计算机应用
Journal of Computer Applications
ISSN 1001-9081,CN 51-1307/TP

《计算机应用》网络首发论文

题目：基于人体骨架特征编码的健身动作识别方法
作者：郭天晓，胡庆锐，李建伟，沈燕飞
收稿日期：2020-07-30
网络首发日期：2020-10-15
引用格式：郭天晓，胡庆锐，李建伟，沈燕飞．基于人体骨架特征编码的健身动作识别方法．计算机应用.
https://kns.cnki.net/kcms/detail/51.1307.TP.20201015.1411.008.html

Journal of Computer Applications ISSN 1001-9081
计算机应用 CODEN JYIIDU http://www.joca.cn

doi:10.11772/j.issn.1001-9081.2020071113

基于人体骨架特征编码的健身动作识别方法

郭天晓[1]，胡庆锐[1]，李建伟[2*]，沈燕飞[2]

(1.北京体育大学 运动人体科学学院，北京 100084； 2.北京体育大学 体育工程学院，北京 100084)

(*通信作者电子邮箱 jianwei@bsu.edu.cn)

健身动作识别是智能健身系统的核心环节，为了提高健身动作识别算法的精度和速度，减少健身动作中人体整体位移对识别结果的影响，提出了一种基于人体骨架特征编码的动作识别方法。本方法包括三个步骤：首先，构建精简人体骨架模型，并利用人体姿态估计技术提取骨架模型中各关节点的坐标信息；其次，利用人体中心投影法提取动作特征区域以消除人体整体位移对动作识别的影响；最后，将特征区域编码为特征向量并输入多分类器进行动作识别，同时通过优化特征向量长度使识别精度和速度达到最优。实验结果表明，本方法在包含 28 种动作的自建健身数据集上的动作识别准确率为 97.24%，能够有效识别各类健身动作；在公开的 KTH 和 Weizmann 数据集上，本方法的动作识别和准确率分别为 91.67%和 90%，优于其他同类型方法。

关键词: 计算机视觉；动作识别；智能健身；骨架信息；姿态估计

中图分类号：TP391.41 **文献标志码：A**

Fitness action recognition method based on human skeleton feature encoding

GUO Tianxiao[1], HU Qingrui[1], LI Jianwei[2*], SHEN Yanfei[2]

(1. School of Sport Science, Beijing Sport University, Beijing 100084, China;
2. School of Sports Engineering, Beijing Sport University, Beijing 100084, China)

Abstract: Fitness action recognition is the core of the intelligent fitness system. In order to improve the accuracy and speed of fitness action recognition algorithm, and reduce the influence of the global displacement of fitness actions on recognition results, a new action recognition method based on human skeleton feature encoding was proposed which included three steps: firstly, the simplified model was constructed and the sequence of skeleton coordinates was extracted through the pose estimation technology; secondly, the feature region of skeleton information was obtained by using the human center projection to eliminate the disturbance of the global displacement of fitness actions; finally, the feature region was encoded into a feature vector with an optimized length for improving recognition speed and precision, and a multi-classifier was designed to achieve the action recognition. The experiment results showed that the proposed method achieved the precision of 97.24% on the self-built fitness dataset with 28 fitness actions, which verified the effectiveness to recognize fitness actions. On the public KTH and Weizmann datasets, the precisions of the proposed method were 91.67% and 90%, respectively, and outperformed other similar methods.

Keywords: computer vision; action recognition; intelligent fitness; skeleton information; pose estimation

0 引言

随着计算机视觉和图像处理技术的高速发展，基于视觉信息处理的智能化训练系统逐渐被应用于运动训练[1]和康复医疗领域[2]。对于初学者而言，接受及时有效的指导和反馈不仅能够帮助其掌握动作[3]，还能够有效避免运动损伤[4,5]。传统的健身动作指导是在教练员的监督引导下纠正错误动作以实现良好的锻炼效果，要求在特定场地下由专人指导进行练习，不适用于居家环境以及利用碎片化时间锻炼的场景。而当前已经出现的依托智能设备的健身指导方案[6]大多缺少对运动过程的有效监控且无法给出反馈和建议，不利于初学者

收稿日期: 2020-07-30; **修回日期**:2020-09-25; **录用日期**: 2020-10-05。
基金项目:国家重点研发计划（2018YFC2000600）；中央高校基本科研业务费专项资金（校 2020056）；中央高校基本科研业务费专项资金（校 2020010）。

作者简介: 郭天晓(1996—)，男，山西大同人，硕士研究生，主要研究方向：智能体育、体育视频分析； 胡庆锐(1996—)，男，安徽滁州人，硕士研究生，主要研究方向：智能体育、体育视频分析； 李建伟(1987—)，女，甘肃兰州人，讲师，博士，主要研究方向：SLAM、计算机视觉、智能体育； 沈燕飞(1976—)，男，江苏靖江人，教授，博士，主要研究方向：人工智能技术、智能视频分析、体育大数据。

掌握动作[7]。运动技能学习过程的起始阶段为泛化阶段[8]，其学习重点为掌握动作要领，需要通过重复观看示范和接收反馈来纠正错误动作[9]。通过技术手段对运动过程进行监控和评估，不仅能够帮助运动者掌握动作，还能够节省人力成本，增加训练过程的趣味性和互动性。

智能健身系统[10, 11]是集成了人体运动信息采集，数据处理与交互，用户终端与设备等模块的综合训练平台。健身动作识别作为其中的核心环节之一，通过采集和分析人体运动特征区分受试者执行的不同动作。目前，人体动作识别主要分为基于惯性传感器[12, 13]和基于视觉特征采集[1, 14]的两大类方法。前者通过可穿戴设备采集人体运动学信息完成动作识别，但在各关节处附着传感器不仅提高了成本，也会影响运动体验。而采集视觉特征进行动作识别的方法能够依托各类相机完成非侵入式[15]的动作识别，更适用于健身场景。

