计模型的方法和基于深度学习的方法进行论述;而在评价方法环节,则以专家知识介入方式为依据,将当前的动作评价相关工作划分为四类并进行了系统梳理。最后对当前存在的瓶颈及未来研究重点进行了总结与展望。

关键词:动作识别;动作评价;特征描述;相似性度量
文献标志码:A　**中图分类号:**TP391.4

Human Action Recognition and Evaluation—Differences, Connections and Research Progress

YANG Gang[1], ZHANG Yushu[1], SONG Zhen[2+]
1. School of Information Science and Technology, Beijing Forestry University, Beijing 100083, China
2. Advanced Research Center for Digitalization of Traditional Drama, The Central Academy of Drama, Beijing 100710, China

Abstract: Human action recognition and action evaluation are hot research issues in recent years. Technologies of action recognition and action evaluation share similarities in terms of data sources, data pre-processing, feature description, etc. In conjunction with the significant recent growth of application requirements, numerous studies on action recognition and evaluation appear. However, the differences and connections between human action recognition and evaluation, as well as their theoretical methods and technical routes, have not been systematically analyzed and summarized. Starting from the perspective of application purpose and technical characteristics, the relationship between the two is discussed, and a clearer concept definition is given. On this basis, action recognition and action evaluation are summarized into a unified technical framework from the perspective of data processing flow. Based on this framework, the research progress and existing problems of all important links involved in motion recognition and evaluation, including data types, pre-processing, feature description, classification methods and evaluation methods, are systematically described. Among them, in the classification method section, the current

基金项目:北京高校卓越青年科学家计划项目(BJJWZYJH01201910048035)。
This work was supported by the Beijing Outstanding Young Scientist Program (BJJWZYJH01201910048035).
收稿日期:2021-10-13　**修回日期:**2022-01-06

action recognition classification methods are divided into statistical model-based methods and deep learning-based methods for discussion; and in the evaluation method section, based on the intervention method of expert knowledge, the current action evaluation related work is divided into four categories and systematically sorted out. Finally, the bottlenecks and the focus of future research are summarized and prospected.

Key words: action recognition; action evaluation; feature description; similarity measurement

人体动作识别和动作评价是当前的研究热点。动作识别是对输入的视频或3D动作数据进行分析处理,以判断不同动作分别属于哪种类别。动作识别技术在人机交互场景[1]、监控视频[2-5]、手势识别[6-8]、康复训练[9]、机器人[10]和行为理解[11-12]等各种行业都有着实际的运用价值。动作评价则是对特定动作的完成质量进行评判。它一般应用于体育、舞蹈、太极拳等专业领域之中,不仅可以辅助裁判、教练进行评分,更重要的是帮助人们进行动作分析与训练。

动作识别与动作评价的区别在于:动作识别其实是一种多分类性质的问题,它的侧重点是实现将输入的数据和作为参考的标准数据进行相似度的对比,然后为不同动作分配所属的类型标签;而动作评价则有更强的专业领域针对性,它必须与领域内的专家经验相结合,构建专业的评价标准,其不仅需要对比动作的外观相似性,还需要对动作的规范性、完成质量甚至艺术性进行评价,从而辅助人们对动作的深度分析。但同时,动作识别与动作评价也有紧密联系,二者在技术流程和方法上也有着很多共通之处。动作评价往往需要在动作识别的基础上完成。

早在20世纪70年代,Johansson[13]的移动光斑的运动感知实验,就证实了可以借助二维模型分析三维的人体运动信息,引发了很多研究人员对人体动作识别的研究兴趣,后续关于动作识别的研究工作大量涌现,并取得了显著成果。另一方面,有关动作评价的研究则还处于起步阶段,虽然有一些成功案例,例如高尔夫挥杆动作[14]、羽毛球挥拍动作[15]等体育运动中的动作,但所能处理的主要是单一且重复度高的动作。而对于更为复杂的动作,比如竞技健美操[16]、舞蹈[17]、24式太极[18]、戏曲[19]等则力不从心。对于这些复杂动作,不应该只是单纯地比较“外观相似度”,还需要在更深层次的“专业相似度”上有所突破。

经过充分而深入的调研,论述了动作识别与动作评价存在的区别与联系,并从完整的数据处理流程的角度出发,归纳了动作识别与动作评价的技术框架。围绕这一框架,从数据类型、预处理、特征描述、识别方法、评价方法等各个环节分析、总结了经典方法以及最新研究进展,并将其按照技术特点分类。最后探讨了当前研究所面临的关键问题及未来发展趋势。

1 相关工作及技术框架概述

动作识别是计算机视觉领域一个重要的研究课题,人们已经开展了大量的研究,并且已经出现了一些相关的综述论文。徐光祐等人[20]主要从视觉处理的角度来分析动作识别,从动作的定义、特征提取和动作表示、动作理解的推理方法三方面对动作识别进行了综述。Wu等人[21]则将重点放在了深度学习上,综述了各种最新的基于深度学习的技术,用于三种类型的数据集:单视点、多视点和RGB-D视频上进行人体动作识别。Presti等人[22]则总结了基于3D骨骼的动作识别的技术和方法,侧重于分析数据预处理、公开可用的3D数据集和精度度量标准等方面,此外他们还提出了基于骨骼的动作特征描述的分类。

上面这些综述工作各有其侧重点,或者聚焦于视觉处理的关键问题,或者聚焦于骨骼数据识别方法,或者聚焦于深度学习方法。而本文的思路与这些综述不同,是从整体的数据处理流程的角度出发进行关键模块的梳理,并将动作识别与动作评价两类问题归纳到了一个统一的技术框架中(图1)。如图1所示,动作识别与动作评价这两类问题既有相同的部分,也有各自独特的部分。其中,数据类型、数据预处理、特征描述三部分是动作识别与动作评价的共通之处,它们对动作识别和评价都有基础意义;而在随后的方法部分,则由于应用需求和研究目标之不同,动作识别与动作评价有显著差异。本文即依据此技术框架对各个模块进行系统的介绍与分析。

值得一提的是,目前尚无动作评价相关的综述,本文首次将这一问题进行了比较系统的介绍和讨论,可以为希望从事相关研究的人员提供一定的参考。

有很多与动作识别有密切关系的概念和技术,如人体姿态估计(human pose estimation)、动作检测(action detection)、行为识别(activity recognition)等。姿态估计是将图像和视频中存在的人物肢体检测出

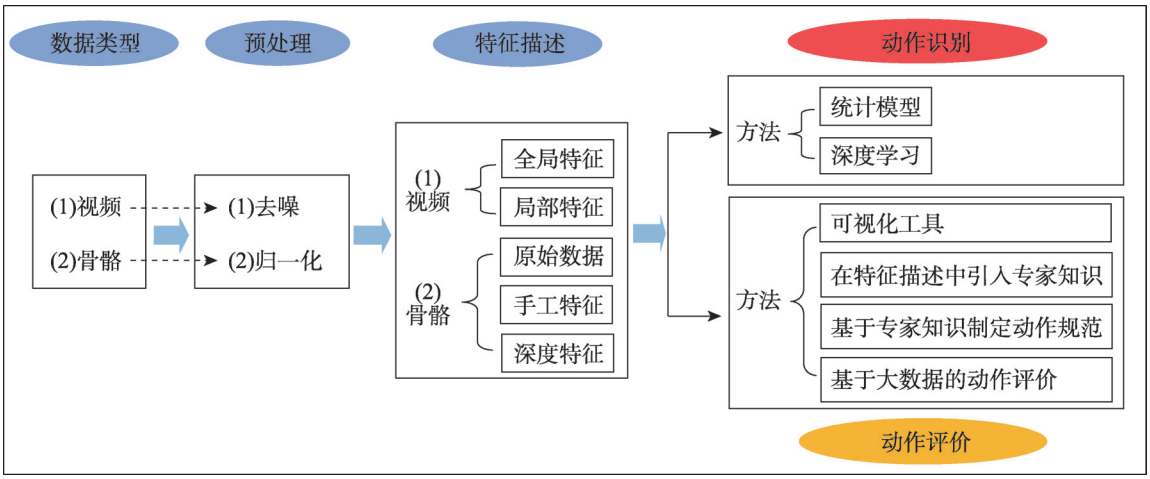


图1　动作识别与动作评价的技术框架图

Fig.1　Technical framework of action recognition and action evaluation

来的技术。姿态估计不仅要检测，还要进一步重建人的肢体和关节，它得到的是重建出的人体关节向量，而不是类别的标签。与之不同，动作识别的目的就是要得到动作的类别标签。姿态估计与动作识别之间有密切联系，很多动作识别算法就是在姿态估计基础上进行特征提取与分类。文献[23-24]等对姿态估计进行了系统介绍，而本文的重点则不放在姿态估计上，而是放在了动作的特征描述与识别、评价上。

动作检测是指从视频中定位出发生特定动作的视频段，并将其分类。标记出目标动作的边界后，再对这种“已修剪”(trimmed)的动作序列进行识别。本文主要讨论的是对已修剪的动作序列进行动作识别，而并不讨论动作边界检测问题。

行为识别与动作识别的区别在于动作(action)比行为(activity)的粒度更细。可以认为一个动作仅包含单人的简单行为；而行为是由一系列动作组成，并可能包含人-人或人-物间的互动。显然，动作识别与行为识别的研究是有交叉的，一些行为识别方法正是基于动作识别技术进行计算的。而相对于动作识别，行为识别更为关注对较长时间内复杂行为序列的理解。本文重点放在动作识别的相关研究上，主要关注单人在较短时间内单位动作的分类与评价。

