

基于人体姿态识别的立定跳远动作智能评估系统

□文 / 李晟¹, 宋可儿², 欧阳柏强¹, 张博中¹, 刘宠¹, 刘朋¹, 刘睿¹

(1. 中移(成都)信息通信科技有限公司, 成都 610096; 2. University of Georgia, Athens GA 30602, USA)

摘要: 针对立定跳远教学场景中的动作不规范及成绩量化不统一等问题, 本文设计并实现了一种基于人体姿态识别技术的立定跳远动作智能评估系统。该系统首先对标准姿态视频进行关键点估计, 并对其进行数据预处理, 在此基础上将立定跳远特征与专家意见相结合, 构建立定跳远标准动作数据信息库; 然后采用摄像头捕获待评估人员姿态, 将其与库中的立定跳远标准姿态进行关键帧匹配; 最后采用皮尔逊积矩相关系数综合各关键点对立定跳远运动的影响程度, 并将其与计算向量相似度及关键特征姿态角度等方法结合, 定义了立定跳远动作规范性评价指标, 以进行姿态的差异化比对。经验证, 该系统可对立定跳远运动姿态的规范性提供有效的评估, 同时为用户提供改进建议, 促使其纠正运动姿态, 提高运动成绩。

关键词: 人体姿态识别; 立定跳远; 关键帧匹配; 姿态比对

中图分类号: TP391; G80 **文献标志码:** A **文章编号:** 2096-5036(2022)02-0075-13

DOI: 10.16453/j.cnki.ISSN2096-5036.2022.02.009

0 引言

随着《关于全面加强和改进新时代学校体育工作的意见》^[1]的提出, 以结合人工智能及深度学习方法为基础的新型体育教学方法不断涌现, 其智能化、数据化、个性化的优点日益显现^[2]。在立定跳远这一传统体育运动项目中, 对其标准姿态的定义大多来源于图像视频及教师的个人经验, 缺乏量化的评估标准, 难以准确有效地评估其动作的规范性并给出合理的改进意见。早期的动作评估智能系统主要通过大量传感器及硬件设备提供支持, 穿戴复杂且会影响学生在运动过程中的发挥, 因此并未广泛普及。计算机视觉的快速发展使得人体姿态识别技术得到了良好的应用, 基于视觉的人体姿态评估以其非接触、低成本等优势, 为立定跳远教学及指导提供了一种全新的交互方式^[3]。

目前, 基于人体姿态的动作识别及对比研究受到了国内外广泛关注。罗大为等^[4]从多个维度对人体姿态识别方法进行研究, 对其疑难点及发展趋势进行了阐述。在人体姿态对比方面, 王建波和邱凯等^[5]基于深度视觉跟踪, 通过异常检测和性能评级, 建立了一种人工智能教练系统 AI Coach。Kim T T 等^[6]提出了一种利用基于视觉的姿态估计

以寻找关键点的分析方法，通过其关键时刻对球手挥杆动作的评估及量化，输出评估结果。胡建朗等^[7]针对运动姿态不标准且无人监督等问题，基于双目变焦伺服系统，结合注意力单元追踪机制，研究实现了一种三维姿态评估方法。王芫^[8]基于相似度和局部评估的概念设计并训练了一种羽毛球运动员姿态评估模型。李睿敏^[3]分别从动作分类、动作检测和动作评估方向出发，设计并实现了基于手工特征的动作分类方法和高精度动作检测方法。杨君等^[9]使用 openpose 技术开发了一套单目图像识别对比指导系统。陈雪梅^[10]以 openpose 提取的二维姿态为基础，利用视差原理恢复三维姿态的方法，建立了一套高尔夫挥杆辅助训练系统。

基于以上所述，针对学校教学智能化、数据化的迫切需求，本文旨在建立一套立定跳远动作智能评估系统。该系统通过提取标准技术动作特征，建立立定跳远标准动作特征库。之后利用摄像头捕获立定跳远视频数据对其中的人体关键点进行估计，通过关键动作的匹配评估人体骨骼关键点在完成各阶段技术动作过程中的参与度，以此作为评估标准，以帮助学生对立定跳远动作规范性的判别及指导，提高其竞技水平，同时避免运动中可能存在的风险。