当前，利用人体视觉特征进行动作识别的方法主要分为基于传统特征提取和基于深度学习的两大类。基于深度学习的动作识别方法构建神经网络[16-20]描述人体运动特征，在大型动作数据集上实现良好的检测效果，此类方法通常依赖大量数据进行模型训练且对计算资源要求较高，限制了其在不同场景下的应用。相对而言，基于传统特征提取的方法[21, 22]对数据量和计算资源的要求较小，能够根据不同需要提取相应动作特征完成识别。在运动训练领域的相关研究中，研究者根据训练目的和项目特点设计动作特征提取方法来完成各类动作的识别和分析。如 Orucu 等人[14]针对上肢力量训练中对动作执行标准程度的评价和指导问题，依托 kinect V2 设计了一套智能训练系统，该系统通过提取受试者上肢各关节点的运动数据监控和评估日常训练过程，实验结果证明该系统能有效改善动作质量。Li 等人[23]为了对比赛视频中运动员的动作进行分析，通过分层提取视频特征获取运动员动作的关键运动学参数并据此完成动作识别，辅助教练员完成比赛录像分析。Ting 等人[1]针对羽毛球运动中复杂技术动作的分类问题，采集各动作 RGB-D 视频并提取四元数特征向量对 10 类羽毛球动作进行识别，所选取的三维动作特征能够有效表示各类羽毛球动作。

针对健身动作的识别问题，除了考虑所选取特征对动作的描述能力外，还应当考虑后续动作评价的可行性。健身动作评价通过捕捉人体各环节间的相对运动来评估动作执行的标准程度。人体整体位移是无关的干扰特征，如跑步时在水平方向的行进位移，跳跃时的垂直高度等。因此，提取健身动作中人体各环节间的相对运动特征不仅有利于区分相似动作，而且能为动作评价创造条件。但在以往基于传统特征提取的动作识别方法中[24-29]，很少考虑到人体运动过程中无关位移对动作识别的影响。此外，健身动作识别场景通常包含多变的背景和光照条件，而基于背景消除提取人体特征的方法对于多变背景的鲁棒性相对较差[30, 31]。随着人体姿态估计技术[32, 33]的发展，语义特征提取方法[34]为人体动作特征的提取提供了新的思路：通过提取图像中的人体骨架信息来描述动作特征并进行动作识别。提取出的人体关节位置信息是具有高度代表性的人体运动特征，能够表示动作视频中的人体活动空间分布[35]，有利于捕捉人体各环节间的组合特征[36]，且在一定程度上避免了传统特征提取方法依赖于图像分割效果的问题，对于视频中多变的背景和光照条件也具有较好的鲁棒性[34]，能够为健身动作识别任务提供具有高度代表性的人体运动骨架信息。

针对上述问题，本文提出了一种基于人体骨架特征编码的健身动作识别方法，包含三个步骤。首先，根据健身动作特点构建包含 15 个关节点的精简人体模型，并利用人体姿态估计技术[33]获取视频中的运动骨架信息。然后，通过人体中心投影消除运动过程中整体位移对识别结果的干扰，并对投影区域的轨迹特征进行缩放以降低人体体型差异对识别结果的影响并提高识别速度，通过优化函数确定缩放比例以在保证识别率的基础上获得有效特征更为集中的动作特征区域。最后，对特征区域进行线性编码以获得描述健身动作的特征向量，并设计了一个基于支持向量机（Support Vector Machine，SVM）[37]的多分类器进行模型训练和识别。为了检验本方法对健身动作的识别效果，构建了一个包含 28 种健身动作的数据集进行实验，结果表明本方法能够有效识别健身动作，识别率达到了 97.24%。在公开的 KTH（Kungliga tekniska högskolan）数据集[26]和 Weizmann 数据集[24]上，本方法的识别准确率分别达到 91.67%和 90%。

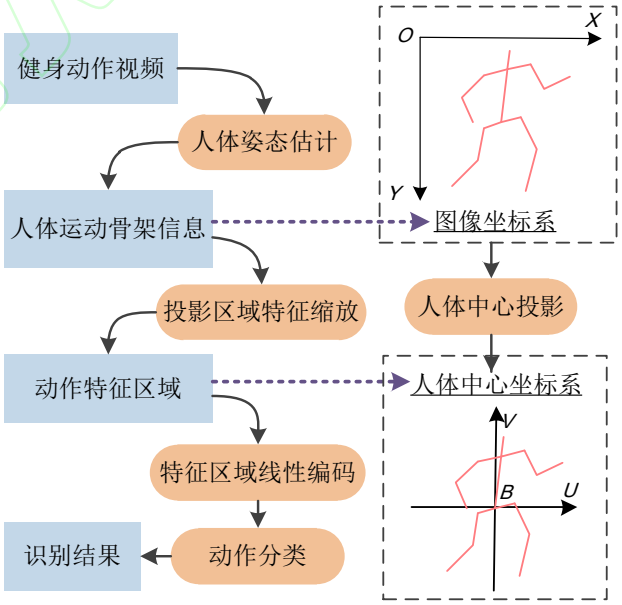

图1 健身动作识别方法流程图

Fig. 1 Pipeline of the proposed fitness action recognition method

本研究的贡献主要体现在以下三个方面：（1）通过人体中心投影法消除健身动作中人体整体位移对动作识别的影响，对利用骨架信息进行动作识别的方法具有普适性；（2）提出一种高效的骨架信息编码方法，能够有效表示健身动作，并使得方法具有较高的识别精度和速度；（3）构建了一个健身

动作数据集，能够支持健身动作识别以及后续动作评价方法的研究。

图 1 所示为本文提出的健身动作识别方法流程图，首先通过人体姿态估计技术提取运动骨架信息，然后通过人体中心投影和缩放投影区域消除干扰，最后将特征区域进行线性编码实现动作分类。本文将在第 1、2 节分别介绍人体动作特征区域的提取过程和特征编码与动作分类过程，在第 3 节介绍在各个数据集上的实验结果和相关分析，最后在第 4 节介绍本研究的结论和展望。

1 基于人体中心投影的动作特征提取

从健身动作视频中提取动作特征区域包括两个步骤：人体运动骨架信息提取和基于人体中心的动作特征区域提取。

1.1 人体运动骨架信息提取

人体骨架信息是具有高度代表性的人体运动特征。本方法利用人体姿态估计技术[33]获取运动过程中人体各环节位置信息并据此提取动作特征。根据健身动作的特点，选取包含 25 个关节点的 Body_25 人体模型进行简化，删除对动作识别贡献有限的双目特征点，双耳特征点，以及足趾和足跟关节点。获得包含 15 个关节点的精简人体模型，相比原模型更关注人体躯干和四肢的动作，有利于提高计算效率。

