

基于端云协同的商用设备运动纠正系统的设计与实现^①

徐 渊^{②*} 刘志刚^{**} 王 卅^{*} 包云岗^{*}

(* 中国科学院计算技术研究所 北京 100190)

(** 中国科学院大学 北京 100049)

摘 要 介绍了一种 GymBeats 运动纠正系统原型,该系统展示了用户如何独自通过商用设备校正运动中身体姿势的错误。具体而言,本文提出了一种新颖的上下文感知运动纠正方法,包括状态识别和错误识别。此外,开发了一套基于端云协同的系统,并通过一种典型的全范围练习运动来描述系统的可行性。本文面向 15 位用户对 GymBeats 的 2 种反馈形式进行评估,即实时语音反馈和可视化报告反馈。实验结果表明,GymBeats 可以以低成本和便捷的方式来提高身体姿势的正确性。

关键词 运动纠正;商用设备传感器;骨架提取;上下文感知识别;实时语音反馈;可视化报告反馈

0 引言

反馈在人们运动锻炼过程中扮演着十分重要的角色,因为它可以识别出运动中身体姿势的正确性并纠正错误。但在多数情况下,人们几乎无法获取系统性的反馈来帮助他们纠正身体姿势,他们获取反馈的唯一信息来源可能是一面无法提供任何专业指导的镜子,他们需要分散注意力来观察镜中的自己,这进一步地降低了运动质量。

尽管近几年有很多运动纠正相关的研究来帮助经验不足的用户提高运动质量,但是它们提供的方法既不便捷、也不精确。一方面,一些现有的工作通过基于深度信息的光学传感器获取骨骼信息,并根据这些信息估计不同部位的关节角度^[1-2]。这种方法通常只能提取人体骨骼正视图,并测量运动涉及到的关节,因此它们不够精确。另一方面,一些基于传统光学跟踪系统^[3-4]和基于惯性测量单元(inertial measurement units, IMU)的跟踪系统^[5-6]通过在重要

的人体解剖位置附加光学标记或者 IMU 传感器,从而提供精确的运动信息。但是,这些方法需要使用的传感器价格昂贵,很难在普通用户中普及。

实现利用智能商用设备上的传感器纠正运动错误姿势的挑战在于运动纠正应用的计算需求与商用移动设备硬件资源的计算能力之间的不匹配。智能手机和平板电脑等移动设备依靠单目相机只能生成不带有深度数据的 2D 图像,而从这些 2D 图像中提取骨架信息需要应用深度神经网络(deep neural networks, DNN),这导致在精确性和响应性的约束下,骨架信息提取过程对硬件资源的需求很高。然而,移动设备为了满足功耗和成本的约束,计算能力往往非常有限,只能进行轻量级的运算。因此,研究人员对在商用移动设备中运行此类人工智能应用程序的性能提出了担忧^[7]。

为了应对这一挑战,本文证明了可以通过构建物端云端协同计算环境来满足上述约束。具体而言,基于云服务商提供的强大的服务器,可以将计算密集型的骨架提取任务卸载到云端服务器中,从而

① 国家重点研发计划(2016YFB1000201),国家自然科学基金(61702480, 61672499)和中国科学院青年创新促进会(2013073, 2020105)资助项目。

② 男,1991 年生,博士;研究方向:计算机系统结构,人工智能,性能优化;联系人,E-mail: xuyuan@ict.ac.cn。
(收稿日期:2021-02-27)

获得实时分析结果。本文的实验进一步表明,尽管本地设备和远程服务器之间的数据传输会引入网络延迟,但是这些额外的开销是可以接受的,并不会打破端到端实时性的约束。

基于上述观察,本文设计了一个基于端云协同的 GymBeats 运动纠正系统,该系统可以为没有经验的用户提供专业的运动纠正反馈。首先将缩放后的 2D 图像从智能手机传输到云服务器,然后从每个图像中提取身体关键点。为了验证动作的正确性,设计了一种新颖的上下文感知运动纠正方法,将从身体关键点获取的特征与本文动作库中每个动作的状态进行匹配。该方法包括 2 个阶段,即状态识别阶段和错误识别阶段。状态识别阶段用于估计练习中每个动作的状态,包括站立、下蹲、底部、站起。错误识别阶段用于检查用户动作与标准规则是否匹配。

受 GymSoles^[8]和 FEMO 启发^[6],本文提供了 2 种反馈机制,即基于语音的实时反馈和基于可视化报告的运动后反馈。本文访问了大学校园内的一群健身爱好者来确定用户需求,并通过咨询一些专业教练来设计不同运动的匹配规则。

1 研究动机

本文建议使用商用设备来进行动作纠正,主要是因为当前的动作纠正方案并不能完全满足用户的需求。本节首先通过调研确定了用户对于运动反馈系统的真实需求,然后逐个分析现有解决方案的问题,最后介绍了本文的 GymBeats 方案。

1.1 用户需求

为了解用户在运动时的实际需求,在大学校园内对一群健身爱好者进行了半结构化访谈,让他们说出理想的运动纠正设备应当具备的功能,最终总结了 3 个最主要的功能如下。

便携的廉价设备(portable cheap device, PCD)。在收到的所有回复中,最关心的问题就是设备的便携性和价格。首先是设备价格不能太贵,假如设备太贵的话,受访人通常情愿花钱请私教,这样可以获取更人性化的指导。此外,他们希望设备具有便携性,以便他们可以在任何地方使用,例如健身房或家

里。

丰富的运动库(rich exercises benchmark, REB)。考虑到单一锻炼对肌肉刺激的局限性以及锻炼种类的多样性,受访人希望随着他们的需求变化,系统可以随之不断丰富锻炼动作类型。因此该软件应当可以识别各种运动类型并提供纠正建议,从而指导用户。

专业的反馈(professional feedback, PF)。本文的目标用户是非专业的初学者,他们往往需要专业的反馈意见,这样可以帮助他们提高动作质量。因此,该软件应当能够准确识别每个错误并提供实时反馈。

1.2 相关工作

先前关于运动纠正的方法主要包括使用光学跟踪相机、基于惯性测量单元(inertial measurement unit, IMU)的传感器和压敏设备。如表 1 所示,本小节根据上述 3 个要求进一步讨论当前技术存在的问题。

表 1 本文方法与其他运动纠正相关工作比较

		PCD	REB	PF
光学运动跟踪	基于光学标记		√	√
	基于光学深度 GymBeats		√	√
基于 IMU 的运动跟踪		√	√	
压力数据跟踪				√

1.2.1 光学运动跟踪

光学运动跟踪系统根据传感源不同可以分为 2 类。第 1 类是基于光学标记的系统,例如 Vicon^[3]和 OptiTrack^[4]。这些商业系统通过将光学标记安装在用户的重要解剖位置(例如关节)来提供高精度的跟踪结果。但是,这种高精度的反馈需要一套昂贵的配件^[9],包括镜头、光学标记及其配套软件。普通用户通常负担不起这么高昂的价格。第 2 类是基于光学深度信息的系统。这些系统使用基于深度信息的设备(例如 ASUS Xtion^[1]和 Microsoft Kinetic^[2])从获取的深度图像中检测骨骼位置,并推断每个像素的身体部位的分布^[10]。基于骨骼信息,这些系统可以提取与关节角度有关的信息,并用这些信