1 系统设计

本文所述的立定跳远动作智能评估系统的实现框架如图 1 所示，主要由 3 个子系统构成——标准动作库子系统、动作采集子系统，以及人体姿态动作评估子系统。其中，标准动作库子系统通过收集立定跳远标准姿态相关数据，采用姿态识别算法提取关键动作特征，构建标准技术动作数据信息库；动作采集子系统主要包括人脸识别模块和动作数据采集模块；人体姿态动作评估子系统则用于对视频中的待测人员进行姿态估计，完成关键帧抽取并与标准姿态关键帧匹配，利用标准动作量化模型进行差异化对比并获得评估结果。

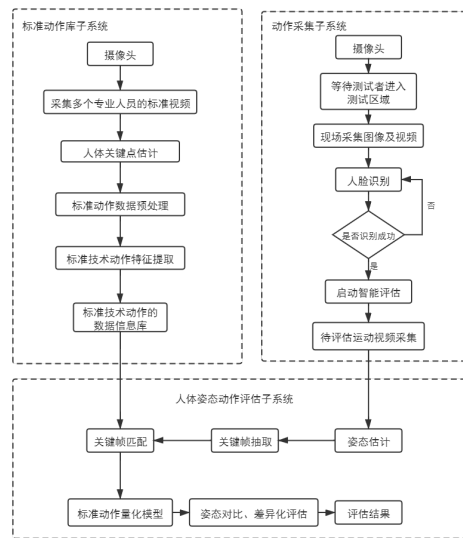


图 1 立定跳远姿态评估系统框架图

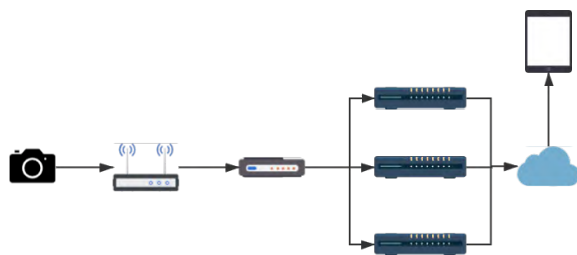


图 2 系统硬件组成图

本文的硬件部分如图 2 所示，本系统采用单目变焦枪式摄像头进行视频图像的采集，设置其分辨率为 1920×1080 ，以提高立定跳远测距的精度，并将其帧率设置为 25FPS，在保证稳定传输的前提下，尽可能提高输出

的频率，避免出现运动模糊的现象。随后，将采集到的视频图像通过路由器及网关转发到远端的边缘云服务器，进行视频图像的分析处理工作，最后将分析处理结果发送到远端云服务器保存，并可实时呈现在远程移动终端设备上。

图3为立定跳远智能评估系统的工作流程，首先，由摄像头采集待测试人员的立定跳远动作视频，并采用自顶向下的方法进行人体检测，之后输出待测人员的人体骨骼关键点。其次，根据立定跳远运动的特点并结合专家意见定义其运动过程中的关键姿态，以此为基础将待测姿态与事先构建的立定跳远标准动作库里的标准姿态进行关键帧匹配，即将两组姿态序列按时序对齐。最后，通过采用皮尔逊积矩相关系数综合度量两组姿态之间的相关性，与特征向量相似度及关键特征姿态角度两种方法相结合，以此为评价指标并输出立定跳远动作的改进建议。

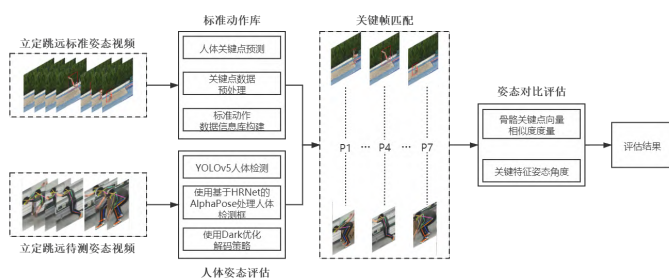


图3 立定跳远姿态评估系统工作流程图

2 立定跳远动作智能评估系统

2.1 标准动作库子系统

标准动作库子系统使用姿态评估算法和关键帧动作特征提取等手段，在立定跳远领域创新性地构建了标准技术动作数据信息库。其核心部分主要分为3个单元——人体关键点预测单元、关键点数据预处理单元，以及标准动作数据信息库构建单元。