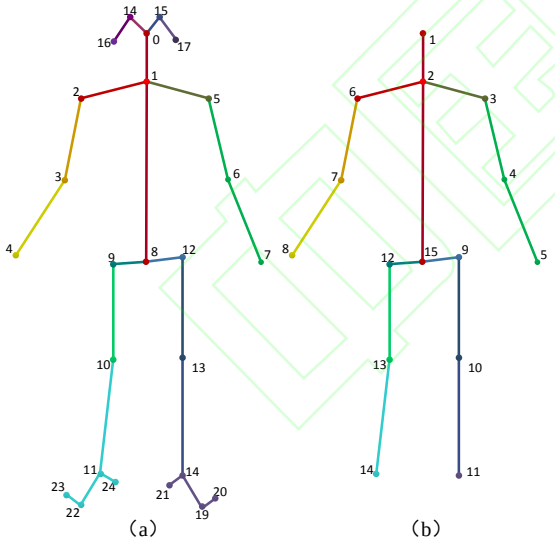

图2 模型对比。(a) Body_25 模型，(b)精简模型

Fig. 2 Comparison of models. (a) The Body_25 model, (b) The simplified model

图 2 所示为 Body_25 模型和精简人体模型和对比图。根据精简人体模型，对包含 N 帧图像的健身动作视频进行姿态估计，提取出人体关节坐标序列 $\{(x_{i,j},y_{i,j})\}$，其中 $1\le i\le N, 1\le j\le 15$，$x_{i,j}\in R$ 和 $y_{i,j}\in R$ 分别为第 i 帧中第 j 个关节点在图像坐标系中的坐标。

1.2 基于人体中心的特征区域提取

动作特征区域提取主要是基于人体中心投影，在人体中心坐标系中提取动作特征所在区域。

1.2.1 人体中心投影法

健身动作一般由人体各环节间的相对运动和人体整体位移两部分组成，其中前者是健身动作识别和评价的主要内容，反映动作执行是否标准有效，而人体整体位移通常不纳入评价体系中，对动作识别而言是无关的干扰特征。另外，健身动作识别可看作相似序列的搜索匹配问题[38]，同类动作不同样本间的时间差异会增加样本的类内差异性，从而影响动作识别结果。

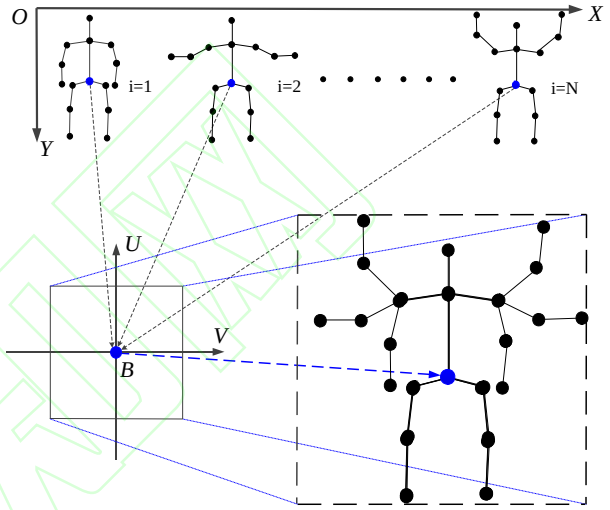

图3 人体中心投影法（*O-XY* 表示图像坐标系，*B-UV* 表示人体中心坐标系）

Fig. 3 Projection based on the body center coordinate system (*O-XY* refers to the Image coordinate system, *B-UV* refers to the body center coordinate system)

综合考虑上述因素，本文提出基于髋关节中点的人体中心投影法消除人体整体位移和动作执行时间差异。如图 3 所示，以位于人体中心的髋关节中点（蓝色关节点）作为坐标系原点建立人体中心坐标系 $B-UV$，通过投影变换获取在人体中心坐标系下的运动骨架信息。在齐次坐标下的人体中心投影过程如式（1）所示：

$$(u_{i,j},v_{i,j},1)=(x_{i,j},y_{i,j},1)\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ -x_{i,hip} & -y_{i,hip} & 1\end{bmatrix} \quad (1)$$

其中，$(u_{i,j},v_{i,j})$ 和 $(x_{i,j},y_{i,j})$ 分别为第 i 帧中第 j 个关节点在人体中心坐标系和图像坐标系下的坐标，$(x_{i,hip},y_{i,hip})$ 为图像坐标系中髋关节中点坐标。

通过人体中心投影，可以将关节坐标序列转换至人体中心坐标系，使得动作轨迹围绕人体髋关节中点分布，消除了人体整体位移和动作执行时间差异对动作识别的影响。

1.2.2 特征区域提取

特征区域提取的目的是获取动作轨迹的空间分布信息，寻找一个最小区域使其能包含全部关节坐标点。在人体中心坐标系中，假设存在一个以坐标系原点为对角线交点的正方形区域 Q，能够包含任一关节坐标 $(u_{i,j},v_{i,j})$，即满足式（2）：

$$u_{i,j}\in[-l/2,l/2],v_{i,j}\in[-l/2,l/2] \tag{2}$$

其中，l 为特征区域的边长，其取值如式（3）所示：

$$l=\max(\max(u_{i,j})-\min(u_{i,j}),\max(v_{i,j})-\min(v_{i,j})) \tag{3}$$

式（3）基于人体最大活动范围获取投影区域，原始尺寸 $l\times l$ 较大，完整保留了不同动作执行者之间的体型差异。为了降低体型差异的影响并提升动作识别算法的效率，对投影区域进一步压缩以获得更为有效的特征区域 Q'。即将投影区域缩放为一个尺寸为 $l'\times l'$ 的特征区域 Q'，在 l' 充分小的情况下能够包含足够的有效特征。假设特征区域 Q' 中包含 K 个动作轨迹特征点，则缩放投影区域变换如式（4）所示：

$$\begin{bmatrix}u_k\\v_k\\1\end{bmatrix}=\begin{bmatrix}l'/l&0&0\\0&l'/l&0\\0&0&1\end{bmatrix}\begin{bmatrix}u_{i,j}\\v_{i,j}\\1\end{bmatrix} \tag{4}$$