息匹配标准运动姿势。开源和无需使用光学标记的特性使此类系统广泛用于各种训练的纠正^[11]。然而基于光学深度的系统具有3个问题。首先,无论是Kinetic还是Xtion都尚未广泛普及,并且由于用户需要将其连接到笔记本电脑上使用,所以并不便携。其次,仅凭借正面测量出的角度信息不足以识别一些运动错误,例如后仰。最后,这种系统提供的是视觉的实时反馈,这会在用户运动时分散注意力。

1.2.2 基于IMU的运动跟踪

基于IMU的运动跟踪系统提供了基于商用可穿戴设备(例如智能手表和手环)的便携式廉价解决方案。通过分析多轴加速度计的测量数据,这些系统能执行活动识别^[12-13]、步态分析^[14-15]和健康监测^[16-17]等任务。此外,一些工作着重于使用IMU数据来检测和协助运动^[18-19]。例如,FEMO使用附着在哑铃上的无源RFID标签的反射后向散射信号的多普勒频移曲线来识别现场的自由重量训练,评估动作质量,并向用户提供有用的反馈。与光学运动跟踪系统不同,基于IMU的运动跟踪系统要实现高精度,就需要在身体上部署多个IMU传感器,并进行一些额外的工作,包括广泛的校准以及精确地将每个传感器安放到身体的特定部位。通常,只从智能手表获取到的IMU的数据不足以准确跟踪身体姿势^[20]。

1.2.3 压力数据跟踪

压力数据跟踪系统通过测量地面反作用力和压力中心(center of pressure, CoP)的位移来分析平衡性和脚部动力学信息。这些数据从压敏设备(如测力板和支撑智能鞋垫)中采集得到。由于成本昂贵

和缺乏便携性,压力数据跟踪系统主要用于实验室的康复研究^[21]和步态分析^[22]。文献[8]提出了一种基于智能鞋垫的运动纠正系统GymSoles,该系统基于脚部CoP提供震动和视觉反馈,从而纠正下蹲和举重时的身体姿势。尽管压力数据跟踪系统可以实现准确的CoP跟踪数据,但此跟踪系统只能用于某些与压力有关的特定运动,例如下蹲和硬拉。然而,大部分其他运动都无法通过CoP进行分析,例如哑铃前举、哑铃侧举和杠铃卧推。

2 系统概述

本节简单介绍了GymBeats及其总体设计架构和运行时的工作流程。

2.1 GymBeats

随着计算机视觉技术的飞速发展,越来越多的研究工作开始关注如何利用卷积神经网络(convolution neural network, CNN)算法从没有深度信息的2D图片中检测出姿势信息^[23-24]。2D图片可以通过智能手机和笔记本电脑等商用设备轻松获取。基于这些工作,本文提出了GymBeats系统(见图1),这是一种使用商用设备的运动纠正系统,并为用户提供专业的反馈。智能手机可以通过将图像发送到云端进行动作分析。GymBeats系统首先识别运动类型,从2D图像中提取骨骼信息。接着,它在丰富的运动动作库中搜索该运动类型的规则信息,并与实时骨骼信息进行匹配。最后,用户可以通过系统中2个反馈机制(实时语音反馈和可视化运动报告反馈)进行身体姿势的纠正。

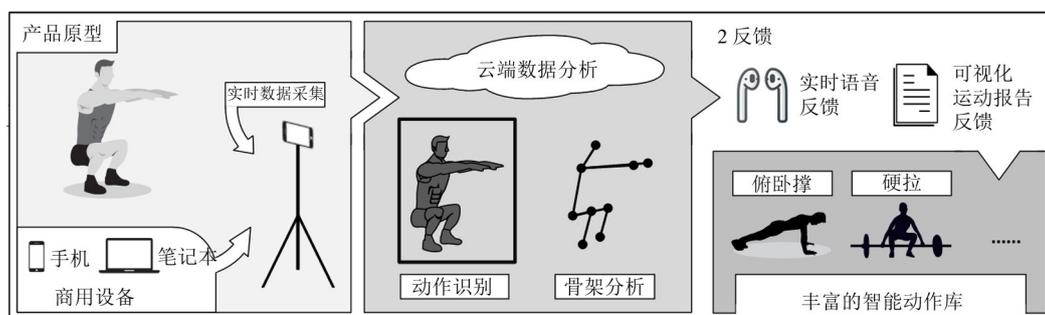


图1 使用智能手机的相机进行动作纠正应用程序实例

2.2 GymBeats 总体架构

GymBeats 框架是一个端-云协同系统(图2),其数据和代码分布在云端服务器(灰色虚线方块)和用户移动设备(灰色实线方块)上。框架中包含3个基础处理阶段:骨架提取阶段、上下文感知识别阶段和评估反馈阶段。在骨骼提取阶段,系统会处理传感器数据,提取出锻炼者的身体关键点位置以及

一些基本运动特征。上下文感知识别阶段会通过提取出的信息验证动作的正确性,并将两种不同形式的反馈结果反馈到用户的移动端设备上。

2.3 GymBeats 工作流程

图2显示了 GymBeats 框架的工作流程。本文详细描述执行的每个步骤(下列数字与图2中的数字一一对应)。

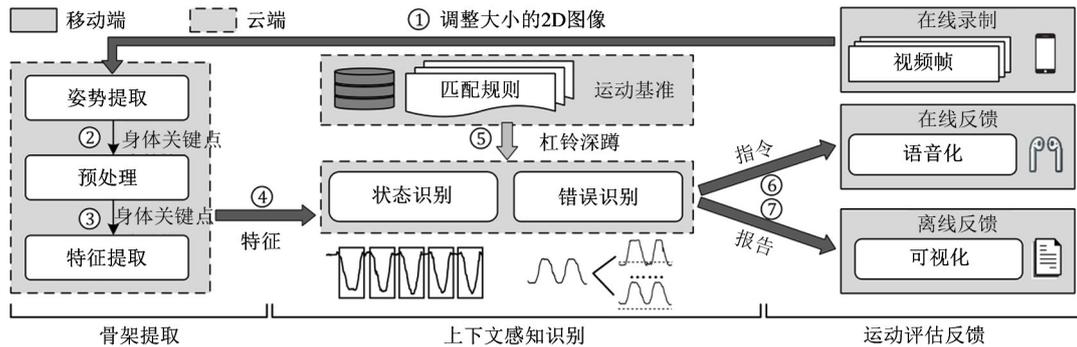


图2 GymBeats 框架总体架构

① 用户在锻炼过程中通过固定的智能手机或笔记本电脑对运动过程中每一个动作进行录制。GymBeats 会将视频中的每一帧从移动端设备发送到云端服务器中进行处理。

② GymBeats 在云服务器中利用 OpenPose^[25] 框架识别出用户每一个动作的姿势,并在每一帧中生成一组用户身体的关键点。OpenPose 是一套基于深度神经网络 DNN 的实时多人 2D 姿势估计框架,集成了许多先进的视觉技术。本文选择该框架出于以下 2 个原因。第一,相较于 Alpha-Pose^[26] 和 Mask R-CNN^[27] 等复杂高精度模型,OpenPose 对多人目标检测进行了优化。这使得 OpenPose 不会像其他复杂模型一样其计算开销会随着场景中人数的增多而呈线性增长,这对于健身房这种频繁有人在镜头前走动的开放场景提供了实时性保障。第二,虽然针对移动端深度优化的 TensorFlow Lite^[28] 框架可以在手机芯片中提供实时的检测信息,但是由于模型精度不高,在本文的实验中发现无法用该模型成功检测出人体侧面和背面的关键点,因此并不适用于本文的需求。综上,本文发现只有 OpenPose 可以在提供高精度的前提下保障实时性。