为了不偏离本文的讨论框架，聚焦于动作识别与动作评价的关键问题，后文将不再对姿态估计、动作检测、行为识别等内容展开叙述。

2　数据类型

当前动作识别与动作评价所处理的动作数据源主要分为两种：视频数据和骨骼数据。数据源的类型不同，则后期的预处理和特征描述等环节将会有显著差别。

2.1　视频数据

视频动作数据是动作识别与评价任务中最常用的一种数据，它是利用相机拍摄的动作视频序列，由于其每帧画面都是由RGB三通道形成的图像，故而也被称为RGB数据。基于视频数据的动作识别方法主要有两种思路：

(1)基于视频数据的直接识别。即直接从视频画面中提取动作序列的时域以及空域特征并进行分类。

(2)先提取骨骼信息再识别。即首先从视频中提取(2D或3D)骨骼信息(如前面第1章所述，这个过程被称为姿态估计)，再进行分类。

近年来，深度摄像头获得了很大发展，利用深度摄像头可使得获取的视频信息中含有场景的深度信息(被称为RGB+D数据)。利用增加的深度信息，姿态估计往往可以取得更好的效果，从而有利于后续的动作识别。

随着设备的进步，视频数据的获取越来越便捷和普遍，这使得基于视频数据的动作识别具有广阔的应用空间，相关工作层出不穷[25]。但采集视频数据时不可避免地会产生遮挡、抖动、明暗变化等噪声，这也为其带来了挑战。

2.2　骨骼数据

骨骼数据出现的时间较晚，相比于视频数据，它可以更加直接地表示身体各部位的运动特征，如关节角度、速度等，从而可以更方便、准确地进行动作识别[26]，因此它成为了近年来人们关注的焦点。它是通过关键点来描述整个人体动作的数据模式，这些

关键点往往依据人体骨骼关节来确定，故而被称为骨骼数据。图2是一种典型的关键点布局图，其中黑色点为骨骼关节点，红色点则用来标识身体主要部位。在动作计算过程中，人们普遍会将模型的盆骨位置作为“根骨骼”，基于根骨骼进行递推，就能得到其他骨骼的相对位置。

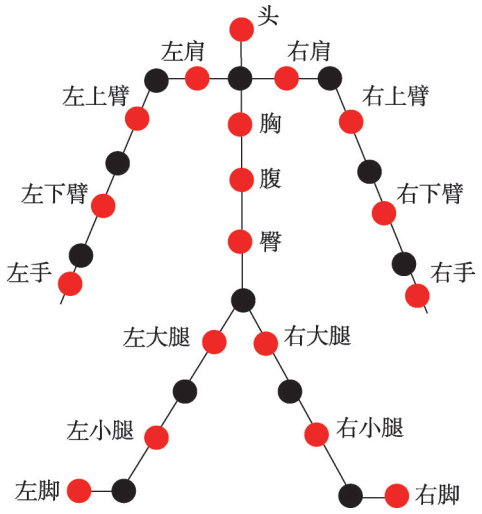


图2 人体骨骼示例图

Fig.2 Sketch map of human skeleton

按照维度不同，可以把骨骼数据分为两类：

（1）2D骨骼。一般是利用姿态估计算法从视频中识别获得的2D骨骼数据。

（2）3D骨骼。根据获取设备或者原始模态的不同，又可以分为两类：①通过光学和惯性动作捕捉设备直接捕捉的人体动作3D骨骼数据。②从视频中提取出2D骨骼数据，再重建为3D骨骼数据。

根据调研，动作评价相关研究工作多数使用的是骨骼数据，因为它更关注动作本身完成的质量；而动作识别问题中使用视频以及骨骼数据的工作都很丰富。骨骼数据相比视频数据优点在于，它包含的信息密集而精炼。但是也有其局限之处：(1)正是由于骨骼数据的冗余较少，对噪声极其敏感，容易影响动作识别和评价的性能。(2)从全局来看，骨骼数据的整体信息量比视频数据少。因为视频数据还会包括环境、物体等，而它们并不存在骨骼信息，因此骨骼信息对于人与物交互的动作的识别不具有优势。

2.3 数据集

目前已经有很多针对动作识别的公开数据集供研究人员使用，表1提到了一些常用的数据集，并列出各数据集的类别数、样本量、数据模态以及数据集内容。

这些数据集都提供了动作类别的标注。从数据规模来看，采集年代较早的数据集，比如UCF101、HMDB51、MSR Action 3D等，一般来说规模普遍较小，场景相对简单，而且视频的分辨率也偏低。但它们应用广泛，在很多研究中被作为基准来使用。而2016年之后的数据集，规模显著增大，比如Activity-Net、NTU RGB+D、NTU RGB+D 120等，它们具有更丰富的类别和更大的数据量。尤其是YouTube-8M，提供了多达4 716的动作类别和800万的样本。这也反映出随着研究的发展，人们能够建立更为复杂的

表1 常用的公开动作识别数据集

Table 1 Commonly used publicly available action recognition datasets

数据集	年份	类别数	样本量	数据模态	主要内容
UCF101[27]	2012	101	13 320	RGB	BBC/ESPN广播电视频道和YouTube的一系列数据集
HMDB51[28]	2011	51	6 766	RGB	从各种互联网资源和数字视频中收集的人类日常行为
YouTube-8M	2016	4 716	8 000 000	RGB	包含800万个YouTube视频，提供视频级别的注释并标记为4 800个知识图谱实体
MuHAVi[29]	2010	17	1 904	RGB	人类动作视频数据，包括手动标注的轮廓数据
ActivityNet	2016	200	20 000	RGB	涵盖了200种不同的日常活动，共计约700 h的视频，平均每个视频上有1.5个动作标注
MSR Action 3D[30]	2010	20	567	RGB+D&skeleton	记录了20个动作，10个主体，每个对象执行每个动作2~3次，共有567个深度图序列，分辨率为640×240像素
NTU RGB+D[31]	2016	60	56 000	RGB&RGB+D&skeleton	主要分为3类：(1)日常行为；(2)医疗卫生相关行为；(3)两人互动
NTU RGB+D 120[32]	2019	120	114 480	RGB&RGB+D&skeleton	主要分为3类：(1)日常行为；(2)医疗卫生相关行为；(3)两人互动
G3D[33]	2012	20	10	RGB&RGB+D&skeleton	包含一系列使用Microsoft Kinect捕获的游戏动作

预测模型，需要更大量的训练数据，同时能够处理更为复杂的任务。此时基于深度学习的动作识别方法逐渐成为主流，这些大规模的数据集也为深度学习在动作识别中的应用提供了有力的支持。

从数据模态来看，表1中给出的所有数据集都提供了RGB数据。RGB视频数据是动作识别领域最常用的数据源，相关的研究工作数量也最多。表1中后4个数据集除了RGB数据外，还提供了深度数据以及骨骼数据。利用这些数据集可以更好地检测出三维骨骼，从而为基于3D骨骼的动作识别与评价提供更好的支撑。

从数据内容来看，有些数据集主要是从公共资源中获得的视频数据，如UCF101、HMDB51、YouTube-8M、ActivityNet等，它们所包含的动作类别主要是各种各样的人类日常活动；这些数据集一般只包含RGB视频。还有一些数据集，如MSR Action 3D、NTU RGB+D、G3D等，则是专门录制或捕获得到的，它们往往包含了针对某些特定应用领域的动作（如医疗卫生、游戏动作等），并具有深度数据和骨骼数据，适用于进行更有针对性的动作识别与评价。

3　数据预处理

在动作识别和动作评价之前，首先要对数据进行预处理。研究中尽管有一些常用的预处理方法，但实际上并没有统一的标准，而且由于任务的不同，所需要的处理方式也有较大差异。一般来说，对于视频数据，预处理的主要任务是去噪；而对于3D骨骼数据，预处理的主要任务则是归一化。

（1）视频数据的去噪

对于动作捕捉设备采集的3D骨骼数据，一般不需要进行去噪处理，因为它很少受环境影响，噪声很小。然而，视频数据必须去噪，因为拍摄过程中受外界不确定因素的影响，原始数据中包含很多不稳定或干扰信息。视频的去噪基于图像去噪技术，但相比于图像多了一个时序维度。

该领域的经典方法是BM3D（block matching 3D）[34]算法，该算法先计算相似性来定位与当前待处理的块相似的二维图像块，然后按照一定的规则将它们堆叠成三维组，最后通过滤波实现降噪。BM3D以及由此延伸出的方法是图像去噪领域公认的效果最好的方法，直到如今对后续研究都有着指导意义。Maggioni等人[35]提出的VBM4D（video block matching 4D）方法即将BM3D方法从图像扩展到时域，从而转变为对视频的去噪。它把连续动作前后帧形成的区域称作补丁（patch），寻找当前待处理补丁的相似补丁，之后通过两种滤波处理并取加权平均，来实现去噪。

这种基于补丁的视频去噪方法（patch-based method）成为传统的主流思路。但近年来，随着深度学习的发展，研究者们开始尝试基于神经网络进行视频去噪。最早用于视频去噪的神经网络方法是递归神经网络，但它只能对灰度图像进行处理而且效果一般，随后出现的VNLnet（non-local video denoising by CNN）[36]、VNLB（video denoising via empirical Bayesian estimation of space-time patches）[37]和DVDNet（fast network for deep video denoising）[38]等算法大大增强了去噪效果。但是已经出现的基于神经网络的视频去噪方法尚无法与最好的patch-based的方法竞争。不过最近，Tassano等人[39]提出了一种最新的基于卷积神经网络结构的视频去噪算法，达到了可与当前最好算法比拟的效果，同时具有更低的计算负载，这表明深度学习方法在视频去噪领域有进一步发展的潜力。