2.1.1 人体关键点预测单元

人体关键点预测单元主要作用是获取标准视频中的人体关键点序列信息。其通过使用高速摄像头采集多个立定跳远标准动作视频（其中视频帧率为25FPS，视频长度为14s左右），进而对采集的视频进行逐帧分析，运用人体姿态估计算法 PoseNet 对人体对象在执行技术动作过程中的动作姿态进行关键点位置估计，以视频帧时间戳为序，输出所有视频帧的人体关键点序列。如图4所示，从左至右，分别展示了立定跳远过程各阶段的人体骨骼关键点。



图4 立定跳远过程图（含骨骼点信息）

2.1.2 关键点数据预处理单元

关键点数据预处理单元主要是对获得的人体关键点序列使用 Savitzky-Golay 滤波器^[11]进行平滑去噪。关键点数据预处理单元的工作流程分如下两步。

第一步，绘制人体骨骼关键点坐标值变化图。为了便于观察立定跳远过程中人体骨

骼关键点的突出特征和变化情况，本单元以2.1.1部分某一人体关键点序列数据为例，绘制出立定跳远过程中人体骨骼关键点坐标X值和Y值的变化情况，分别如图5和图6所示。

在立定跳远领域，本文主要聚焦于人体的手腕、手肘、肩膀、脚踝、脚尖、脚后跟、膝盖和髌等骨骼关键点的坐标变化情况（以人体左边骨骼点为例）。由图5及图6可以看出，以上关键点的整体移动可以指向立定跳远运动的某种状态。如在人体关键点X值迅速减小之前，手腕关键点的Y值上下明显波动，表明目标人员正在进行起跳前的蓄力准备工作。

第二步，本文使用了Savitzky-Golay滤波器进行数据流平滑除噪。Savitzky-Golay滤波器是一种在时域内基于局域多项式最小二乘法拟合的滤波方法，其最大的特点在于在滤除噪声的同时可以确保信号的形状、宽度不变。如图5所示，人体关键点坐标的X、Y值在250帧至350帧这个区间，有一定程度上的不确定性，这是PoseNet模型是否可以准确找到身体部位关键点的能力导致的。为了对抗这些突然的变化，本文使用了Savitzky-Golay滤波器进行平滑去噪，降低噪声的干扰，过滤掉立定跳远运动中人体关键点的一些极端异常值。

2.1.3 标准动作数据信息库构建单元

标准动作数据信息库构建单元将“自定义-关键帧”和“专家-关键帧”结合，得到关键帧对应的人体姿态特征向量和各姿态角度等信息，构建标准动作的数据信息库。

2.1.3.1 自定义-关键帧

本部分使用Python定义了一个“Get Custom keyframes”函数，重点关注的上述骨骼关键点Y值的相交点和极值点，获得自定义关键帧。

在相交点方面，得到视频中某一关键点与一个或多个其他关键点相交的视频帧。例如，当左手腕Y值与左肩膀Y值曲线相交时，得到所有交叉的视频帧。然后，进一步过滤视频帧，计算视频帧中所有部位相交的次数，并在执行“群交叉”（多个部位的相交）计算时选择发生次数最多的帧，从而获取关键人体部位运动的重要特征，然后将他们过