③ 由于用户在拍摄过程中可能会存在噪声和

遮挡,例如一帧中存在其他目标或者器械挡住用户,因此 GymBeats 对每一帧目标用户的身体关键点进行预处理操作,从而保证运动纠正过程的精确性。

④ GymBeats 会从预处理后的身体关键点中提取运动相关的特征值并发送给识别模块。这些特征值与运动的类别相关,例如膝盖角度和髌骨角度等。

⑤ GymBeats 通过在云端建立一个运动基准库存储每一个类型的动作及其与该动作相应的正确性匹配规则。这些从每帧中提取的特征会与运动基准库中的正确性动作规则进行匹配,从而验证每个动作的正确性。匹配过程包含特定运动的状态识别和错误识别。

⑥ 对于每一帧,GymBeats 都会提供实时的语音反馈,从而帮助用户及时纠正其动作,防止运动造成的身体损伤。

⑦ 在用户完成所有动作后,GymBeats 还会把一份可视化运动报告发送到用户的移动端设备上。该报告包含该运动锻炼过程中的各个性能指标,帮助用户系统性地了解自己每个动作的规范性。

3 系统实现

本节详细描述 GymBeats 框架的系统设计和各

个模块面临的挑战和解决方案。3.1节重点展示预处理组件如何处理实际环境中噪声和遮挡问题。3.2节通过深蹲实例介绍本文的状态识别和错误识别组件如何利用基准库中的规则进行匹配。3.3节将介绍运动评估反馈模块如何实现实时语音反馈和运动后的可视化报告反馈。

3.1 骨架提取模块

骨架提取模块负责从用户上传的视频中的每一帧提取目标身体的关键点并转换为特征值。在这个过程中最大的挑战就是噪声和遮挡问题。

(1) 噪声问题。在真实运动环境中, GymBeats通过商用设备获取到的数据可能存在两种噪声干扰。第1种噪声是多用户干扰。因为在健身房中经常会有大量人群来回走动, 本文必须要能够从这些运动的人中识别出目标用户, 并在整个过程中跟踪目标的运动动作。第2种噪声是运动方向干扰。由于不同运动匹配的规则依赖于运动的方向(例如正面或者侧面相机), 因此, 必须可以自动地确定用户运动的方向并提取相应的身体关键点。

针对多用户干扰噪声, 本文通过估计用户的相对相机的位置来识别出目标用户。这是由于目标用户不同于来回走动的噪声人群, 其通常在一个地方相对静止地完成一系列动作。因此, 目标用户相对相机的距离也基本没有太大变化。针对运动方向噪声, 本文通过分析提取到的身体关键点的分布来确定目标相对相机的方向。例如, 相比于正面, 侧面方向识别出双肩关键点的距离要小得多。

(2) 遮挡问题。遮挡是现实中最常见的问题。例如, 当用户左侧面向相机录制视频时, 可以从目标用户的右肩和右膝盖处提取到身体关键点。但是, 当用户下蹲时, 杠铃片会遮盖用户的小腿, 这大大降低该部位关键点提取的准确性。

针对遮挡干扰, 本文采用自适应关键点的预测方法解决。该方法是基于提取到的关键点置信度、特定运动的先验信息和用户方向这3个信息。具体来说, 当身体关键点模糊或不可见时, DNN模型生成的置信度将会变得很小, 甚至为零。因此, 当置信度降至某个阈值以下时, 可以认为关键点无效。由于杠铃只覆盖部分身体关键点(例如下蹲时的肩膀

关键点), 因此可以利用其他高置信度的关键点来预测被遮挡的关键点位置。此外, 本文还选择与用户方向相同的身体关键点作为主要参考依据。例如, 当智能手机位于用户左侧时, 可以认为从身体左侧部位提取的关键点更为可信。

3.2 上下文感知识别模块

上下文感知识别模块负责将骨架提取模块识别出来的特征值与云端基准库中相应运动的规则进行匹配, 从而识别出用户动作的正确性。本文以杠铃深蹲为例来描述此前相关工作的缺陷并介绍如何解决这些问题。选择该运动的原因有两个。首先, 杠铃深蹲是一种全方位的力量运动, 可以直接训练核心和背部肌肉群, 包括脊柱直肌、绳肌、臀大肌、内收肌、股四头肌和斜方肌^[29]。由于杠铃深蹲非常复杂, 可以借此对本文的方法进行全面的描述。其次, 杠铃深蹲是一种危险的力量运动, 动作错误和杠铃重量过大都会对锻炼者造成严重损伤。因此, 杠铃深蹲是最近的工作中主要关注的锻炼之一^[8,11]。

尽管有许多书籍介绍杠铃深蹲的动作规范, 但是本文遵循权威的杠铃训练书籍, 即《Starting Strength Basic Barbell Training》^[30]。在描述需求规范之前, 首先对负重下的身体运动系统做一个必要介绍。当把用户和杠铃视为同一个系统时, 系统的重心(center of mass, COM)将从髋骨的中部位置移动到背部负重杠铃中部。当用户本身负载更重的杠铃时, 用户-杠铃系统的COM可以近似为杠铃自身的中心点。因此, 本文要对杠铃进行很大范围的移动时, 考虑杠铃本身的平衡性就非常重要。

脚中点位于足弓下方, 是身体保持平衡的关键位置。图3显示了在用户-杠铃系统中背部杠铃和

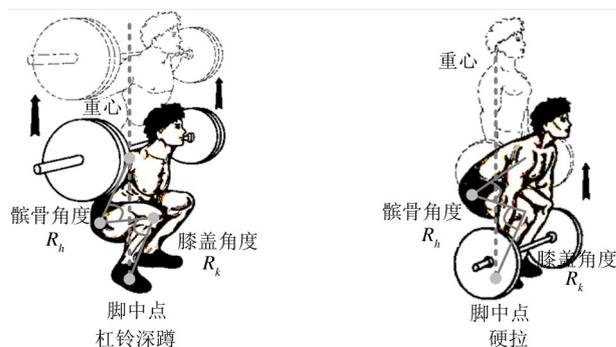


图3 杠铃系统的诊断信息

脚中点之间的关系,其中, R_h 和 R_k 分别表示运动员的髋骨角度和膝盖角度。显然,当脚中点位于 COM 的垂直线上时,系统将达到平衡。此外,任何杠铃运动的本质都是抵抗重力的向下做功。因此,最高效的杠铃移动姿势是保持 COM 与脚中点形成一条直线,仅垂直移动杠铃。