其中，(u_k,v_k) 表示第 k（$1\le k\le K$）个动作特征点坐标。通过投影变换，不仅获得了有效特征更为集中的特征区域表示人体动作，而且可以降低体型差异对动作识别的影响。

2 基于特征区域编码的健身动作识别

从动作特征区域中提取特征向量进行动作识别包括两个步骤：动作特征区域线性编码和基于 SVM 的健身动作识别。

2.1 特征区域编码

将特征区域进行线性编码的目的是提取动作特征向量。令 $S=\{s_k\}$ 表示含有 K 个元素的集合，s_k 为第 k 个动作特征点 (u_k,v_k) 在特征区域中的位置编码。集合大小 K 随特征区域 Q' 中包含动作特征点多少而变化，s_k 取值如式（5）所示：

$$s_k=l'\times v_k+u_k,\ 1\le k\le K \tag{5}$$

则集合 S 中包含特征区域中动作特征点的位置分布信息。基于集合 S 继续构造一个长度为 $l'\times l'$ 的特征向量 $\boldsymbol{Z}$ 表示特征区域 Q'。特征向量 $\boldsymbol{Z}$ 的初值为全零向量，根据特征区域中动作特征点的位置分布更新各元素：将 $\boldsymbol{Z}$ 中 s_k 位置的值置 1 以表示特征向量中的运动轨迹信息，其余值不变表示背景区域。经过以上步骤，可以获得一个固定长度的特征向量 $\boldsymbol{Z}$ 来表示一次动作特征。

2.2 动作识别与特征向量长度优化

2.2.1 基于 SVM 的健身动作分类

本文基于 SVM 设计了一个多类分类器对特征向量集进行分类。假设 $\boldsymbol{D}=\{(\boldsymbol{Z}_a,L_a)\}$ 是一组含有 n 个样本的特征向量集，其中 $(1\le a\le n)$，$\boldsymbol{Z}_a\in\mathbb{R}^{l'\times l'}$ 是第 a 个样本的特征向量，L_a 是第 a 个样本的类别。对样本的分类识别可以等价于一个约束最优化问题，如式（6）所示：

$$\begin{aligned}&\min\frac{1}{2}\|w\|^2+C\sum_{a=1}^{n}\xi_a\\&s.t.\ L_a(w\boldsymbol{Z}_a+b)\ge1-\xi_a,\xi_a\ge0,1\le a\le n\end{aligned} \tag{6}$$

其中，ξ_a 为松弛变量。C 为惩罚因子，取值越大对误差的容忍程度越低，相对来说容易出现过拟合，反之则容易欠拟合。针对数据集的样本量和特征向量维度，选用径向基函数（Radial Basis Function，RBF）作为核函数[37]。惩罚因子 C 和核函数参数 *gamma* 的选择决定分类器的性能，前者调整拟合和预测样本的能力，后者则与样本划分有关。在本文的实验中，通过对特征向量训练集进行网格寻优以获取最优参数（$C=64$，$gamma=0.0078125$）完成模型训练，实现动作识别。

2.2.2 特征向量长度优化

由于不同长度的特征向量中包含的动作特征点数量不同，会对识别精度和速度产生影响。在本方法中，特征区域 Q' 的尺寸 $l'\times l'$ 决定特征向量的长度。为了兼顾识别精度与速度，需要对 l' 取值进行优化。

本方法预设了一系列 l' 的离散取值，通过比较实验结果进行参数选择。对于特征向量集 $\boldsymbol{D}$，当 l' 取一定值时，将 $\boldsymbol{D}$ 中第 a 个样本识别为类别 $\tilde{L}_a$ 且 $\tilde{L}_a=L_a$ 的概率为 $p_a(l')=p(\tilde{L}_a=L_a|l')$，识别该样本的时间为 $t_a(l')$。l' 的取值应当在保证识别率的基础上提高检测速度，其优化函数如式（7）所示：

$$F(l')=\min\left(\left[1-\frac{1}{n}\sum_{a=1}^{n}p_a(l')\right]+\beta\frac{1}{n}\sum_{a=1}^{n}t_a(l')\right) \tag{7}$$

其中，β 为平衡识别精度和识别速度的权重值，在本实验中取值为 0.5。

3 实验与讨论

本节首先介绍所使用的三种数据集，然后介绍在不同数据集上的实验结果及讨论。实验均在 Inter@CoreTM i7-7700 CPU 3.60GHz 处理器，Ubuntu16.04 系统的计算机上实现。

3.1 动作识别数据集

本实验使用的三个动作识别数据集分别是健身动作数据集，KTH 数据集和 Weizmann 数据集。

健身动作数据集：该数据集使用两台主光轴相互垂直的 GoPro Hero 7 Black 对 15 名运动者进行同步拍摄，分别命名为主机位和副机位，主机位相机用于拍摄主动作特征平面。在执行不同的动作时，根据动作特点决定使用主机位拍摄运动者的矢状面或冠状面。每位受试者执行 28 种健身动作，动作分类和部分动作示例如表 1 和图 4 所示，这些动作包含力量练习，拉伸练习和综合练习，进一步可细化为器械和徒手练习，静态和动态练习。选择主机位和副机位相机拍摄的 24 人次共 5854 组（每组包含主副机位）视频作为视频数据集，数据集拍摄及受试者相关信息如表 2 所示。

表1 数据集动作分类对照

Tab. 1 Classification of actions in fitness action dataset

一级分类	二级分类	动作序号
力量练习	哑铃练习	12，13，14，15，16，25，26，28
	徒手练习	3，5，6，8，10，24
拉伸练习	动态练习	17，18，20，21
	静态练习	22，23
综合练习		1，2，4，7，9，11，19，27