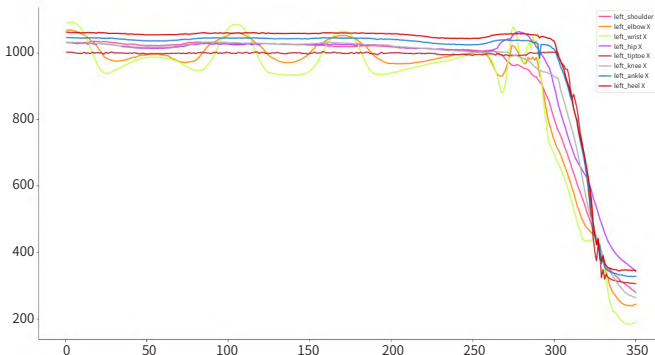


图5 立定跳远关键点坐标X值变化图

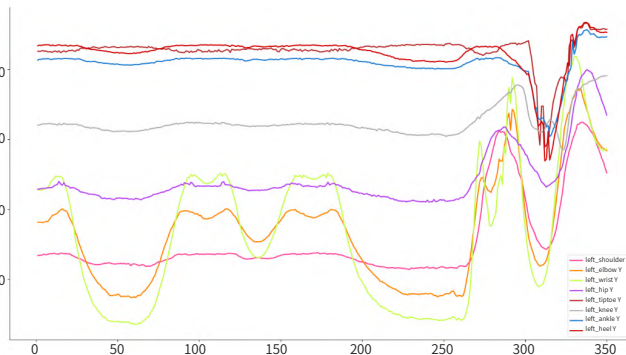


图6 立定跳远关键点坐标Y值变化图

滤到“自定义 - 关键帧”中。

在极值点方面，通过寻找上述重点部位的极大值和极小值确定一些关键运动特征。如图 7 所示，在标点 1 之前的左手腕的极大值代表达到手臂预摆的最高点。一个标准立定跳远动作的 X/Y 曲线图会使临界点的寻找变得简单，但这对 PoseNet 有着极高要求。若姿态估计在精度上存在误差，则视觉上恒定的姿势其表现结果可能会存在较大波动。因此，在此步骤之前需采用上文的数据预处理单元进行降噪过滤处理。

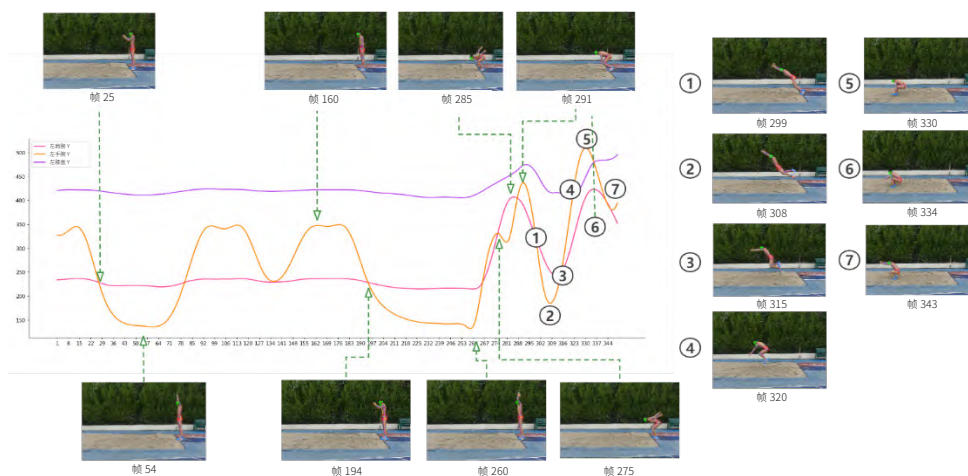


图 7 取自定义 - 关键帧简单示意图 (仅使用手腕、膝盖、肩膀部位)

2.1.3.2 专家 - 关键帧

本部分是基于国家体育锻炼标准和专家建议，将跳远过程拆分成 P1 至 P7 共 7 个“专家 - 关键帧”。

为了建立立定跳远的标准技术动作特征库，需要选取特定的立定跳远者的关键特征。这些关键特征可以影响视频分析的准确性以及影响为跳远者提供性能指标和姿势纠正建议的可靠度。因此，关键帧的抽取需要十分科学合理，根据国家体育锻炼标准和专家建议，对立定跳远整个动作进行分解，得到了 7 种动作特征，通过这 7 种动作特征进行关键点匹配，如图 8 所示。