杠铃深蹲 本文在图 4 中总结了 4 个杠铃深蹲的基本要求规范。其中,虚线表示 COM,黑色实心圆圈表示杠铃。

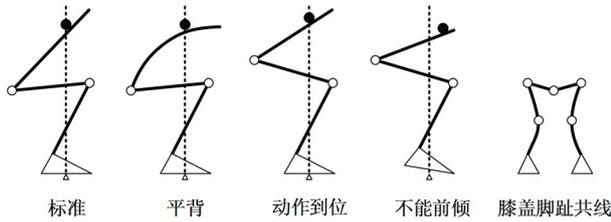


图 4 杠铃深蹲的基本要求规范

(1)平背。深蹲最大危险是腰部受伤。这些损伤是由于椎骨在负重作用下运动不当,向前或者向后挤压椎间盘造成的。用户应当保证背部平坦,因为平坦的背部可以防止腰椎在负重过程中过度伸展或弯曲,并确保力在椎间盘上正确分布。

(2)动作到位。深蹲的深度是影响关节稳定性与安全性的的重要因素。正确的下蹲动作要求臀部下降到膝盖骨顶端以下的位置,从而使股四头肌产生向前力与腘绳肌产生的向后力保持平衡。在此深度以上的深蹲称为部分深蹲。该动作不会给臀大肌、内收肌或腘绳肌施加压力,而是给膝盖和四头肌施加压力。此外,由于运动范围短,部分深蹲会让用户负载更大的重量,从而增大脊柱过度负重导致的背部损伤风险。

(3)不能前倾。在下蹲过程中,用户的膝盖会

向前弯曲,目光前视。因此,当身体失去平衡时,用户会本能地前倾。失衡会使 COM(杠铃)和脚中点之间的水平偏移产生力矩(旋转力)。用户必须提供一定的力来抵消旋转力的影响并重新获得平衡性。因此,在整个负载运动中,用户必须确保 COM 始终与脚中点在一条垂直线上。

(4)膝盖脚趾共线。杠铃深蹲过程中和膝盖相关的错误类型通常有两种:膝盖内扣和前伸。本文着重于第 1 类错误类型。因为第 2 类错误类型取决于用户的股骨、胫骨和躯干尺寸。对于一些股骨短而胫骨长的人,膝盖是可以在脚趾的前面。而对于第 1 类错误类型,当膝盖彼此向内移动时,股骨内侧和外侧的肌肉功能会减弱,这会导致用户失去平衡并磨损膝盖软骨。因此,正确的动作应是保持与脚趾在一条直线上。

基准库动作规则 尽管此前有相关工作^[11]通过考虑髋骨角度、膝盖角度以及这两个角度之间的比率来分析运动姿势错误。但是这些粗粒度方法无法准确识别出所有动作错误。例如,针对膝盖脚趾共线的错误识别,则需要考虑膝盖之间的距离而不是角度。此外,很多动作在不同状态下匹配的规则也是不同的。例如,动作到位错误的识别首先需要检查用户是否已经蹲到底部。为了解决这些问题,本文设计了一种新颖的上下文感知运动纠正方法(参见图 5)。如表 2 所示,除了传统的膝盖角度 R_k 和

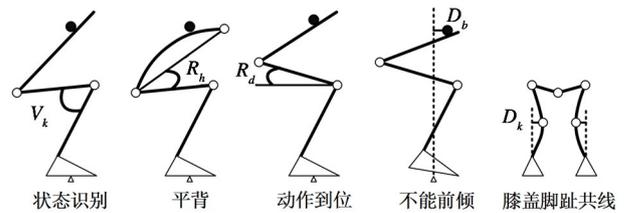


图 5 针对不同错误的上下文感知运动纠正方法

表 2 上下文感知运动纠正方法的度量值

度量值	功能	描述
R_k	状态识别	膝盖角指胫骨和股骨的平面夹角
V_k		膝盖角速度定义为膝盖角的变化率
R_h	平背	髋骨角指躯干和股骨的平面夹角
R_d	动作到位	深度角指躯干的平面和膝盖的水平线夹角
D_b	不能前倾	杠铃距离定义为杠铃与标准 COM 的偏移量
I_b		杠铃距离指数定义为杠铃距离与人的身高比
D_k	膝盖脚趾共线	膝盖距离定义为膝盖关节到脚趾垂直线的偏移量
I_k		膝盖距离指数定义为 D_k 与人的身高比

髌骨角度 R_h 外,本文还提出了一些新的识别指标,包括膝盖角速度 V_k 、深度角 R_d 、杠铃距离 D_b 和膝盖距离 D_k 。这些指标在被提取后会与基准库中相应动作的规则进行匹配。整个匹配过程包含两个阶段:状态识别和错误识别。

状态识别 仍然以杠铃深蹲为例。通常在整个运动过程中,用户可能处于以下 4 种状态之一:站立、下蹲、底部、站起。本文同时使用膝盖角度和膝盖角速度来识别负重过程中的每个状态。具体而言, V_k 为正值表示用户处于站起状态,为负值表示处于下蹲状态。当用户开始/结束下蹲并处于站立状态时, V_k 和 R_k 分别接近 0 rad/s 和 180° 。如果在上述条件下 R_k 远小于 180° ,则意味着用户处于底部状态。

错误识别 对于每一种错误类型,采用以下纠正方法。

(1) 平背。通过追踪底部状态下髌骨角 R_h 识别。如果髌骨角接近 40° 或更大,则认为该动作符合平背要求。当背部向内弯曲时,胫骨平面的高度将会低于标准高度,使得髌骨角明显减小。

(2) 动作到位。通过追踪底部状态下深度角 R_d 识别。如果深度角接近 10° 或更小,本文认为该动作满足动作到位要求。这里选择 10° 而非负值,是因为相对真实髌关节位置,从 2D 图像中提取的髌关节关键点存在一定的偏移。

(3) 不能前倾。通过追踪站起状态下杠铃距离指数 I_b 识别。如果杠铃距离指数在 $-0.05 \sim 0.05$ 范围内波动,本文认为该动作满足不能前倾的要求。正如 3.2 节讨论的那样,杠铃和脚中点之间距离偏移会导致用户失去平衡。由于杠铃距离 D_b 会随着相机和用户之间的距离发生变化,因此需要用该距离与用户身高的比值。

(4) 膝盖脚趾共线。通过追踪用户全程状态下膝盖距离指数 I_k 识别。如果膝盖距离指数达到 0.05 或者更大,本文认为该动作满足膝盖脚趾共线的要求。这里不直接使用膝盖距离 D_k 来估计该误差的原因与 I_b 相同。

3.3 运动评估反馈模块

运动评估反馈模块是向用户的移动端设备提供

运动的正确性匹配结果。根据 3.2 节中讨论的需求调查,本文发现专业的运动评估反馈是帮助用户改善运动姿势的关键要素之一。因此,GymBeats 提供 2 种反馈。