表2 健身数据集信息

Tab. 2 Information of fitness action dataset

数据集参数	参数值
分辨率	1920×1440
帧率	60fps
受试者身高分布	171±10.19cm
受试者体重分布	66.79±13.46kg

KTH 数据集：包含六种人体动作（行走、慢跑、奔跑、拳击、挥手和鼓掌），由 25 名受试者在四种不同的场景下完成：室外环境、室外环境（缩放镜头）、室外环境（不同着装）和室内环境。共包含 598 段平均时长为 4 秒的视频，由固定相机拍摄完成，拍摄帧率为 25fps，分辨率为 160×120。

Weizmann 数据集：包含十种人体动作（弯腰、开合跳、跳跃移动、原地跳跃、奔跑、侧向跨步移动、单腿跳跃移动、行走、单侧挥手、双侧挥手）。该数据集由 9 名受试者拍摄完成，共包含 90 段视频，拍摄帧率为 50fps，分辨率为 188×144。

图4 健身动作数据集部分动作

Fig. 4 Sample actions in fitness action dataset

3.2 实验结果

3.2.1 健身动作数据集上的实验结果

为了验证本方法各模块对识别结果的影响，在健身数据集上分别进行了以下三部分实验：1）分别在包含主副机位拍摄动作的数据集上使用不同长度的特征向量表示动作，观察其对识别结果的影响并验证算法对拍摄视角和背景变化的鲁棒性；2）应用人体中心投影法，观察其对识别结果的影响；3）使用不同数据量的训练集训练模型，观察其对识别结果的影响，并验证本方法在较小样本量数据集上的可迁移性。

特征向量长度对识别结果的影响：为了探究特征向量长度对识别结果的影响并验证算法对拍摄视角和背景变化的鲁棒性，在包含主副机位动作视频的数据集上进行实验。随机选取 9 人次共 2062 组视频作为测试集，其余 15 人次共 3792 组视频作为训练集。分别使用长度为 16，64，144，256，400，576，784，1024 的特征向量表示动作。实验结果如图 5，图 6 所示。

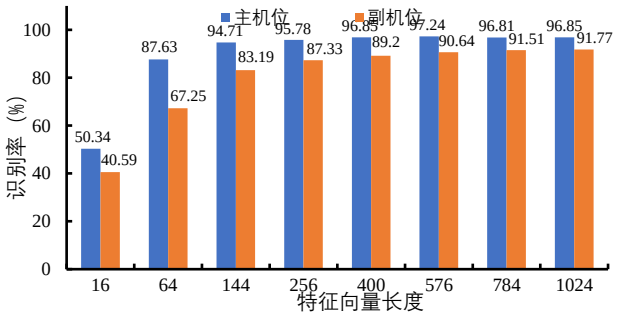

图5 不同长度特征向量的识别率对比

Fig. 5 Comparison of recognition rates with different feature lengths

对比不同特征向量长度下的识别率，长度为 16 的特征向量描述动作特征的能力较弱，难以捕捉一些位于四肢环节处的动作区分特征导致识别结果欠佳。当特征向量的长度增加到 64 时，对于大多数动作都能够较好的识别，同时主机位拍

摄动作的识别率达到接近 90%。使用长度为 144 及以上的特征向量时，识别主机位动作的正确率基本稳定在 95%以上，副机位识别率也超过 80%。当长度为 576 的特征向量作为分类器输入时，主机位识别率最高达到了 97.24%，副机位识别率也超过 90%。测试不同长度特征向量下的识别速度结果如图 6 所示，识别样本的时间与特征向量长度同向变化且上升趋势明显。识别主副机位动作在使用相同长度特征向量时识别速度相同，故图 6 只显示主机位数据集上的实验结果。综合识别精度和速度，根据式（7）特征向量长度优化函数确定参数 l' 为 24，对应特征向量长度为 576。

对比算法对主副机位拍摄动作的识别率，对副机位拍摄动作的识别率总体低于主机位。其原因有两部分：（1）相比主机位拍摄健身动作的主特征平面，副机位所拍摄平面中关节遮挡较为严重，造成提取动作特征更加困难；（2）主副机位的拍摄背景不同也会对识别结果造成影响。尽管如此，算法对副机位拍摄动作的识别率最高仍能达到 91.77%，证明方法对相机视角的变化和背景改变具有一定鲁棒性。

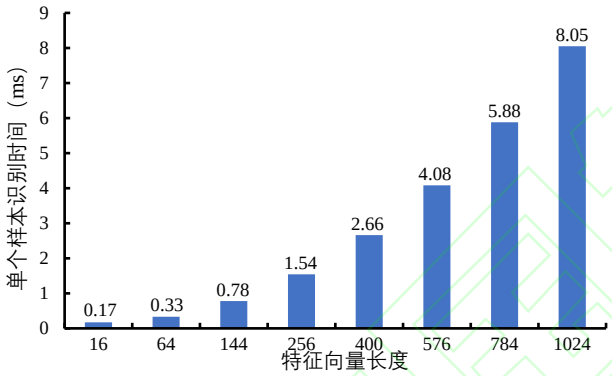

图6 不同长度特征向量的识别时间对比

Fig. 6 Comparison of recognition time with different feature lengths

人体中心投影法对识别结果的影响：为了验证所提出的人体中心投影法对识别结果的影响，对比了应用人体中心投影法和图像坐标系投影识别主机位拍摄健身动作的结果。对比结果如图 7 所示，相比图像坐标系投影，应用人体中心投影法使得动作识别率在大部分情况下都有所提升。

由于健身动作数据集中所有动作均在原地执行，运动过程中整体无关位移对动作识别造成的影响较小，故人体中心投影法对动作识别性能的提升幅度有限。

训练集大小对识别结果的影响：为了验证本文方法在数据量较小的数据集上的识别能力，在主机位拍摄的健身动作数据集中进行实验。分别使用包含 1 人次，2 人次，3 人次，7 人次，10 人次，13 人次和 15 人次动作视频的训练集进行训练，仍用包含 9 人次视频的测试集进行测试。