（2）3D骨骼数据的归一化

对于3D数据来说，不同人体的骨骼尺寸及骨骼比例都不相同，在对骨骼数据进行比较、匹配时，需要首先对骨骼数据进行转换，使不同的骨骼具有相同的比例或尺度。这种处理被称为骨骼数据的归一化。比如，Ping等人[40]将四肢和肩膀作为基准，来使人体骨骼标准化；Wu等人[41]以髋关节为原点，进行对齐和比较；Wang等人[42]则是选择头部位置为原点对齐。归一化不是简单地同比例缩放，而是根据各自不同的方法需求实施适宜的归一化策略。总结来看，3D骨骼数据的归一化一般首先选定基准点进行位置的对齐，然后需要选定基准长度进行关节长度的归一。不过，人体各关节的长度比例存在个体差异，这种个体差异有可能会对后面的动作评价产生影响，是否应当在归一化阶段将所有人体归一化到相同的长度比例，这还是一个待探讨的问题。

4　特征描述的方法

特征描述是指将原始动作序列数据构建成具有显著物理或统计意义的特征，提炼出的特征通常被称为特征描述符。可以说，选择合适的特征描述符是动作识别的关键[24]。而动作评价问题则在此基础上需要进一步将专家知识引入特征描述中，以达到评价目的。

视频数据与3D骨骼数据的数据结构、信息模式

差别很大,导致其特征描述方式也显著不同。综合分析当前的特征描述相关工作,对视频数据和骨骼数据分别进行归类与总结。对于视频数据,从特征区域的角度出发,将其特征描述划分为全局描述和局部描述两大类。对于3D骨骼数据,则从特征抽取手段的角度,将其特征描述划分为三类:(1)原始数据(角度、坐标等);(2)手工特征;(3)深度特征[43]。下面具体地介绍各种典型的特征描述方法。

4.1 视频数据的特征描述

全局特征描述将要识别或评价的目标作为一个整体来考虑[44],其覆盖人体姿态的全部信息;而局部特征描述则是在选定的特征点周围划分出一块局部几何区域,然后生成一个能够表示这块区域特征的标识性向量[45]。

4.1.1 全局特征描述

常见的全局特征有颜色特征、纹理特征和形状特征等。Bobick等人[46]最早采用轮廓和能量来描述人体的运动信息,提出运动能量图(motion energy image,MEI)和运动历史图(motion history image,MHI)两个模板结合起来表示对应的一个动作信息。方向梯度直方图(histogram of oriented gradient,HOG)是另一种非常经典的全局图像特征描述方式。Dalal等人[47]首先使用HOG进行行人识别,并取得了很好的效果。后来的很多研究工作都是基于HOG来进行的。

全局特征描述具有稳定性好、简洁直观等优点,但它也有一些缺点,比如容易受到背景负面影响、计算量大等。

4.1.2 局部特征描述

局部特征是从局部区域中抽取的特征,包含边缘、角点、曲线等类别。一般来说,局部特征的提取分为局部特征区域检测和对局部特征区域描述两部分[24]。文献[24]认为局部特征区域检测是为了找出能标识动作信息的特征点,并将其称作“时空兴趣点”。人们发现人体动作特征往往反映在突变状态时,因此这些兴趣点通常在运动发生突变时产生的点中选取。角点检测是最早提出的特征点检测之一。Moravec角点检测算法把那些与周围像素的特征都有很大差异的像素,认为是“角”[45],这就属于发生了突变的点。Laptev[48]提出的3D Harris算子对Moravec算子进行了改进,将2D Harris角点检测扩展到了时序和空序中,能够捕捉到运动目标同时在局部的时空域里,都产生了剧变的点。

在检测得到特征区域后,即可对局部特征区域进行描述。常用的特征包括梯度和光流信息等。Laptev等人在文献[49]中使用了局部梯度直方图(HOG)和光流直方图(histograms of oriented optical flow,HOF),将本是全局特征的描述方法转换为局部特征描述。Wang等人[50]则是将各种局部描述符进行了总结和比较,他们认为,描述效果最好的是同时采用了梯度和光流信息的方法。

与全局特征比起来,局部特征的优点是可获得的数量丰富,特征之间的相互约束弱,因此受遮挡影响小、稳定性高。相对地,它涵盖的范围不够全面,可能漏掉重要信息。

4.2 3D骨骼数据的特征描述

4.2.1 原始数据

3D骨骼数据由关键点的三维信息组成,所谓原始数据特征是指将这些关键点本身的一些属性,比如坐标、角度、变化速率等作为动作特征。它们通常可以表示为绝对[51]或相对[52]的关节坐标向量。使用原始骨架数据特征非常直接,但其对动作语义特征的表达不足,并且数据量过大,因此除了用于基线评估外很少被使用[43]。

4.2.2 手工特征

手工特征(hand-crafted features)是指在原始数据基础上,通过描述关节间的某些关系,人为定义的一些特征。这些手工特征经常会利用不同关节间的相对旋转和平移等信息[43]。Masood等人[53]通过测量关节对之间的距离来表示身体姿势。Müller等人[54]则利用布尔特征来表达身体几何关系,通过描述不同身体部位之间的几何关系来表示人体骨架,可以使得对特征的描述不受骨骼大小的影响。不过,目前这些手工特征都没有考虑时域信息,对动作的描述不够充分。而且手工特征的另一个问题是,不同领域提取的手工特征往往具有特殊性,在另一个领域的数据上可能无法适用,使得基于此特征的动作识别算法难以推广应用。

4.2.3 深度特征

手工特征的发展逐渐进入瓶颈,而深度学习的发展为动作数据的特征提取带来了新的可能。深度神经网络能够从复杂数据中自动学习出特征,从而可用于动作识别。近年来,人们使用RNN(recurrent neural network)、CNN(convolutional neural network)[55]和GCN(graph convolutional network)[56]等开展了骨骼数据的特征描述工作[57]。

RNN将数据相邻时刻整合成递归结构,因此它很

适合描述动态数据。文献[58]提出在RNN网络中加入注意力机制，使之成为EleAttG（elementwise-attention gate）结构，给输入数据里不同元素赋予不同的重要程度，并将之用于动作识别。作者在NTU RGB+D数据集中的骨骼以及视频数据都进行了测试，对骨骼数据的识别率由基线方法的75.2%提升到了80.7%，对视频数据的识别率由基线方法的81.5%提升到了88.4%，结果表明加入这个模块后RNN的性能得到了极大的提升。

以前CNN通常被用于图像处理，它学习、描述高层语义的能力十分强大[57]，将其作为一种骨骼特征提取方式可以极大提高识别效率。但图像问题与时序无关，因此基于CNN的方法进行骨骼特征的描述，并用于动作识别，必须思考如何更好地加入时域信息。

GCN将CNN拓展到了任意结构的图（graphs）结构上来，并且在诸如图像分类、半监督学习任务中得到了广泛的应用[56]，但之前尚未有人将GCN应用于人体骨骼序列的特征描述中。最近，Yan等人[56]提出的时空图卷积网络模型（spatial temporal graph convolutional networks，ST-GCN）首次使用GCN方法对骨骼信息进行时空特征描述。一般来说，骨骼信息只包含各个关节点坐标和它们的连线。而该方法将骨骼序列作为输入，将人体骨骼作为图结构进行描述，即由关节点、关节间连线以及时序上对应的关节点连成的虚拟的“时间边”组成。该方法在NTU-RGB+D数据集上，将当时的最高识别率提高了近4个百分点，效果显著。

深度特征的优势是无需手工参与，而能提取到较高层次的特征；而且，借助于大量的训练数据，深度特征受光照、姿态等影响较小。不过，深度特征的提取类似于一种“黑盒”计算模式，无法得到其显式的特征表达方式。

上面介绍了针对不同数据的多种类别的特征描述方式。目前并没有特别主流的、占主导优势的特征描述方法，不同特征描述最终能达到的效果与数据集特点、要识别的目标以及所采用的动作识别方法等都有很大关系。往往需要根据所处理的动作对象特点进行有针对性的特征描述。这一点对于“动作评价”而言更为重要，必须根据所评价的对象，增加专家知识，制定有针对性的特征。

5　动作识别的分类方法

在明确了动作数据的特征描述之后，即可进行动作识别或动作评价工作，本章介绍动作识别相关方法。动作识别的下一步就是构建分类器进行动作的分类。分类算法是动作识别过程中最后，同时也是最关键的一部分，它依据特征向量进行训练，从而输出每一个识别对象的类别标签[24]。至今已经出现了很多有关动作的分类算法，本章将它们分成两大类进行介绍：基于统计模型的方法和基于深度学习的方法。基于统计模型的方法包括隐马尔可夫模型、动态贝叶斯网络、支持向量机、模板匹配等；而基于深度学习的方法则是目前的主流方法，这里介绍当前三类主流的动作识别深度学习框架。

5.1　基于统计模型的方法

5.1.1　模板匹配法

动作识别中最简单、直接的方法是模板匹配法，这种方法首先将一些人体动作作为模板库[23]，然后计算待识别的动作与模板之间的相似度，如达到某阈值即可判定为此动作类型。用于动作识别的典型模板有ASM（active shape models）[59]、AAM（active appearance models）[60]、MHI[61]、MEI[46]等，它们采取的有形状、外观、历史图、能量图等各种特征模态。模板匹配法有着思想容易理解，模板设计复杂度低的优势，但也存在着易受噪声和持续的动作变化影响，鲁棒性不强，识别准确度不高的缺陷[23]。