(1) 语音反馈。GymBeats 会将上下文感知识别阶段的生成结果转化为语音指令并发送到用户的移动端设备中。例如,当识别出运动动作不到位时,GymBeats 会发送一条语音指令到用户手机,从而提醒其需要再向下再蹲一点。收到指令后,本地智能手机上的应用程序将调用文本语音引擎 (android text to sound, TTS) 接口来发出语音提示。

在运动期间,为了确保安全性和人机交互的友好性,GymBeats 需要能够实时识别错误并对错误做出足够迅速的响应。假如用户在杠铃深蹲时接收到过时的语音命令,并按照命令指示进行动作,则很可能导致受伤。基于 DNN 的 2D 姿态提取模型需要的强大计算能力,但是智能手机中有限的硬件资源并不能予以支撑。如图 6 所示,在本地智能手机上处理一帧时,端到端延迟高达 15.7 s 。此外,分辨率为 1920×1080 的原始 RGB 图像的大小为 5.93 MB ,在每秒 30 FPS 的处理频率下会浪费过多的带宽。

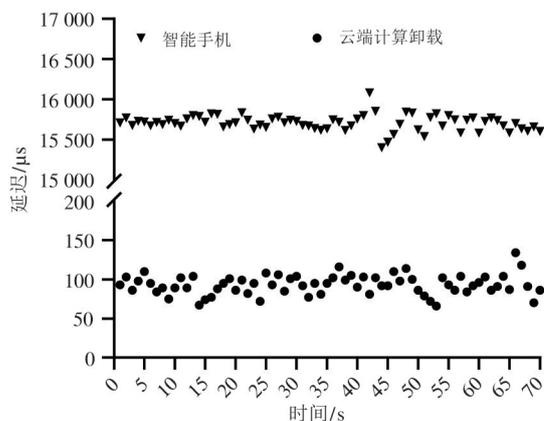


图 6 从发送一个视频帧到接收到一条语音反馈的端到端延迟

为了实现实时语音看反馈,需要减少每帧的处理时间并缩短两帧之间的间隔时间(即提高帧速率)。一方面,GymBeats 将姿势提取的所有计算工作转移到功能强大的远程服务器上,并将端到端的等待时间显著减少到 93.2 ms 。另一方面,姿势提取

模型所需的图像分辨率固定为 300×300 。在此基础上,GymBeats 首先使用 OpenCV 库将原始图像尺寸从 1920×1080 调整为 300×300 分辨率,然后将原始图像压缩为 40.76 kB jpeg 格式的图像。如图 7 所示,GymBeats 在本地智能手机上以 30 FPS 的速率发送视频帧,并测量手机发送帧的带宽和在远程服务器中接收帧的速率。从图中可以看出,调整大小操作可以将带宽从 28.1 Mbps(三角点)显著减少到 7.4 Mbps(圆点),并将帧率从 4.3 FPS(三角点线)增加到 24.5 FPS(圆点线)。

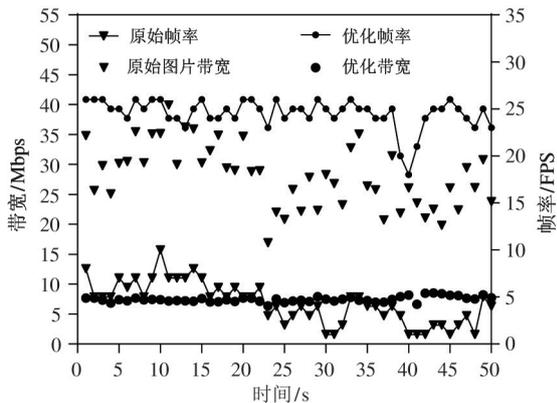


图7 本地手机发送视频帧带宽和在云端服务器接收帧率

(2) 可视化反馈。本文的可视化反馈以视频报告的形式生成,并在用户运动完后可以下载。具体而言,在处理完每个视频帧并在上下文感知识别阶段生成结果之后,本文将结果附加到每个视频帧上,并将其存储在云中。图 8 显示了视频报告中的 2 个帧,在每帧左上角 1/4 处显示用户当前运动状态和错误识别结果。本文还添加了辅助线,以帮助用户纠正错误姿势,例如标准的重心 COM 线和杠铃的真

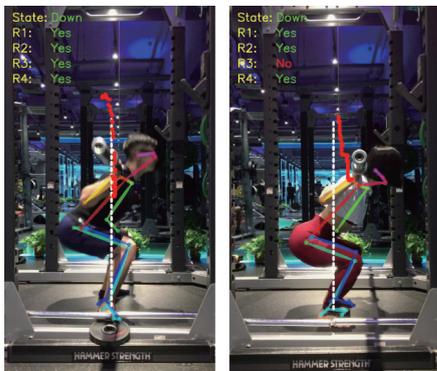


图8 可视化视频报告反馈

实轨迹。用户可以浏览包含这些诊断信息的完整视频以及动作明显出错的关键视频帧。

4 实验

本节通过一项用户研究来评估 2 种反馈机制在深蹲运动中的作用。

4.1 实验配置

本文使用商用智能手机和远程功能强大的服务器进行实验(实验视频发布在网站 <https://www.bilibili.com/video/BV1q54y1t7FQ>)。智能手机是配备高通骁龙 845 CPU@2.8 GHz、6 GM RAM 和超光敏 AI 后置摄像头的 Android 手机,该摄像头可以在 120 FPS 的速率下录制 1080p 视频。本文使用的无线耳机是 Sumsang Galaxy Buds,用于发出语音指令。智能手机通过无源 5 GHz 频段无线网络连接到实验室中功能强大的远程服务器(英特尔 Xeon Silver 4114 CPU@2.20 GHz,31 G RAM 和 Titan X Pro GPU)。

选择 15 位年龄分布在 18 ~ 35 岁之间的志愿者,在填写同意后,使用 GymBeats 系统录制他们的运动视频。由于 GymBeats 旨在帮助初学者用户提高运动表现,因此本文选择的所有志愿者都没有足够的杠铃深蹲经验。其中 10 名志愿者几乎没有去过健身房,而其中 5 名志愿者每周去健身房花费 3 ~ 5 h 进行有氧运动。所有的 15 名志愿者从未练习过杠铃深蹲,但让一位经验丰富的教练在锻炼前进行了标准动作示范。志愿者将在以下 3 种情况进行杠铃深蹲训练:(1) 无反馈(no feedback, NF); (2) 实时语音反馈(real-time voice feedback, RVF); (3) 运动后可视化反馈(post-exercise visual feedback, PVF)。在确保所有其他条件相同的前提下,每个志愿者在每种情况下都必须重复杠铃深蹲 5 组。在动作结束后,给每位志愿者一份问卷,询问他们关于这 3 种类型的反馈的有用性和偏好。此外,还将要求他们提供有关当前 GymBeats 系统设计和交互模式的其他建议。

4.2 状态和错误识别

本实验重点是测试 3.2 节中讨论的状态和错误识别阈值的有效性。如图 9 所示,本文进行了一组

对照实验,观察在运动过程中标准动作和错误动作的指标数据分布。首先收集了示范教练的指标数据,作为基本事实对照组。然后跟踪了 15 名志愿者的指标数据,并选择了不符合这 4 个标准要求的动作数据。