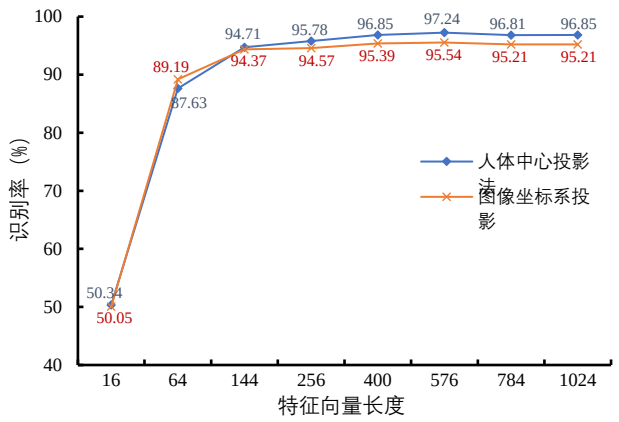

图7 人体中心投影法与图像坐标系投影识别率对比

Fig. 7 Comparison of recognition rates with the body center projection and the image coordinate system projection

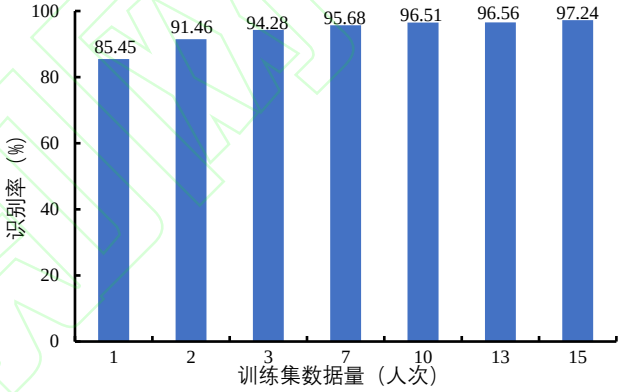

图8 不同数据量训练集的识别率对比

Fig. 8 Comparison of recognition rates with different training data sizes

实验结果如图 8 所示：识别率与训练集数据量同向变化，但随着数据量的增长，增加训练样本对于识别正确率的提升幅度逐渐下降。当使用 1 人次健身动作视频（每类动作含有 1-10 个样本不等）作为训练集时，测试能够获得 85.45%的正确率，这是由于一方面本文方法使用了最优化长度的特征向量表示动作，另一方面所使用的分类器对于小样本量数据集也有较好的识别效果。结果说明本方法在较小样本量的数据集上拥有较好的识别能力。

3.2.2 公开数据集上的实验结果

KTH 数据集：按照 Schüldt 等人[26]的方法划分测试集和训练集，训练集包含 16 名受试者的动作视频，测试集包含 9 名受试者的动作视频，用本研究提出的方法对其进行数据处理，训练和测试，实验结果如表 3 所示。

表3 KTH 数据集识别率（%）

Tab. 3 Recognition rates(%) on KTH dataset

真值	预测值					
	行走	拳击	慢跑	奔跑	挥手	鼓掌
行走	**100**	**0**	**0**	**0**	**0**	**0**

拳击	0	80	2.5	17.5	0	0
慢跑	0	0	90	10	0	0
奔跑	0	0	12.5	87.5	0	0
挥手	0	0	0	0	100	0
鼓掌	0	0	0	0	7.5	92.5

应用人体中心投影对识别结果的改善对比如图 9 所示。结果显示，对于 KTH 数据集中的六个动作，应用人体中心投影能够对除拳击动作外的五个动作进行更为精确的识别，特别是对行走，慢跑，奔跑三个动作的改进效果更为明显。这是由于在图像坐标系中，行走、慢跑、奔跑动作中大范围的人体整体位移覆盖了动作间的有效区分特征，在将其消除后识别率得到大幅提高。最终在 KTH 数据集上算法识别率达到 91.67%，应用人体中心投影法使得整体识别率提升了 14.57%。

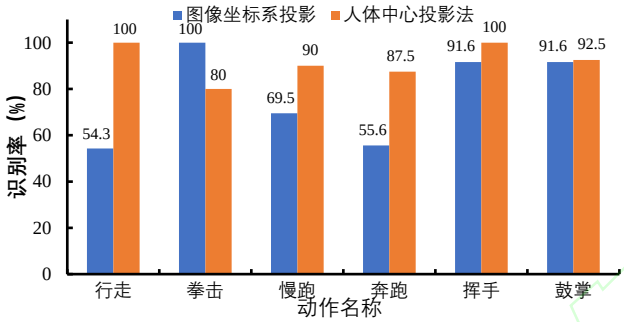

图9 应用人体中心投影法前后识别率对比

Fig. 9 Comparison of recognition rate before and after applying the body center projection

拳击动作是唯一在应用人体中心投影后识别率下降的动作类别，误识别为慢跑或奔跑动作。其原因是：在消除了较大范围的整体位移后提取出奔跑和慢跑动作的特征与拳击动作发生混淆。通过观察动作识别结果，发生误检的拳击动作中受试者出拳幅度往往较小且方向水平，与奔跑和慢跑动作的上肢环节运动轨迹相似，从而引起误检。针对此类个别动作类别间的混淆问题，后续可通过增加局部运动特征权重予以解决。

Weizmann 数据集：将数据集划分为包含 6 名受试者的训练集和包含 3 名受试者的测试集，用本文的方法进行数据处理，训练和测试，实验结果如表 4、5 和图 10 所示。除奔跑和原地跳跃两个动作外，对其余 8 个动作的识别率均为 100%，在 Weizmann 数据集上平均精度为 90%。

图 10 中呈现的四类动作在应用人体中心投影前识别率较低，通过消除运动中的整体位移使得各自的识别率得到了较大程度改善，其中跳跃移动和侧向跨步移动的识别率达到了 100%。对该数据集总体识别率提升了 16.67%。

表4 Weizmann 数据集识别率（%）I

Tab. 4 The 1st part of recognition rates(%) on Weizmann dataset

真值	预测值				
	弯腰	开合跳	行走	单侧挥手	双侧挥手
弯腰	100	0	0	0	0
开合跳	0	100	0	0	0
行走	0	0	100	0	0
单侧挥手	0	0	0	100	0
双侧挥手	0	0	0	0	100

表5 Weizmann 数据集识别率（%）II

Tab. 5 The 2nd part of recognition rates(%) on Weizmann dataset

真值	预测值				
	跳跃移动	奔跑	侧向跨步移动	单腿跳跃移动	原地跳跃
跳跃移动	100	0	0	0	0
奔跑	0	33.3	0	67.7	0
侧向跨步移动	0	0	100	0	0
单腿跳跃移动	0	0	0	100	0
原地跳跃	0	0	0	33.3	67.7