5.1.2　状态空间法

该方法将每个动作定义为一个状态，通过概率来描述状态和状态之间的转移，因此一个动作序列可以表示为一系列状态的转移过程。典型的状态空间模型有隐马尔可夫模型（hidden Markov models，HMMs）和动态贝叶斯网络（dynamic Bayesian network，DBN）。

经典的隐马尔可夫模型（HMMs）是一种基于时序、转移概率和传输概率的随机模型[23]。在确定了特征向量之后，根据训练的模型参数获得状态序列，然后进行动作的分类。HMMs模型最早是一种数学统计概念，而Yamato等人[62]首先将其用于动作识别，经过几十年的发展，已经在语音识别、故障诊断和动作识别等领域成功实现应用，甚至成为了人体动作识别的主流方法之一。在这之后又出现了HMMs的各种改进模型，比如Nguyen等人[63]提出的分层隐马尔可夫模型（hierarchical hidden Markov models，HHMMs）。作者使用该模型，依据运动轨迹学习和识别动作，取得了良好的效果。近年仍然有人在改进HMMs模型，梅雪等人[64]提出了一种基于多尺度特征的双层隐马尔可夫模型，在双层HMMs模型中添加运动轨迹和人体姿态边缘小波矩，提供更为丰富的层次信

息。仿真实验的结果证明,此模型达到了很高的识别准确率。

动态贝叶斯网络(DBN)是一种考虑了相邻变量转化的贝叶斯网络,它的框架简洁合理,逻辑关系更加清晰、更易于理解。相比HMMs,DBN的表达能力更强,因此DBN对于需要多信息交叉融合的场景识别效果更佳[23]。HMMs模型需要巨大的训练样本量,而DBN因为自身的结构的优势,训练复杂度要低很多,但正因如此,它比HMMs的设计复杂度要高。Du等人[65]提出了一种新的带有状态持续时间的DBN模型结构,将全局特征和局部特征协调地结合起来,以模拟人类的交互活动,达到了很好的效果。Oliver等人[66]则探讨了使用DBN模型进行动作识别时的几个重要问题:(1)观测到变量的可能性;(2)数据是否存在内在的联系;(3)进行实际应用的复杂程度。

5.1.3 支持向量机法

支持向量机(support vector machine,SVM)是一种经典的机器学习分类方法,它是一种广义的线性分类器,通过监督学习的方式进行数据的二元分类[67-68]。但SVM使用的是一对一识别策略,将其应用在动作识别上,输出结果需要经历多次筛选,会降低识别效率。为了提高识别性能,相关研究都致力于找到更好的方法来表示关节特征。比如,Pontil等人[69]使用SVM在高维空间上处理图像的像素点,以此来进行动作识别。Manzi等人[70]采用X-means方法进行特征描述,最后运用SVM进行动作分类。Schuldt等人[71]则是将时域和空域特征结合起来,使用SVM方法,对动作进行局部表征,最后实现动作识别。

5.2 基于深度学习的方法

近年来,一些研究者将深度学习方法应用于动作识别,使得动作识别的准确率有了显著提升。目前,深度学习方法已成为动作识别研究中的主流方法。下面介绍三种最典型的用于动作识别的深度学习算法框架:CNN、双流网络框架(two-stream network)以及融合CNN-LSTM(convolutional neural network-long short term memory network)结构。

4.2.3小节提到了用CNN进行动作特征的描述。显然,在采用CNN进行特征描述的基础上,可以进一步完成动作识别任务。Mohamed等人[72]将SVM和CNN两种方法进行了比较,用它们来处理RGB-D相机采集到的同一套但数据类型不同的数据。SVM处理3D骨骼数据,而CNN则是处理2D深度图数据。实验发现,这两种方法性能相差不多,但CNN方法在深度图像上的效果更佳。

4.2.3小节介绍的Yan等人[56]提出的ST-GCN模型,能够更好表示人体重要关节之间的空间关系和时序关系,从而可以用于3D骨骼数据的动作识别。在此基础上,刘锁兰等人[73]提出了一种ST-GCN方法的新型分区策略,相比于之前的工作加强了骨骼关节点信息在时间和空间上的联系,然后通过迭代学习率进一步提升识别精度的目的。结果在Kinetics和NTU RGB+D数据集上比现有方法识别效果均有显著提高。

运用CNN进行动作识别取得了不错的效果,不过当前方法在投入应用时存在的问题在于:很多方法都对应用场景进行了一些实际生活中难以满足的假设,比如视角或背景固定不变、无遮挡等。针对这个问题,Ji等人[74]提出了一个新的用于运动识别的3D CNN模型。该模型从连续视频帧中产生多通道的信息,然后在每一个通道都分离地进行卷积和下采样操作,最后将所有通道的信息组合起来得到最终的特征描述。而李元祥等人[75]提出一种基于深度运动图(depth motion maps,DMMs)和密集轨迹的人体动作识别算法。作者利用CNN训练DMMs数据并提取高层特征作为静态特征描述符,使用密集轨迹作为动态特征描述符,最后整合静态和动态特征作为整体特征描述符,取得了良好的识别结果。这两种模型都通过计算高层运动特征来增强特征提取能力,并综合了多种特征去判断识别结果,因此可适用于各种不同环境,一定程度上解决了对场景要求比较严苛的问题。不过,多通道特征的学习和融合也在一定程度上增大了计算复杂度,降低了识别效率。

双流网络框架通过模仿人体视觉形成过程,来理解视频信息,以达到更好的视频内容理解能力。双流网络将分类任务分成两个模块,一个处理图像RGB信息,另一个处理光流信息,然后联合训练CNN模型,融合两个网络的训练结果,得到动作的类别。Simonyan等人[76]最先使用了双流网络进行动作识别,他们的方法后来成为相关研究的基准之一。Feichtenhofer等人[77]在双流网络结构的基础上,改进了融合空域和时域的方法,以便更好地理解双流框架中的时空信息。

双流网络的应用使得动作识别精度上了一个台阶。然而,双流网络也存在一定的问题,比如它不会专门分辨不同通道的差异性,不能很好区分冗余帧和背景等信息,而减弱了其整体特征表达能力。石祥滨等人[78]提出了一种基于双流时空注意力机制的端到

端的动作识别方法（end-to-end action recognition model based on two-stream network with spatio-temporal attention mechanism，T-STAM）。首先在双流结构中加入通道注意力来校准包含的信息，然后设计时间注意力模型和多空间注意力模型来对关键帧上的动作显著区域进行重点关注。实验表明，该方法在数据集UCF101和HMDB51上比近年来提出的其他先进方法，取得了更高的精度。这也说明，有效地区分不同通道特征，将注意力集中在关键时空信息上，能够进一步提高双流网络的效率。

动作识别问题最重要的任务之一就是对时域维度的处理。如果能很好地处理时域信息，识别效果一定会显著提升。而RNN能够很好表达时序特征，适于处理动态动作序列。在各种RNN模型中，LSTM性能优异，可以完整地学习序列的空域和时域特征。Donahue等人[79]将CNN与LSTM相结合来提取视频数据中的时空信息。该CNN-LSTM框架首先基于CNN来提取每帧图像的特征，之后用LSTM挖掘特征之间的时序关系来完成动作识别，这种方法不仅精度高，速度也快。大多数之前的动作识别方法，如卷积神经网络、双流网络，使用的特征仅包含全局时域信息，而忽略了局部时序特征。为了解决这个问题，杨珂等人[80]提出了一种基于时序交互感知模块的长短时序关注网络（long and short sequence concerned networks，LSCN），通过融合时序信息，利用不同卷积层时序特征的交互加强，来表示不同时长的动作，在长动作和短动作的识别上均有很好的效果。实验结果证明，此方法在UCF101和HMDB51两个公共数据集上，比基础的方法在精度上分别有0.4个百分点和2.9个百分点的提升。

5.3　动作识别方法总结

由以上论述可见，不同的动作识别方法的算法结构及所采用的特征描述各有不同，导致其适用范围各有差别，并不存在可以解决所有的分类问题的完美算法。表2列举了以上提到的各种动作分类方法，并总结了它们的优缺点。表中重点比较了各种

表2　动作分类方法总结

Table 2　Summary of action classification methods

类别	方法分类	相关工作与方法		优缺点
基于统计模型的方法	模板匹配法	ASM、AAM、MHI、MEI 基于二维网格模板特征的匹配方法 DTW（动态时间规整算法）		实现简单，计算复杂度低，但精度低，鲁棒性差
	状态空间法	HMMs	HMMs HHMMs S-HSMM 基于多尺度特征的双层隐马尔可夫模型	精度较高，但鲁棒性差，计算复杂度高
		DBN	Du等人[65] Oliver等人[66]	精度较高，计算复杂度较低，但设计复杂度高，鲁棒性差
	支持向量机法	Pontil等人[69] Manzi等人[70] Schuldt等人[71]		精度高，设计复杂度低，但鲁棒性差，对大规模训练样本难以实施
基于深度学习的方法	CNN	Mohamed等人[72] ST-GCN[56] 基于ST-GCN方法的新型分区策略[73] Ji等人[74] P3D、T3D、R3D 基于深度运动图和密集轨迹的动作识别算法[75]		精度非常高，鲁棒性强，处理高维数据能力强，但计算复杂度高，需要调参数
	双流网络	Simonyan等人[76] Feichtenhofer等人[77] T-STAM [78] Two stream fusion（双流融合） 3D ResNet		精度非常高，鲁棒性强，但计算复杂度高，速度慢
	CNN-LSTM结构	Donahue等人[79] LRCN Unsupervised+LSTM（无监督的LSTM模型） LSCN[80]		精度非常高，鲁棒性强，且计算速度快