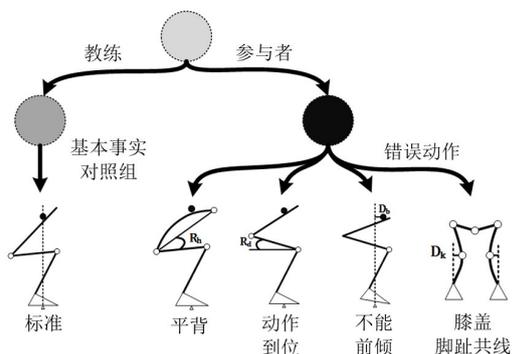


图 9 错误识别流程

图 10 展示了状态识别的指标结果。从图 10 中可以观察到 5 个虚线灰色方形中的数据具有相同的分布,这表明志愿者进行了 5 次重复的深蹲运动。本文在第 1 个数据波形中标记了深蹲运动的 4 个状态。志愿者蹲下时膝盖角速度为负,而志愿者站起时,膝盖的角速度为正值。当膝盖角速度接近于 0 rad/s 时,如果膝盖角接近于 180° ,说明志愿者处于站立状态。如果膝盖角远小于 180° ,则说明志愿者处于底部深蹲状态。

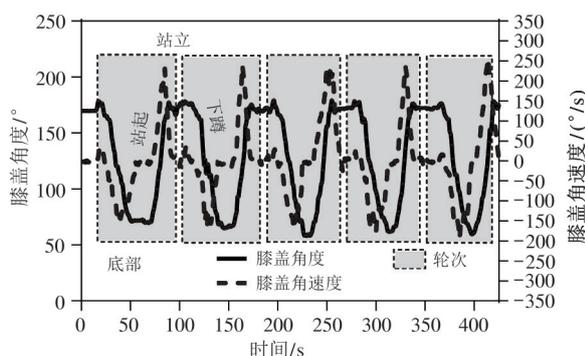


图 10 状态识别度量值测量

图 11 分别展示了错误动作的指标值分布,这些动作分别无法满足平背、动作到位、不能前倾以及膝盖脚跟贴线的要求。三角组成的波形表示志愿者的标准指标分布,而圆形组成的波形表示志愿者的错

误分布。本文是用虚线表示每个指标中的阈值,从而识别动作是否满足要求。可以清晰地看到,志愿者背部向后弯曲时,髌部角低于 60° (图 11(a))。当志愿者深蹲不到位时,深度角则超过了 10° (图 11(b))。当志愿者前倾时,杠铃距离指数超过 0.05 或者低于 -0.05 (图 11(c))。当志愿者膝盖和脚趾不处于一条直线时,膝盖距离指数超过 0.05 (图 11(d))。

发现 1 上下文感知方法对于杠铃深蹲细粒度的状态识别和错误识别来说十分必要。这种运动纠正方法也可以很好地应用到其他运动类型中。

4.3 运动质量提高

为了验证 GymBeats 在深蹲运动中改善身体姿势的效果,本文统计了 15 名志愿者在无反馈、实时语音反馈和运动后可视化报告反馈条件下的动作指标。具体来说,每位志愿者总共进行 9 次深蹲运动。其中前 3 次运动在无反馈下进行,中间 3 次运动在语音反馈下进行,最后 3 次运动在可视化报告反馈下进行。本文仅关注特定状态下的指标数据,例如平背错误识别中深蹲底部状态的髌部角度量,因为这些数据对于准确分析身体姿势的改善至关重要。

图 12 展示了本文深蹲实验中的指标分布,虚线表示每个度量值的阈值。对于平背错误识别(图 12(a)),本文观察到实验数据分布分别是无反馈($M = [34.2, 33, 32.8]$)、实时语音反馈($M = [56, 56.4, 53.6]$)和运动后可视化反馈($M = [36.2, 43.3, 52.7]$)。对于动作到位错误识别(图 12(b)),本文观察到实验数据分布分别是无反馈($M = [6.9, 7.5, 13.6]$)、实时语音反馈($M = [3.4, 1.2, 1.2]$)和运动后可视化反馈($M = [10.4, 5.1, 1.4]$)。对于不能前倾错误识别(图 12(c)),本文观察到实验数据分布分别是无反馈($M = [0.06, 0.04, 0.09]$)、实时语音反馈($M = [0.02, 0.03, 0.01]$)和运动后可视化反馈($M = [0.11, 0.03, 0.01]$)。最后,对于膝盖脚跟贴线错误识别(图 12(d)),本文观察到实验数据分布分别是无反馈($M = [0.05, 0.06, 0.06]$)、实时语音反馈($M = [0.03, 0.04, 0.04]$)和运动后可视化反馈($M = [0.06, 0.05, 0.04]$)。