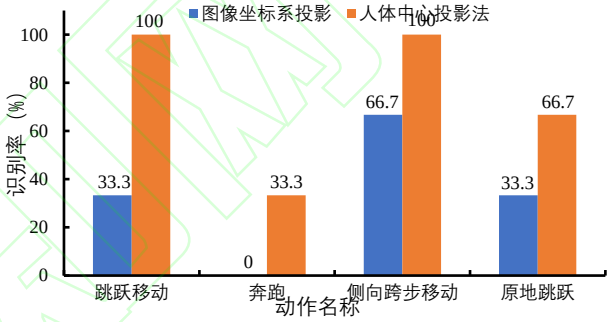

图10 应用人体中心投影法前后识别率对比

Fig. 10 Comparison of recognition rate before and after applying the body center projection

在 KTH 数据集和 Weizmann 数据集上的实验结果表明，本方法对于同类型的数据集具有较好的泛化能力，能够应用于类似的动作识别任务。另外，应用人体中心投影法能够消除健身动作中的人体整体位移从而提升识别率。尽管改进投影策略使得对 KTH 数据集中拳击动作识别率降低，但从整体上较大幅度的提高了算法对各类动作的识别率。另外，消除人体整体位移使得算法能够提取各关节间的相对运动特征完成动作识别，这对于进一步提取人体局部运动特征和进行动作质量评价是非常重要的。

本文方法与其他方法实验比较：为了验证本文方法在公开数据集上的识别能力，与其他同类型方法进行了比较。表 6 共列出四种动作识别方法在两个公开数据集上的识别结果，均为基于特征提取的动作识别方法。四类方法分别利用局部时空特征[26]，时空兴趣点[27]，随机时间规划[28]和方向梯度直方图[29]描述动作特征，并结合分类算法完成动作识别。如表 6 所示，本文提出的方法在两个公开动作识别数据集上的识别精度均高于同类型其他方法。

表6 本文方法与其他方法识别率对比

Tab. 6 Comparison of recognition rates of the proposed method with other methods

数据集	方法	平均精度
KTH	局部时空特征[26]	71.71%
	时空兴趣点+隐主题模型[27]	83.33%
	随机时间规划+格拉斯曼判别[28]	83.96%
	本文方法	91.67%
Weizmann	方向梯度直方图+K 均值聚类[29]	86.66%
	时空兴趣点+隐主题模型[27]	90%
	本文方法	90%

4 结论与展望

针对健身动作识别场景，本文结合人体中心投影法和运动骨架编码提出了一种高效的动作识别方法，能够有效且快速识别健身动作。该方法首先基于精简人体骨架模型提取人体运动骨架信息；然后通过人体中心投影提取运动特征区域，消除健身动作中人体整体位移的干扰；最后进行骨架信息编码和动作识别。在自建健身数据集和公开数据集上均获得了较好的识别效果，并证明本方法在由固定相机位拍摄的动作数据上有较好的可迁移性。所提出的人体中心投影法能够消除运动过程中人体无关位移的影响从而改善动作识别效果。在今后的研究中，将考虑关节点之间的相对关系以及人体局部运动特征，进一步提高动作识别精度并为动作评估创造条件。未来的工作将在本文的研究基础上对健身动作进行相应的评级和评分，完善智能健身指导系统。

参考文献

[1]TING H Y, SIM K S,and ABAS F S. Automatic Badminton Action Recognition Using RGB-D Sensor[J]. Advanced Materials Research, 2014, 1042: 89-93.

[2]KUO Y, LEE J,and CHUNG P. A Visual Context-Awareness-Based Sleeping-Respiration Measurement System[C]//Proceedings of the 2010 international conference of the IEEE engineering in medicine and biology society. Piscataway: IEEE, 2010: 255-265.

[3]VANDO S, HADDAD M, MASALA D,et al. Visual feedback training in young karate athletes[J]. Muscles, ligaments and tendons journal, 2014, 4(2): 137-140.

[4]JONES C, GRIFFITHS P,and MELLALIEU S D. Training Load and Fatigue Marker Associations with Injury and Illness: A Systematic Review of Longitudinal Studies[J]. Sports Medicine, 2017, 47(5): 943-974.

[5]HALSON S L. Monitoring training load to understand fatigue in athletes[J]. Sports medicine, 2014, 44(2): 139-147.

[6]吕咏. 从符号互动理论看当今社会运动健身类 APP 热的现象——以 keep app 为例[C]//第十一届全国体育科学大会论文摘要汇编. 中国江苏南京：中国体育科学学会，2019: 4330-4332. (LU Y. On the phenomenon of sports and fitness apps in current society from the perspective of symbolic interaction Theory -- Take Keep APP as an example[C]//Proceedings of the 11th national convention on sports science of China. NanJing JiangSu China:China Sport Science Society, 2019:4330-4332.)

[7]SIGRIST R, RAUTER G, RIENER R,et al. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review[J]. Psychonomic Bulletin & Review, 2013, 20(1): 21-53.

[8]王瑞元,苏全生.运动生理学[M]. 北京：人民教育出版社，2012: 295-296. (WANG R U, SU Q S. Sports Physiology[M]. Beijing: People's Education Press,2012: 295-296)

[9]LUCAS T. Exploring the effect of realism at the cognitive stage of complex motor skill learning[J]. E-Learning and Digital Media, 2019, 16(4): 242--266.

[10]HUANG C C, LIU H M,and HUANG C L. Intelligent scheduling of execution for customized physical fitness and healthcare system[J]. Technology and Health Care, 2016, 24(s1): 385-392.

[11]HUANG C C, HUANG C L,and LIU H M. Fool-proofing design and crisis management for customized intelligent physical fitness and healthcare system[J]. Technology and Health Care, 2016, 24(s1): 407-413.

[12]QI J, YANG P, HANNEGHAN M,et al. A Hybrid Hierarchical Framework for Gym Physical Activity Recognition and Measurement Using Wearable Sensors[J]. IEEE Internet of Things Journal, 2019, 6(2): 1384-1393.