方法在精度、鲁棒性、计算复杂度、计算速度等方面的表现，同时也保留了不同方法的一些其他特点，简洁明了地展现了各种动作分类方法的优势与缺陷。而目前看来，基于深度学习的方法和传统方法相比，具有更高的精度和计算性能，例如，文献[70]中开发的双流网络方法，在UCF101数据集上取得了88%的准确率，比当时最先进的算法又提高了0.1个百分点；文献[56]中采用的ST-GCN方法，在NTU-RGB+D数据集上，在指标cross-subject上将当时的最高准确率提高了2个百分点左右，在cross-view上将当时的最高准确率提高了近4个百分点，可以说是巨大的提升。而Peng等人[81]在NTU RGB+D和Kinetics数据集上，首次基于神经架构搜索自动生成图卷积结构，甚至将准确率刷新到了95.7%。

6 动作评价的研究现状

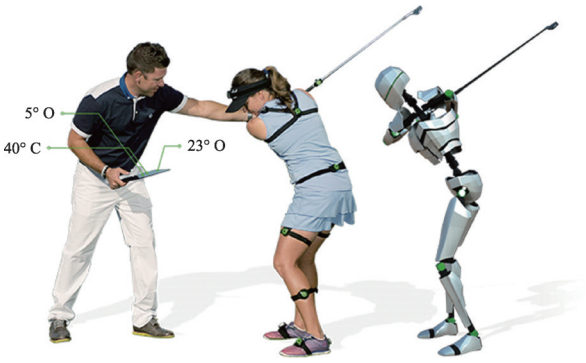
动作评价是最近几年逐步受到关注的研究课题，但目前尚未有明确的概念定义和理论阐述。从动作评价的目的和主要处理过程来看，将动作评价描述为：将输入的“学习者”数据经过动作识别之后，与相对应的“教师”数据进行对比，结合定量指标及专家知识，评价“学习者”动作的完成质量，并给予“学习者”以动作改进的反馈。目前动作评价相关的研究还比较少，但其在体育训练、医疗康复、艺术表演等真实场景下的迫切需求，使其逐渐成为新的研究热点。

动作评价和动作识别在整体处理流程上有共通之处（如图1），也需要经历数据预处理和特征描述等步骤。并且，动作识别往往是动作评价的基础和前提。但是，与动作识别最大的不同在于：动作评价不仅需要对动作外观进行相似性判定，还需要专家知识的介入，对动作的规范性、流畅性、艺术性等一些内在的、隐含的特征进行评价。可以认为：正是因为增加了专家的经验，才使得对动作的处理从分类问题向评价问题转变。因此，在深入分析相关工作之后，本文采取了以专家知识介入方式为依据的分类方式，将当前的动作评价相关工作划分为如下几类：(1)为专家提供可视化工具，构建专家经验与定量参数间的联系；(2)在特征描述中引入专家知识；(3)基于专家知识制定动作规范；(4)基于大数据的动作评价，采用大数据分析来替代专家知识。下面分别进行介绍。

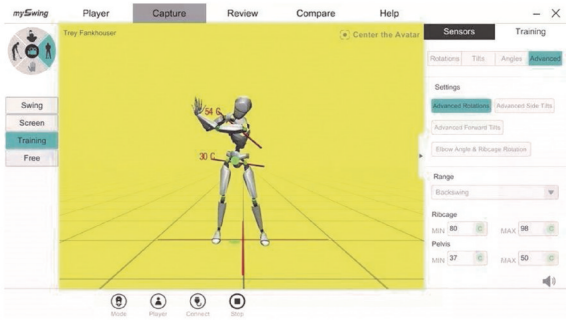
6.1 动作评价的可视化工具

想要在动作识别的基础上加入专家知识其实是很困难的，这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，专家可能也不清楚影响动作质量的具体参数。因此，动作评价的第一阶段不是随意增加专家知识，而是为专家提供工具，使他们能够更加全方位地、定量地、可视化地观察各种动作参数，从而辅助专家发现规律。

近年来，不少研究者们开发出了各种动作评价系统。这些动作评价系统无一例外都采用了对三维动作数据的可视化手段。如陈学梅[14]所开发的高尔夫挥杆动作评价系统，能够对比训练者进行挥杆动作时的关节角度与标准挥杆动作的差异，并直观地将差异展现出来，辅助球员进行练习。图3则提供了诺亦腾公司开发的高尔夫评估和训练系统的应用场景和软件界面，从图中可见，其将动作数据三维可视地显示，用户可以360°观察动作骨骼，并获得重要动作关节的数值。图3(a)为运动员佩戴动作捕捉设备进行训练的实景展示，图3(b)为高尔夫评价系统的界面。它可以提供运动员的关节角度、挥杆速度、加速度、动力链等多项数据，并可以与其他运动员进行



(a) 实景展示
(a) Display of actual scene



(b) 高尔夫评价系统的界面
(b) Interface of golf evaluation system

图3 诺亦腾开发的高尔夫评估和训练系统：mySwing

Fig.3 Golf evaluation and training system developed by Noitom: mySwing

对比，帮助运动员更好地训练和提高。

京剧是一种非常复杂的艺术表演形式，很难进行定量化动作评价。最近，一些研究者在京剧动作评价方面进行了探索，他们充分利用了可视化工具来发现动作规律。王台瑞[19]基于3D动捕设备采集的数据，分析了京剧表演中专业表演者与学习者动作的异同。他将表演者的三维动作数据可视化为三维空间中的离散点集，通过研究点集的分布规律来进行动作评价。研究有9个受试者，其中既有科班学生、戏曲学校学生（非科班）、有扎实舞蹈基础的学生，也有其他的普通学生。结果发现，得到京剧专家很高评分的学生，通过动捕获得的骨骼数据与专家之间的相似性并不一定高。因此要把数据和人类感受很好地结合起来还是很具有挑战性的。将动作序列中的关键参数（如关节角度、关节变化速度、运动轨迹等）进行可视化，并以直观的方式进行对比，可以为专家提供有力的分析工具，有望辅助于将定性的专家知识转化为定量的动作标准，并发现动作的内在规律。这项工作可以作为动作评价的必要模式，而其中复杂运动参数的可视化方法及分析策略可作为进一步研究的要点。

6.2　在特征描述中引入专家知识

特征描述方法对于动作评价具有重要意义。与动作识别不同的是，动作评价的特征描述不仅仅用来评价动作外观的相似性，更要能反映出此类动作的专业特征。因此，动作评价的关键就是要引入更有科学性、专业性的特征描述。在这个问题上，专家知识必不可少。

在很多体育运动的动作评价中，都可以在特征描述阶段引入专家知识。例如，各种运动都有比较固定的评价规则，这些规则代表了裁判或专家在进行动作评价时所关注的重点，可以将这些规则转化成容易评价的定量指标，从而用于动作的相似性度量。

所谓相似性度量，即综合评定两个事物之间相近程度的一种度量。将相似性度量引申运用在人体动作评价中，就是基于定量的评价指标，对“学习者”动作与“教师”动作进行相似性比较，从而实现对动作完成质量的评价。这其中的关键点是：(1)由领域专家确定应采用哪些特征描述符作为动作评价的指标；(2)如何定义样本之间的相似性测度。

上节中的陈学梅[14]所研制出的高尔夫挥杆评价系统，主要使用了和挥杆动作联系最紧密的关节角度的指标。李奎[15]的工作则根据对羽毛球挥拍动作的研究，使用非定长稠密轨迹算法来表征这些动作，然后计算待分析动作与标准动作之间的切比雪夫距离来衡量它们的相似度。张晓莹等人[82]对两名男子竞技健美操世界冠军完成难度动作C289不同技术的运动学特征进行深入分析与量化研究，并进行相应的技术诊断，揭示完成此难度的运动学特征与核心技术，为运动员提高难度动作成功率奠定基础，同时也为难度动作的科学训练提供可靠的理论依据和实践参考。Alexiadis等人[83]采用关节旋转的四元数特征对舞蹈动作进行评价，并基于此实现了动作序列的评估。

人们发现对于不同的专业动作，各个身体关节在动作中起到的作用是不同的，因此在动作评价中，应给各个关节赋予一个权重，由此可突出重点关节的作用。各关节的权重参数一般就需要根据专家经验来设置，这种设置方式显然具有一定的主观性。也有人通过对动作的分析来自动为骨骼关节计算权重。如Patrona等人[26]提出了一种自动和动态加权的方法，根据动作参与程度的差异，赋予关节相应的权重，再整合基于动能的描述符采样，进行相似性度量。随后利用模糊逻辑提供语义反馈，指导用户如何更准确地执行操作。

由上述研究可以发现，速度、加速度、关节角度等基本动作参数，往往并不能满足动作评价的需求，而需要在这些参数基础上结合专家经验进行综合分析与特征描述，以得出综合的评价指标。

6.3　基于专家知识制定动作规范

确定动作的特征描述方式之后，可以更进一步基于专家知识建立动作规范：即可依据此规范评估动作做到何种程度可以被认为是合格的、优秀的或者是错误的。

在医疗康复训练的动作评价中，制定动作规范的方式比较常见。李睿敏[84]针对发展性协调障碍疾病，提出了一种基于时域滤波卷积神经网络的动作检测方法，实现了交互过程中的精细动作评估。Richter等人[85]针对髋部外展、髋关节伸展和髋部弯曲这三种运动错误进行了研究。他们定义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分别对应不同的运动错误，以此分析病人的动作执行情况，针对性地给出评价和指导。在康复医疗场景下，有很多与动作相关的障碍性疾病，此类疾病的临床诊断通常是由专业医师通过观察和分析病人在一些特定动作评估任务中的表现给出的。但