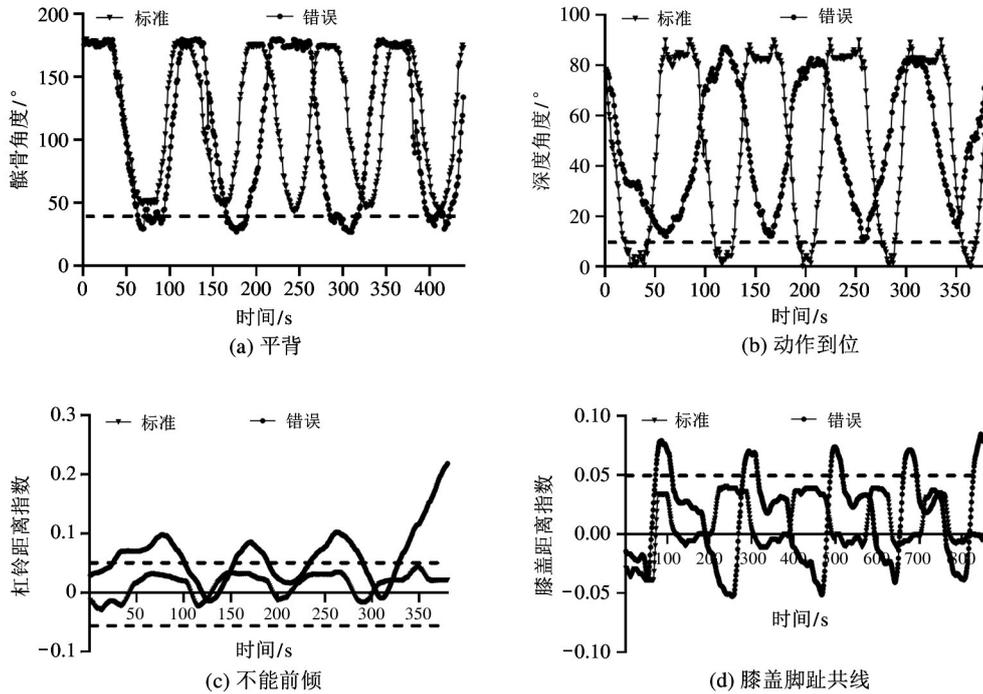


图 11 错误识别度量值测量

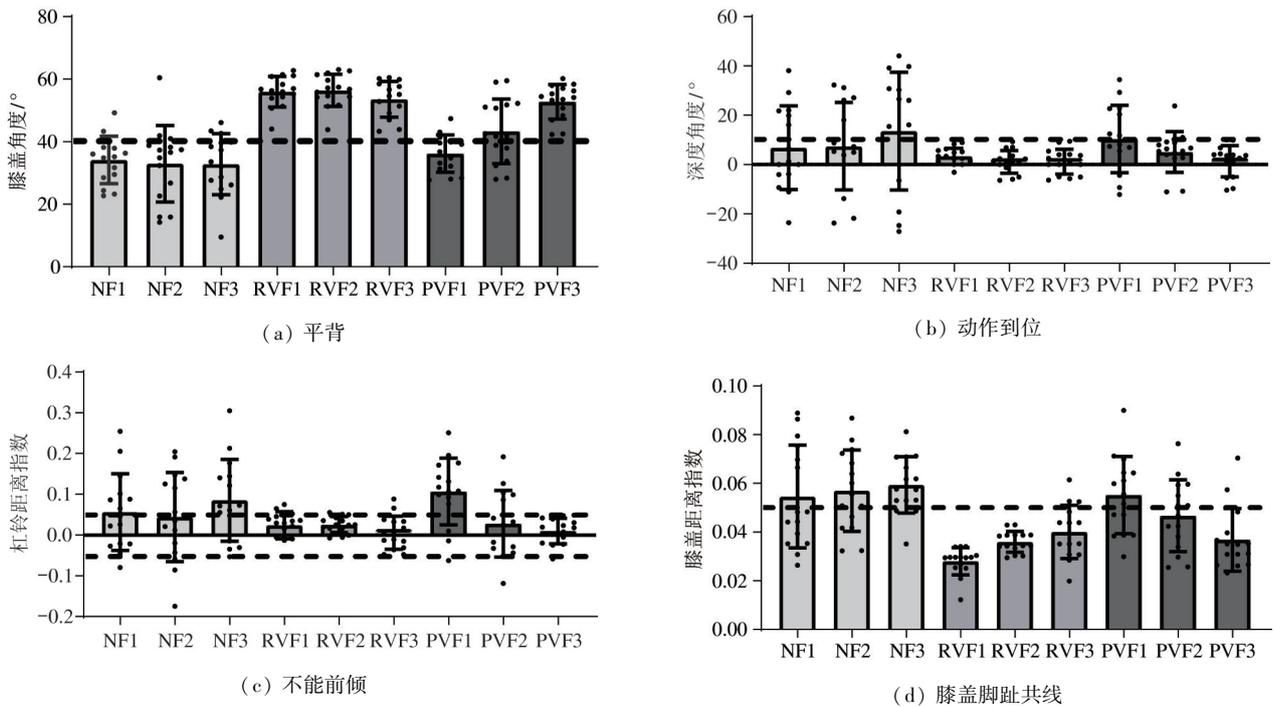


图 12 无经验志愿者在无反馈、实时语音反馈和运动后可视化反馈 3 种情况下的上下文感知度量值分布

通过对比每个度量值与虚线所表示的阈值,可以发现实时语音反馈 RVF 是纠正无经验志愿者错误身体姿势的最佳方法。这是因为志愿者在发现错误后可以立即纠正。另一方面,运动后可视化反馈 PVF 在每次运动结束后起作用。志愿者在看过报

告后可以找到姿势错误的地方,并在下一组运动中更正错误。但即使志愿者从 PVF 意识到了自己所有的姿势错误,也不一定能在下一组中全部纠正。

发现 2 志愿者可以从 GymBeats 中的实时语音反馈和运动后可视化反馈中受益。这两个反馈之间

的区别在于志愿者是否实时获取到运动纠正信息。

4.4 反馈偏好

本文分别对 5 位教练和 15 位志愿者进行主观评价偏好的问卷调查。问卷的形式为针对每个反馈的 5 分李克特量表(1 分表示非常低,5 分表示非常高),该调查表的结果如图 13 所示。

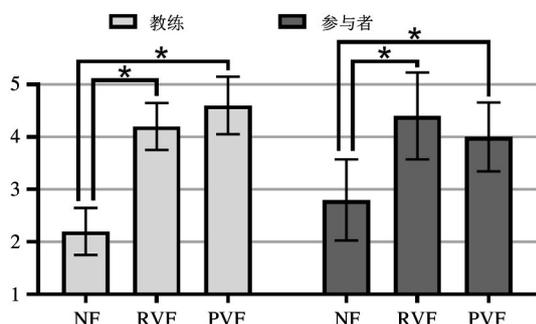


图 13 教练与无经验志愿者对不同反馈的主观评价

对于教练,就有效性而言,3 种反馈中均值最高的是运动后可视化反馈($M = 4.6, SD = 0.5$),其次分别是实时语音反馈($M = 4.2, SD = 0.4$)和无反馈($M = 2.2, SD = 0.4$)。单项方差分析测试显示主要影响数据为 $F_{2,12} = 35.43, p < 0.00001$ 。本文通过 Tukey's HSD 测试分别识别 RVF-NF 和 PVF-NF 的差异性。数据显示有经验的教练更喜欢 PVF 而不是 RVF,这是因为他们熟悉标准的深蹲动作,因此 RVF 几乎没有帮助。而更直观的运动后可视化报告反馈 PVF 对于他们而言更为有效。

对于无经验志愿者而言,结果却并不一样。3 种反馈中均值最高的是实时语音反馈($M = 4.4, SD = 0.8$),其次分别是运动后可视化反馈($M = 4, SD = 0.6$)和无反馈($M = 2.8, SD = 0.8$)。同时,单项方差分析测试显示主要影响数据为 $F_{2,42} = 18.2, p < 0.00001$ 。与教练数据相似, Tukey's HSD 测试表明,与 NF 相比, RVF 和 PVF 都被认为有效。

显然,教练和无经验志愿者对于不同反馈的主观性偏好各有不同。首先,由于无经验志愿者缺乏深蹲练习的经验,他们希望 GymBeats 可以实时纠正他们的错误,因此更喜欢 RVF 而不是 PVF。其次,志愿者的 3 种反馈的标准差和无反馈的均值都大于教练。这是因为志愿者缺乏专业知识,并没有意识到深蹲运动中非标准动作的危害性。最后,志愿者

之间的个体差异比教练之间的个体差异数值更显著,这导致 3 种反馈的接受范围变化幅度更大。

发现 3 对于无经验志愿者,实时反馈对于改善运动表现更为有效,而运动后反馈对于有经验的教练而言更为有效。与无反馈情况相比,另两种类型反馈都认为是有效的。

5 结论

本文提出了一种新颖的基于端云协同的运动纠正框架 GymBeats,该框架可以使用智能手机和笔记本电脑等商用设备进行运动纠正。具体来说,本文描述了现有解决方案存在的不足,并提出了一种细粒度的上下文感知运动纠正方法,包括状态识别和错误识别。在此基础上,设计了基于端云协同的概念验证原型,可以生成实时语音和运动后的可视化反馈。GymBeats 共有 3 个处理阶段,包括骨骼提取、上下文感知识别和评估反馈阶段。本文与 15 位志愿者一起评估了 GymBeats。结果表明,实时语音和运动后的可视化反馈都显著改善了用户在深蹲运动中的身体姿势。下一步, GymBeats 将在智能手机上建立一个有前景的健身生态系统,并为人们的整体运动和健康做出贡献。