[13]HAUSBERGER P, FERNBACH A,and KASTNER W. IMU-based smart fitness devices for weight training[C]//Proceedings of the 42nd Annual Conference of the IEEE Industrial Electronics Society. Piscataway: IEEE, 2016: 5182-5189.

[14]ORUCU S,SELEK M. Design and Validation of Rule-Based Expert System by Using Kinect V2 for Real-Time Athlete Support[J]. Applied Sciences, 2020, 10(2): 611.

[15]SHIH H. A Survey of Content-Aware Video Analysis for Sports[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 28(5): 1212-1231.

[16]TRAN D, BOURDEV L, FERGUS R,et al. Learning Spatiotemporal Features with 3D Convolutional Networks[C]//Proceedings of the 2015 IEEE international conference on computer vision. Piscataway: IEEE, 2015: 4489-4497.

[17]SIMONYAN K,ZISSERMAN A. Two-Stream Convolutional Networks for Action Recognition in Videos[C]//Proceedings of the 2014 advances in neural information processing systems. Montreal: The MIT Press, 2014: 568-576.

[18]KARPATHY A, TODERICI G, SHETTY S,et al. Large-Scale Video Classification with Convolutional Neural Networks[C]//Proceedings of the 2014 IEEE conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2014: 1725-1732.

[19]WANG L, XIONG Y, WANG Z,et al. Temporal segment networks: Towards good practices for deep action recognition[C]//Proceedings of the 2016 European conference on computer vision. Cham: Springer, 2016: 20-36.

[20]ZHU W, HU J, SUN G,et al. A key volume mining deep framework for action recognition[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 1991-1999.

[21]左国玉，徐兆坤，卢佳豪,等. 基于结构优化的 DDAG-SVM 上肢康复训练动作识别方法[J]. 自动化学报，2020, 46(03): 549-561. (ZUO G Y, XU Z K, LU J H, et al. A Structure-optimized DDAG-SVM Action Recognition Method for Upper Limb Rehabilitation Training[J]. Journal of Automatica Sinica, 2020. 46(03): 549-561.)

[22]闫航，陈刚，佟瑶,等. 基于姿态估计与 GRU 网络的人体康复动作识别[J]. 计算机工程，2020. https://doi.org/10.19678/j.issn.1000-3428.0058201. (YAN H, CHEN G, TONG Y, et al. Rehabilitation Action Recognition Based on Pose Estimation and GRU Network[J]. Computer Engineering, 2020.)

[23]LI H, TANG J, WU S,et al. Automatic Detection and Analysis of Player Action in Moving Background Sports

Video Sequences[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2010, 20(3): 351–364.
[24]BLANK M, GORELICK L, SHECHTMAN E,et al. Actions as space-time shapes[C]//Proceedings of the 10th IEEE International Conference on Computer Vision. Piscataway: IEEE, 2005: 1395-1402.
[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]SCHüLDT C, LAPTEV I,and CAPUTO B. Recognizing human actions: A local SVM approach[C]//Proceedings of the 2004 International conference on pattern recognition. Piscataway: IEEE, 2004: 32-36.
[27]NIEBLES J C, WANG H,and FEIFEI L. Unsupervised Learning of Human Action Categories Using Spatial-Temporal Words[J]. International Journal of Computer Vision, 2008, 79(3): 299-318.
[28]SOUZA L S, GATTO B B,and FUKUI K. Enhancing discriminability of randomized time warping for motion recognition[C]//Proceedings of the 2017 international conference on machine vision. Piscataway: IEEE 2017: 77-80.
[29]THURAU C. Behavior histograms for action recognition and human detection[C]//Proceedings of the 2007 Workshop on Human Motion. Cham: Springer, 2007: 299-312.
[30]KUMARI S,MITRA S K. Human Action Recognition Using DFT[C]//Proceedings of the 2011 computer vision and pattern recognition. Piscataway: IEEE, 2011: 239-242.
[31]CHERLA S, KULKARNI K, KALE A,et al. Towards fast, view-invariant human action recognition[C]//Proceedings of the 2008 computer vision and pattern recognition. Piscataway: IEEE, 2008: 1-8.
[32]FANG H, XIE S, TAI Y,et al. RMPE: Regional Multi-person Pose Estimation[C]//Proceedings of the 2017 IEEE international conference on computer vision. Piscataway: IEEE, 2017: 2353-2362.
[33]CAO Z, SIMON T, WEI S,et al. Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields[C]//Proceedings of the 2017 computer vision and pattern recognition. Piscataway: IEEE, 2017: 1302-1310.
[34]BUX A, ANGELOV P,and HABIB Z. Vision based human activity recognition: a review[C]//Proceedings of the 2017 advances in Computational Intelligence Systems. Cham: Springer, 2017: 341-371.
[35]YAO A, GALL J, FANELLI G,et al. Does human action recognition benefit from pose estimation?[C]//Proceedings of the 2011 British Machine Vision Conference. Piscataway: IEEE, 2011: 1-11.
[36]YAN S, XIONG Y, LIN D,et al. Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition[C]//Proceedings of the 2018 national conference on artificial intelligence. New Orleans: AAAI, 2018: 7444-7452.
[37]HSU C, CHANG C,and LIN C. A practical guide to support vector classification[J]. BJU International, 2008, 101(1): 1396-1400.
[38]RAKTHANMANON T, CAMPANA B, MUEEN A,et al. Searching and mining trillions of time series subsequences under dynamic time warping[C]//Proceedings of the 18th international conference on Knowledge discovery and data mining. New York: ACM, 2012: 262-270.

This work was supported by National Key Research and Development Project (2018YFC2000600), the Fundamental Research Funds for Central Universities (2020056) and the Fundamental Research Funds for Central Universities (2020010).

GUO Tianxiao, born in 1996, M. S. candidate. His research interests include intelligent sports and sports video analysis；

HU Qingrui, born in 1996, M. S. candidate. His research interests include intelligent sports and sports video analysis.

LI Jianwei, born in 1987, Ph. D., lecturer. Her research interests include SLAM, computer vision and intelligent sports.

SHEN Yanfei, born in 1976, Ph. D., professor. His research interests include Artificial Intelligence Technology, Intelligent video analysis and sports big data.