医师评估的花费的时间长、费用昂贵，很难大规模筛查，因此，进行自动化动作评价既能满足计算机领域对动作评价的研究需要，又能推进自动化医疗辅助诊断的发展[84]。

制定动作规范的方式在其他领域中也有应用。徐铮[86]提出了一种24式太极拳动作评价方法。他首先通过与太极专家的交流沟通，建立了太极拳动作原语库，然后据此制定了太极拳动作相似度金字塔模型以及相应的动作规范。在此基础上，采用典型相关分析（canonical correlation analysis，CCA）方法对动作数据进行局部关节特征向量相似度量，并依据所制定的动作规范对用户的太极拳动作给出评价和指导建议。

在运动或表演领域，“动作评价”一般都是一种主观方式，而基于专家知识来制定定量化动作评价标准及动作规范，则可以将原本比较模糊的评价任务变得清晰明确，使评价具有更好的客观性和科学性。

6.4 基于大数据的动作评价

在复杂表演动作的评价方面，专家知识具有主观性、模糊性和隐含性，很难获得显式的、定量化的表达。事实上，目标动作的特征都蕴含在其动作数据中，如果采用大数据分析的方式，通过对大量教师动作（或专家动作）的数据分析，也许能发现动作中的合理评价标准。这种方式相当于是采用大数据分析的手段来替代专家的主观评价，也许能够为专业动作评价提供一种新的有效手段。

现有的数据集中记录的数据多为简单的日常动作，并不能满足专业领域动作的识别和评价，因此需要构建专用的动作数据集来实现专业动作的大数据分析。吕默等人[16]采集了大量高水平运动员的标准动作，扩充了MSR Action3D数据集，再结合健美操国际权威标准制备对比数据库，然后将骨骼特征与深度局部特征进行傅里叶金字塔过滤并融合，根据融合特征进行动作的识别与评价。基于此方法开发的健美操辅助评审系统可以有效帮助裁判对竞技健美操难度动作给出正确的分数。

基于大数据的动作评价相关工作目前还非常少，吕默等也只是采用了传统的分类方法来对动作数据进行分类与识别；尚需进一步解决的问题包括：专业动作数据集的建设、适用于专业动作评价的网络构建、评价结果的合理性评估等诸多问题，有待研究者的进一步探索。

综上所述，表3列出了动作评价相关方法的类别、内容和方法。

表3 动作评价方法总结

Table 3 Summary of action evaluation methods

方法类别	相关工作	评价对象	标准/方法
动作评价的可视化工具	陈学梅[14]	高尔夫挥杆动作	关节角度
	李奎[15]	羽毛球挥拍动作	切比雪夫距离
	王台瑞[19]	京剧	专家和机器分别打分
在特征描述中引入专家知识	陈学梅[14]	高尔夫挥杆动作	关节角度
	Zhang等人[82]	竞技健美操	人体动力学
	Alexiadis等人[83]	舞蹈	四元数特征
	Patrona等人[26]	医疗训练	动态加权、动能描述符
基于专家知识制定动作规范	李睿敏[84]	发展性协调障碍症	基于时域滤波的CNN
	Richter等人[85]	髋外展、髋伸展和髋弯曲	基于规则和标签
	徐铮[86]	24式太极拳	CCA
基于大数据的动作评价	吕默等人[16]	体操	大数据

7 结束语

近年来，人体动作识别和动作评价的相关研究获得了长足发展。本文首先给出了二者较为明确的概念定义，探讨了二者之间存在的区别与联系。以此为基础，从数据处理流程的角度出发系统地梳理了两者的技术模块，并将这两类问题归纳到了一个统一的技术框架中。之后，依据该技术框架，对各个技术模块的相关工作进行了系统的介绍与分析。

在动作识别问题上，随着深度学习的应用，普通动作的识别精度已经可以达到相当高的程度，如前文提到的，Peng等人[81]在NTU RGB+D和Kinetics数据集上，已经将识别准确率刷新到了95.7%。虽然可以取得如此优异的实验结果，但人体运动的高复杂性和多变化性使得当前的识别方法并没有完全满足实际应用需求。当前存在的瓶颈及未来的研究重点包括：

（1）缺乏标注良好的大型数据集。虽然表1给出了不少动作识别相关数据集，但与图像处理领域的诸多经典数据库（如ImageNet、MS-COCO、Open Images等）相比，其数据集的完备性和标注程度还有待提高，动作识别领域依然缺乏大规模且标注良好的基准数据集。在深度学习成为主流方法的当今时代，标注良好的大型数据集对动作识别领域的发展具有十分关键的作用。今后在数据集建设中，一方面可以考虑进一步细化动作粒度，将数据集中的动作进行子动作划分及标注；另一方面需要提供更丰富的

标注标签，例如对于视频数据不仅提供动作类别标签，还可进一步提供人体部位、骨架甚至与人体进行互动的环境物体等标注。

（2）大部分研究仍处于实验室阶段，在实际应用场景中的鲁棒性不强。在实际应用环境中所采集的数据，大都存在着多人体目标、遮挡、摄像机移位等干扰因素，目前方法对这些实际数据中的干扰能力还不够强，导致其实用化程度十分有限。一个可行的解决策略是采用多特征融合的方法，提高模型泛化能力，解决多样化场景下的人体动作识别问题。一些研究者已经在这方面做出了初步尝试，例如文献[74]采用了多通道特征融合的方式，文献[75]则综合考虑静态、动态和高层次特征，文献[80]则融合了不同时长的动作特征。利用多特征融合的策略，这些方法在抗环境干扰方面都取得了不错的效果。该思路依然值得进一步深入探讨。

（3）对于速度很快的动作，尚无法达到满意的识别效果。在一些专业运动领域，例如健美操等，其动作密集而快速（如健美操中的各种空翻动作），准确识别出每一次的动作难度依然很大。对于这种数秒内完成多次的动作，需要应用更细粒度的数据标签进行训练，而另一个值得考虑的思路是结合注意力机制，对关键帧中的快速动作区域进行重点关注，以提高识别效果。

（4）尚缺乏对动作中的语义信息的理解。如在京剧表演中相似的腿部姿态却可能代表着不同的自然语义。但目前的动作识别技术仅通过当前动作外观进行分类，很难对这种动作的语义差别进行区分。因此，借助上下文及环境等对动作的语义信息进行识别理解是一个重要的研究点，该问题研究也能为动作评价打下良好的基础。

在动作评价问题上，当前的研究还比较初步，目前出现的针对羽毛球、高尔夫球、康复医疗等专业领域的动作评价工作，所针对的都是比较简单、标准化的动作；其所采用的指标也比较单一，主要考虑关节角度、速度、加速度等基本方位指标。分析来看，动作评价研究所面临的关键问题包括：

（1）构建符合专业评价要求的数字化评价标准。这是进行专业动作评价的关键问题，其重点是需要将专业动作规范及专家的感性认知转化为量化的指标。目前虽然已经有了一些相关的工作，但其方法主要针对特定的动作领域，很难推广，在这方面还没有特别成熟而系统的方法。一种值得探索的方式是利用相似性度量算法自动发现“学习者”动作与“教师”动作的差异之处，再进一步结合专家知识或者直接启发专家形成定量化动作规范。

（2）从“形似”到“神似”。当前的动作评价工作仅仅局限在外在动作相似度的比较上。而一些专业领域的动作，如京剧表演、舞蹈等，讲究“以形传神，形神兼备”，其不仅要求在身段、身法上“形似”，还需要通过动作、表情等将内在的“神韵”表达出来。对这类动作的评价不能仅仅停留在动作相似性的度量上，还需有平衡性、流畅性、稳定性等更高级别特征的评估，需要思考如何在有形的“数据”和无形的“美感”之间搭建桥梁，实现更能反映艺术性的定量化评价。这方面的工作尚未见开展，却有重要研究意义。采用深度学习方法对大量表演数据进行分析，从中发现高层次的艺术特征，也许可以作为一条可探索的思路。

参考文献：

[1] DURIC Z, GRAY W, HEISHMAN R, et al. Integrating perceptual and cognitive modeling for adaptive and intelligent human-computer interaction[J]. Proceedings of the IEEE, 2002, 90(7): 1272-1289.

[2] KWAK S, HAN B, HAN J H. Scenario-based video event recognition by constraint flow[C]//Proceedings of the 24th IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, Jun 20-25, 2011. Washington: IEEE Computer Society, 2011: 3345-3352.

[3] GAUR U, ZHU Y, SONG B, et al. A “string of feature graphs” model for recognition of complex activities in natural videos[C]//Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Nov 6-13, 2011. Washington: IEEE Computer Society, 2011: 2595-2602.

[4] PARK S, AGGARWAL J K. Recognition of two-person interactions using a hierarchical Bayesian network[C]//Proceedings of the 2003 ACM SIGMM International Workshop on Video Surveillance. New York: ACM, 2003: 65-76.

[5] JUNEJO I, DEXTER E, LAPTEV I, et al. View-independent action recognition from temporal self-similarities[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(1): 172-185.

[6] THANGALI A, NASH J P, SCLAROFF S, et al. Exploiting phonological constraints for handshape inference in ASL video[C]//Proceedings of the 24th IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, Jun 20-25, 2011. Washington: IEEE Computer Society, 2011: 521-528.

[7] 樊景超, 周国民. 基于Kinect骨骼跟踪技术的手势识别研

究[J]. 安徽农业科学, 2014, 42(11): 3444-3446.
FAN J C, ZHOU G M. The research of gesture recognition based on kinect skeleton tracking technology[J]. Journal of Anhui Agricultural Sciences, 2014, 42(11): 3444-3446.