参考文献

- [1] ASUS. Asus Xtion Pro [EB/OL]. https://www.asus.com/3D-Sensor/Xtion_PRO/; ASUS Corporation, [2020-12-31]
- [2] Microsoft. Microsoft Kinect [EB/OL]. <https://developer.microsoft.com/en-us/windows/kinect/>; Microsoft Corporation, [2020-12-31]
- [3] Vicon. Vicon [EB/OL]. <https://www.vicon.com/>; Vicon Corporation, [2020-12-31]
- [4] OptiTrack. OptiTrack [EB/OL]. <https://www.optitrack.com/>; OptiTrack Corporation, [2020-12-31]
- [5] SUNNY C, DAVID W M, TAMMY T, et al. Activity sensing in the wild: a field trial of ubifit garden[C] // Computer Human Interaction, Florence, Italy, 2008: 1797-1806
- [6] HAN D, LONGFEI S, ZHENG Y, et al. FEMO: a platform for free-weight exercise monitoring with RFIDs[C]

- //ACM International Conference on Embedded Networked Sensor Systems, Seoul, Korea, 2015: 141-154
- [7] TIANSHI C, ZIDONG D, NINGHUI S, et al. DianNao: a small-footprint high-throughput accelerator for ubiquitous machine-learning [C] // International Conference on Architectural Support for Programming Languages and Operating Systems, Salt Lake City, USA, 2014: 269-284
- [8] DON S E, DENYS J C M, LÖIC D, et al. GymSoles: improving squats and dead-lifts by visualizing the user's center of pressure [C] // Computer Human Interaction, Glasgow, UK, 2019: 1-12
- [9] OptiTrack. OptiTrack Accessories [EB/OL]. <https://www.optitrack.com/accessories/>; OptiTrack Corporation, [2020-12-31]
- [10] JAMIE S, ANDREW W F, MAT C, et al. Real-time human pose recognition in parts from single depth images [C] // Computer Vision and Pattern Recognition, Colorado Springs, USA, 2011: 1297-1304
- [11] CALEB C, GENE M P. Correcting exercise form using body tracking [C] // CHI Extended Abstracts, San Jose, USA, 2016: 3028-3034
- [12] SAMANTA R, GABRIELLA B, MARCO K. Comparison of different sets of features for human activity recognition by wearable sensors [J]. *Sensors*, 2018, 18(12): 1-16
- [13] 张春祥, 赵春蕾, 陈超, 等. 基于手机传感器的人体活动识别综述 [J]. *计算机科学*, 2020, 47(10): 1-8
- [14] ARIF R A, HONGNIAN Y, MICHAEL V. An automatic gait feature extraction method for identifying gait asymmetry using wearable sensors [J]. *Sensors*, 2018, 18(2): 1-28
- [15] 张文超, 魏东岩, 袁洪, 等. 基于惯性递推原理的行人自主定位方法综述及展望 [J]. *导航定位与授时*, 2021, 8(3): 109-122
- [16] SUMIT M, TAPAS M, M. J D. Wearable device-based gait recognition using angle embedded gait dynamic images and a convolutional neural network [J]. *Sensors*, 2017, 17(1): 1-20
- [17] 李延军, 李莹辉, 余新明. 穿戴式健康监测设备的现状与未来 [J]. *航天医学与医学工程*, 2016, 29(3): 229-234
- [18] KENG-HAO C, MIKE Y C, JOHN F C. Tracking free-weight exercises [C] // Ubiquitous Computing (UbiComp), Innsbruck, Austria, 2007: 19-37
- [19] SUNNY C, DAVID W M, TAMMY T, et al. Activity sensing in the wild: a field trial of ubifit garden [C] // Computer Human Interaction, Florence, Italy, 2008: 1797-1806
- [20] CHENGUANG S, BO-JHANG H, MANI B S. MiLift: efficient smartwatch-based workout tracking using automatic segmentation [J]. *IEEE Transactions on Mobile Computing*, 2018, 17(7): 1609-1622
- [21] HAWKAR O, ANURAG S, MIN-HYUNG C, et al. VRInsole: an unobtrusive and immersive mobility training system for stroke rehabilitation [C] // International Conference on Wearable and Implantable Body Sensor Networks, Las Vegas, USA, 2018: 5-8
- [22] STACY J M B, ARI Y B, DONNA M S, et al. Gait analysis using a shoe-integrated wireless sensor system [J]. *IEEE Transactions on Information Technology in Biomedicine*, 2008, 12(4): 413-423
- [23] CHUNYU W, YIZHOU W, ALAN L Y. An approach to pose-based action recognition [C] // Computer Vision and Pattern Recognition, Portland, USA, 2013: 915-922
- [24] 杨君, 张素君, 张创豪, 等. 基于 OpenPose 的人体动作识别对比研究 [J]. *传感器与微系统*, 2021, 40(1): 5-8
- [25] CAO Z, HIDALGO G, SIMON T, et al. OpenPose: real-time multi-person 2D pose estimation using part affinity fields [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, 43(1): 172-186
- [26] FANG H S, XIE S, TAI Y W, et al. Rmpe: regional multi-person pose estimation [C] // Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017: 2334-2343
- [27] KAIMING H, GEORGIA G, PIOTR D, et al. Mask R-CNN [C] // IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017: 2980-2988
- [28] ABADI M, BARHAM P, CHEN J, et al. TensorFlow: a system for large-scale machine learning [C] // The 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI), Savannah, USA, 2016: 265-283
- [29] GLASSMAN G. The crossfit training guide [EB/OL]. <http://library.crossfit.com/free/pdf/CFJ/Seminars/TrainingGuide/012013-SDy.pdf>; Crossfit Corporation, [2020-12-31]

- [30] RIPPE TOE M, KILGORE L. Starting Strength Basic Barbell Training [M]. Wichita Falls: Aasgaard Company Publishers, 2001

Design and implementation of cloud-edge based exercise correction system with commodity devices

XU Yuan^{***}, LIU Zhigang^{***}, WANG Sa^{*}, BAO Yungang^{*}

(^{*} Institute of Computing Technology, Chinese Academy of Science, Beijing 100190)

(^{**} University of Chinese Academy of Sciences, Beijing 100049)

Abstract

This paper introduces an exercise correction prototype, GymBeats, and shows how a user can correct errors in body posture during exercise with a commodity devices. Specifically, a novel context-aware exercise correction methodology is proposed, including state recognition and mistake recognition. A proof-of-concept cloud-edge based system is designed and a typical full-range-of-motion-exercise is used to describe the proposed methodology. GymBeats is evaluated with 15 users in two feedbacks, real-time voice feedback and visual report feedback. The experimental results show that GymBeats can significantly improve body posture in a low-cost and convenient way.

Key words: exercise correction, commodity device sensor, skeleton extraction, context-aware recognition, real-time voice feedback, visual report feedback