[8] COOPER H, BOWDEN R. Large lexicon detection of sign language[C]//LNCS 4796: Proceedings of the 2007 IEEE International Workshop on Human-Computer Interaction, Rio de Janeiro, Oct 20, 2007. Berlin, Heidelberg: Springer, 2007: 88-97.

[9] CHANG Y J, CHEN S F, HUANG J D. A kinect-based system for physical rehabilitation: a pilot study for young adults with motor disabilities[J]. Research in Developmental Disabilities, 2011, 32: 2566-2570.

[10] 李少波. 机器人的人体姿态动作识别与模仿算法[D]. 上海:上海交通大学, 2013.
LI S B. Algorithm of human posture action recognition and imitation for robots[D]. Shanghai: Shanghai Jiaotong University, 2013.

[11] REHG J M, ABOWD G D, ROZGA A, et al. Decoding children's social behavior[C]//Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, Jun 23-28, 2013. Washington: IEEE Computer Society, 2013: 3414-3421.

[12] PRESTI L L, SCLAROFF S, ROZGA A. Joint alignment and modeling of correlated behavior streams[C]//Proceedings of the 2013 IEEE International Conference on Computer Vision, Sydney, Dec 1-8, 2013. Washington: IEEE Computer Society, 2013: 730-737.

[13] JOHANSSON G. Visual perception of biological motion and a model for its analysis[J]. Perception & Psychophysics, 1973, 14: 201-211.

[14] 陈学梅. 基于人体三维姿态的动作评价系统[D]. 杭州: 浙江大学, 2018.
CHEN X M. An action evaluating system based on 3D human posture[D]. Hangzhou: Zhejiang University, 2018.

[15] 李奎. 羽毛球运动员挥拍动作的捕捉、识别与分析[D]. 成都: 电子科技大学, 2017.
LI K. Capture, recognition and analysis of badminton player's swing[D]. Chengdu: University of Electronic Science and Technology of China, 2017.

[16] 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]. 电子设计工程, 2019, 27(16): 6-10.
LV M, WAN L C. Design of sports competition aided evaluation system based on big data and motion recognition algorithm[J]. Electronic Design Engineering, 2019, 27(16): 6-10.

[17] KAO C I, SPIRO I, SEUNGKYU L, et al. Dancing with turks[C]//Proceedings of the 23rd Annual ACM Conference on Multimedia Conference, Brisbane, Oct 26-30, 2015. New York: ACM, 2015: 241-250.

[18] SCOTT J, COLLINS R, FUNK C, et al. 4D model-based spatio-temporal alignment of scripted Taiji Quan sequences[C]// Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Oct 22-29, 2017. Washington: IEEE Computer Society, 2017: 795-804.

[19] 王台瑞. 以三维撷取探索戏曲动作教与学之异同[J]. 艺术教育研究, 2018(35): 69-92.
WANG T J. USING 3D motion capture study Chinese opera performance movements[J]. Research in Arts Education, 2018 (35): 69-92.

[20] 徐光祐, 曹媛媛. 动作识别与行为理解综述[J]. 中国图象图形学报, 2009, 14(2): 189-195.
XU G Y, CAO Y Y. Action recognition and activity understanding: a review[J]. Journal of Image and Graphics, 2009, 14(2): 189-195.

[21] WU D, SHARMA N, BLUMENSTEIN M. Recent advances in video-based human action recognition using deep learning: a review[C]//Proceedings of the 2017 International Joint Conference on Neural Networks, Anchorage, May 14-19, 2017. Piscataway: IEEE, 2017: 2865-2872.

[22] PRESTI L L, MARCO L C. 3D skeleton-based human action classification: a survey[J]. Pattern Recognition, 2016, 53: 130-147.

[23] 黄国范, 李亚. 人体动作姿态识别综述[J]. 电脑知识与技术, 2013, 9(1): 133-135.
HUANG G F, LI Y. A survey of human action and pose recognition[J]. Computer Knowledge and Technology, 2013, 9(1): 133-135.

[24] 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(4): 1-8.
TIAN Y, LI F D. Research review on human body gesture recognition based on depth data[J]. Computer Engineering and Applications, 2020, 56(4): 1-8.

[25] 黄晴晴, 周风余, 刘美珍. 基于视频的人体动作识别算法综述[J]. 计算机应用研究, 2020, 37(11): 3213-3219.
HUANG Q Q, ZHOU F Y, LIU M Z. Survey of human action recognition algorithms based on video[J]. Application Research of Computers, 2020, 37 (11): 3213-3219.

[26] PATRONA F, CHATZITOFIS A, ZARPALAS D, et al. Motion analysis: action detection, recognition and evaluation based on motion capture data[J]. Pattern Recognition, 2018, 76: 612-622.

[27] SOOMRO K, ZAMIR A R, SHAH M, et al. UCF101: a dataset of 101 human actions classes form video vision the wild[J]. arXiv:1212.0402, 2012.

[28] KUEHNE H, JHUANG H, GARROTE E, et al. HMDB: a large video database for human motion recognition[C]// Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Nov 6-13, 2011. Washington: IEEE

Computer Society, 2011: 2556-2563.

[29] SINGH S, VELASTIN S A, RAGHEB H. MuHAVi: a multi-camera human action video dataset for the evaluation of action recognition methods[C]//Proceedings of the 2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance, Washington, Aug 29-Sep 1, 2010. Washington: IEEE Computer Society, 2010: 48-55.

[30] LI W Q, ZHANG Z Y, LIU Z C, et al. Action recognition based on a bag of 3D points[C]//Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, Jun 13-18, 2010. Washington: IEEE Computer Society, 2010: 9-14.

[31] SHAHROUDY A, LIU J, TIAN-TSONG N G, et al. NTU RGB+D: a large scale dataset for 3D human activity analysis[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Jun 27-30, 2016. Washington: IEEE Computer Society, 2016: 1010-1019.

[32] LIU J, SHAHROUDY A, PEREZ M, et al. NTU RGB+D 120: a large-scale benchmark for 3D human activity understanding[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 42: 2684-2701.

[33] BLOOM V, MAKRIS D, ARGYRIOU V. G3D: a gaming action dataset and real time action recognition evaluation framework[C]//Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Providence, Jun 16-21, 2012. Washington: IEEE Computer Society, 2012: 7-12.

[34] DABOV K, FOI A, KATKOVNIK V, et al. Image denoising by sparse 3D transform-domain collaborative filtering[J]. IEEE Transactions on Image Processing, 2007, 16: 2080-2095.

[35] MAGGIONI M, BORACCHI G, FOI A. Video denoising using separable 4D nonlocal spatiotemporal transforms[J]. Proceedings of SPIE-The International Society for Optical Engineering, 2011, 7870(3): 1-12.

[36] DAVY A, EHRET T, MOREL J M. et al. Non-local video denoising by CNN[J]. arXiv:1811.12758, 2018.

[37] ARIAS P, MOREL J M. Video denoising via empirical Bayesian estimation of space-time patches[J]. Journal of Mathematical Imaging & Vision, 2018, 60(1): 70-93.

[38] TASSANO M, DELON J, VEIT T. DVDNet: a fast network for deep video denoising[C]//Proceedings of the 2019 IEEE International Conference on Image Processing, Taipei, China, Sep 22-25, 2019. Piscataway: IEEE, 2019: 1805-1809.

[39] TASSANO M, DELON J, VEIT T. FastDVDnet: towards real-time deep video denoising without flow estimation[J]. arXiv:1907.01361v2, 2019.

[40] PING W, ZHENG N, ZHAO Y, et al. Concurrent action detection with structural prediction[C]//Proceedings of the 2013 International Conference on Computer Vision, Sydney, Dec 1-8, 2013. Washington: IEEE Computer Society, 2013: 3136-3143.

[41] WU D, SHAO L. Leveraging hierarchical parametric networks for skeletal joints based action segmentation and recognition [C]//Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, Jun 23-28, 2014. Washington: IEEE Computer Society, 2014: 724-731.

[42] WANG C, WANG Y, YUILLE A L. An approach to pose-based action recognition[C]//Proceedings of the 2013 Conference on Computer Vision and Pattern Recognition, Portland, Jun 23-28, 2013. Washington: IEEE Computer Society, 2013: 915-922.

[43] SEDMIDUBSKY J, ELIAS P, BUDIKOVA P, et al. Content-based management of human motion data: survey and challenges[J]. IEEE Access, 2021, 9: 64241-64255.

[44] 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: 971-987.

[45] 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, 2020, 12(3): 261-273.
TANG C, TANG L G, LIU B. A survey of image feature detecting and matching methods[J]. Journal of Nanjing University of Information Science & Technology, 2020, 12(3): 261-273.

[46] 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: 257-267.

[47] DALAL N，TRIGGS B. Histograms of oriented gradients for human detection[C]//Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, Jun 20-25, 2005. Washington: IEEE Computer Society, 2005: 886-893.

[48] LAPTEV I. On space-time interest points[J]. International Journal of Computer Vision, 2005, 64(2/3): 107-123.

[49] LAPTEV I, MARSZALEK M, SCHMID C, et al. Learning realistic human actions from movies[C]//Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Anchorage, Jun 24-26, 2008. Washington: IEEE Computer Society, 2008: 1-8.

[50] WANG H, ULLAH M M, KLÄSER A, et al. Evaluation of local spatio-temporal features for action recognition[C]//Proceedings of the 2009 British Machine Vision Conference, London, Sep 7-10, 2009. London: The British Machine Vision Association, 2009: 1-11.

[51] BAUMANN J, WESSEL R, KRÜGER B, et al. Action graph a versatile data structure for action recognition[C]//Proceedings of the 9th International Conference on Computer Graphics Theory and Applications, Lisbon, Jan 5-8, 2014: 325-334.

[52] BARNACHON M, BOUAKAZ S, BOUFAMA B, et al. A real-time system for motion retrieval and interpretation[J]. Pattern Recognition Letters, 2013, 34(15): 1789-1798.

[53] MASOOD S Z, MASOOD S Z, TAPPEN M F, et al. Exploring the trade-off between accuracy and observational latency in action recognition[J]. International Journal of Computer Vision, 2013, 101(3): 420-436.

[54] MÜLLER M, RÖDER T, CLAUSEN M. Efficient content-based retrieval of motion capture data[J]. ACM Transactions on Graphics, 2005, 24(3): 677-685.

[55] CHERON G, LAPTEV I, SCHMID C. P-CNN: pose-based CNN features for action recognition[C]//Proceedings of the 2015 IEEE International Conference on Computer Vision, Santiago, Dec 7-13, 2015. Washington: IEEE Computer Society, 2015: 3218-3226.

[56] YAN S, XIONG Y, LIN D, et al. Spatial temporal graph convolutional networks for skeleton-based action recognition[C]//Proceedings of the 32nd AAAI Conference on Artificial Intelligence, the 30th Innovative Applications of Artificial Intelligence, and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence, New Orleans, Feb 2-7, 2018. Menlo Park: AAAI, 2018: 7444-7452.

[57] REN B, LIU M, DING R, et al. A survey on 3D skeleton-based action recognition using learning method[J]. arXiv: 2002.05907, 2020.

[58] ZHANG P, XUE J, LAN C, et al. Adding attentiveness to the neurons in recurrent neural networks[C]//LNCS 11213: Proceedings of the 15th European Conference on Computer Vision, Munich, Sep 8-14, 2018. Cham: Springer, 2018: 136-152.

[59] LANITIS A, TAYLOR C J, COOTES T F. Automatic interpretation and coding of face images using flexible models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 743-756.

[60] COOTES T F, EDWARDS G J, TAYLOR C J. Active appearance models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(6): 681-685.

[61] BOBICK A F, WILSON A D. A state-based approach to the representation and recognition of gesture[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(12): 1325-1337.

[62] YAMATO J, OHYA J, ISHII K. Recognizing human action in time sequential images using hidden Markov model[C]//Proceedings of the 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Champaign, Jun 15-18, 1992. Washington: IEEE Computer Society, 1992: 379-385.

[63] NGUYEN N T, PHUNG D Q, VENKATESH S, et al. Learning and detecting activities from movement trajectories using the hierachical hidden Markov model[C]//Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, Jun 20-26, 2005. Washington: IEEE Computer Society, 2005: 955-960.

[64] 梅雪, 胡石, 许松松, 等. 基于多尺度特征的双层隐马尔可夫模型及其在行为识别中的应用[J]. 智能系统学报, 2012, 7(6): 512-517.
MEI X, HU S, XU S S, et al. Multi-scale feature based double-layer HMM and its application in behavior recognition[J]. CAAI Transactions on Intelligent Systems, 2012, 7(6): 512-517.

[65] DU Y T, CHEN F, XU W L, et al. Recognizing interaction activities using dynamic Bayesian network[C]//Proceedings of the 18th International Conference on Pattern Recognition, Hong Kong, China, Aug 20-24, 2006. Washington: IEEE Computer Society, 2006: 618-621.

[66] OLIVER N, HORVITZ E. A comparison of HMMs and dynamic Bayesian networks for recognizing office activities [C]//LNCS 3538: Proceedings of the 10th International Conference on User Modeling, Edinburgh, Jul 24-29, 2005. Berlin, Heidelberg: Springer, 2005: 199-209.

[67] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.
ZHOU Z H. Machine learning[M]. Beijing: Tsinghua University Press, 2016.

[68] 李航. 统计学习方法[M]. 北京: 清华大学出版社, 2012.
LI H. Statistical learning methods[M]. Beijing: Tsinghua University Press, 2012.

[69] PONTIL M, VERRI A. Support vector machines for 3D object recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(6): 637-646.

[70] MANZI A, CAVALLO F, DARIO P. A 3D human posture approach for activity recognition based on depth camera [C]//LNCS 9914: Proceedings of the 14th European Conference on Computer Vision, Amsterdam, Oct 8-10, 15-16, 2016. Cham: Springer, 2016: 432-447.

[71] SCHÜLDT C, LAPTEV I, CAPUTO B. Recognizing human actions: a local SVM approach[C]//Proceedings of the 17th International Conference on Pattern Recognition, Cambridge, Aug 23-26, 2004. Washington: IEEE Computer Society, 2004: 32-36.

[72] MOHAMED E, ISMAIL C, WASSIM B, et al. Posture recognition using an RGB-D camera: exploring 3D body modeling and deep learning approaches[C]//Proceedings of the 2018 IEEE Life Sciences Conference, Montreal, Oct 28-30, 2018. Piscataway: IEEE, 2018: 69-72.

[73] 刘锁兰, 顾嘉晖, 王洪元, 等. 基于关联分区和ST-GCN的人体行为识别[J]. 计算机工程与应用, 2021, 57(13): 168-175.
LIU S L, GU J H, WANG H Y, et al. Human behavior recognition based on associative partition and ST-GCN[J]. Computer Engineering and Applications, 2021, 57(13): 168-175.

[74] JI S W, XU W, YANG M, et al. 3D convolutional neural networks for human action recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1): 221-231.

[75] 李元祥, 谢林柏. 基于深度运动图和密集轨迹的行为识别算法[J]. 计算机工程与应用, 2020, 56(3): 194-200.
LI Y X, XIE L B. Human action recognition based on depth motion map and dense trajectory[J]. Computer Engineering and Applications, 2020, 56(3): 194-200.

[76] SIMONYAN K, ZISSERMAN A. Two-stream convolutional networks for action recognition in videos[C]//Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, Montreal, Dec 8-13, 2014: 568-576.

[77] FEICHTEMHOFER C, PINZ A, ZISSERMAN A, et al. Convolutional two-stream network fusion for video action recognition[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Jun 27-30, 2016. Washington: IEEE Computer Society, 2016: 1933-1941.

[78] 石祥滨, 李怡颖, 刘芳, 等. T-STAM: 基于双流时空注意力机制的端到端的动作识别模型[J]. 计算机应用研究, 2020, 38(3): 1235-1239.
SHI X B, LI Y Y, LIU F, et al. T-STAM: end-to-end action recognition model based on two-stream network with spatio-temporal attention mechanism[J]. Application Research of Computers, 2020, 38(3): 1235-1239.

[79] DONAHUE J, HENDRICKSL A, GUADARRAMA S, et al. Long-term recurrent convolutional networks for visual recognition and description[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 677-691.

[80] 杨珂, 王敬宇, 戚琦, 等. LSCN: 一种用于动作识别的长短时序关注网络[J]. 电子学报, 2020, 48(3): 503-509.
YANG K, WANG J Y, QI Q, et al. LSCN: concerning long and short sequence together for action recognition[J]. Acta Electronica Sinica, 2020, 48(3): 503-509.

[81] PENG W, HONG X P, CHEN H Y, et al. Learning graph convolutional network for skeleton-based human action recognition by neural searching[C]//Proceedings of the 34th AAAI Conference on Artificial Intelligence, the 32nd Innovative Applications of Artificial Intelligence Conference, the 10th AAAI Symposium on Educational Advances in Artificial Intelligence, New York, Feb 7-12, 2020. Menlo Park: AAAI, 2020: 2669-2676.

[82] 张晓莹, 刘莉, 赵轩立. 竞技健美操难度动作C289不同技术特征的运动学分析[J]. 北京体育大学学报, 2017, 40(10): 99-105.
ZHANG X Y, LIU L, ZHAO X L. Kinematic analysis on different technical characteristics of C289 in aerobic gymnastics [J]. Journal of Beijing Sport University, 2017, 40(10): 99-105.

[83] ALEXIADIS D S, DARAS P. Quaternionic signal processing techniques for automatic evaluation of dance performances from MoCap data[J]. IEEE Transactions on Multimedia, 2014, 16(5): 1391-1406.

[84] 李睿敏. 基于视觉数据的人体动作精细分类及评估方法研究[D]. 北京: 中国科学院大学, 2020.
LI R M. Research on fine classification and evaluation of human action based on visual data[D]. Beijing: University of Chinese Academy of Sciences, 2020.

[85] RICHTER J, WIEDE C, HEINKEL U, et al. 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 VISIGRAPP 14th International Conference on Computer Vision Theory and Applications, Prague, Feb 25-27, 2019: 497-504.

[86] 徐铮. 基于全身动捕的太极拳辅助教学与评价方法[D]. 郑州: 郑州大学, 2018.
XU Z. Taiji boxing assist teaching and evaluation method based on whole body motion capture[D]. Zhengzhou: Zhengzhou University, 2018.

杨刚(1977—),男,山西长治人,博士,副教授,CCF会员,主要研究方向为计算机图形学、虚拟现实等。

YANG Gang, born in 1977, Ph.D., associate professor, member of CCF. His research interests include computer graphics, virtual reality, etc.

张宇姝(1997—),女,湖北十堰人,硕士研究生,CCF会员,主要研究方向为计算机科学与技术(虚拟现实方向)。

ZHANG Yushu, born in 1997, M.S. candidate, member of CCF. Her research interest is computer science and technology (direction of virtual reality).

宋震(1976—),男,北京人,博士,教授,博士生导师,主要研究方向为传统戏剧数字化、戏剧人工智能、数字演员与未来戏剧等。

SONG Zhen, born in 1976, Ph.D., professor, Ph.D. supervisor. His research interests include digitalization of traditional drama, drama artificial intelligence, digital actors and future drama, etc.