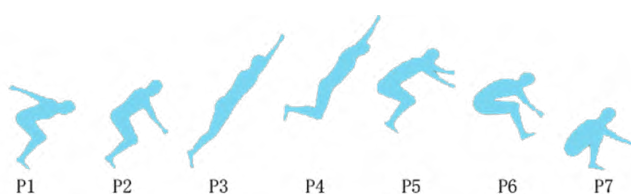


图 8 立定跳远标准动作特征 (P1: 双手向后与身体夹角呈 45 度左右，双腿弯曲 100 度左右蓄力; P2: 双手向前摆出，与身体呈 90 度左右，双腿弯曲 110 度左右用力蹬出; P3: 双手、双腿与身体保持一条直线 (身体呈 180 度左右)，向前荡出; P4: 双腿弯曲 90 度左右，双手仍与身体保持直线 (180 度左右)，达到最高点; P5: 双腿向前弯曲 90 度，双手向前保持平衡 (与身体存在的夹角不做特定要求); P6: 双腿尽量向前伸直 (大于 90 度)，尽可能跳出更远距离，双手保持平衡; P7: 双腿弯曲 (小于 90 度)，与地面垂直 (膝盖与脚踝关键点 y 轴相近)，完成立定跳远动作，双手保持平衡)

综上所述，本单元将“自定义 - 关键帧”和“专家 - 关键帧”两部分结合，剔除重复的关键帧后，获得共 10 个关键帧。然后，针对每个关键帧，提取各个骨骼关键点特征向

量和姿态中的各个角度大小，构成本文的标准技术动作的数据信息库。

我们利用视频采集器，采集标准动作，运用 PoseNet 对人体对象进行关键点位置估计，再结合数据去噪、关键帧提取等手段，分析不同性别的专业学生的标准技术动作，形成结构化标准技术动作的数据信息库。

2.2 动作采集子系统

动作采集子系统主要是负责采集测试者运动数据、对测试者进行人脸核对，以及开启算法模块。其中，立定跳远项目按图 9 方式进行部署，将摄像头置于测试区域左侧，枪头下倾 30 度左右，距测试区域约 4 米，测试区域应为摄像头采集的中间区域，最好为无干扰的封闭区域，以便于摄像头对焦与采集。

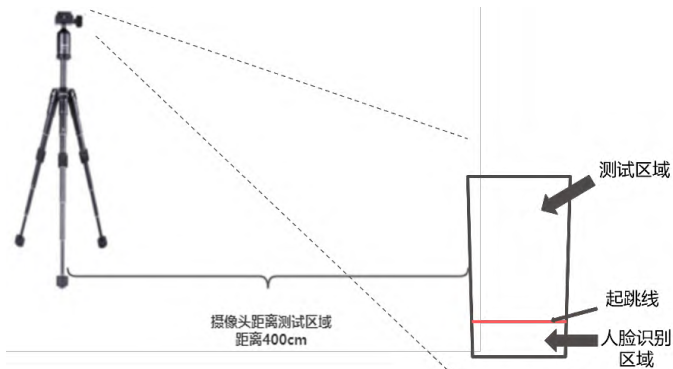


图 9 立定跳远项目部署示意图

在现场立定跳远智能评估系统部署好之后，待评估人员根据指示进入人脸识别区域，系统调用人脸识别算法进行身份验证，随后启动智能评估，摄像头持续采集测试区域的视频数据，并将其传输给边缘云服务器，为后续立定跳远标准姿态和待评估姿态的差异化对比做准备。

2.3 人体姿态动作评估子系统

首先，人体姿态动作评估子系统使用人体姿态评估模型进行姿态估计；其次，根据上文所定义的立定跳远标准技术动作特征，抽取标准姿态视频与待评估姿态视频中的相似帧进行人体姿态匹配；最后，根据人体姿态的特征选取合适的骨骼点组成特征向量及关键特征姿态，通过对比向量的相似性及特征姿态的角度，结合其与标准姿态的相关系数评估待评估姿态的准确率，并分析它们之间的差异。

2.3.1 人体姿态评估模型

人体姿态评估模型不断捕获测试区域待评估人员的运动视频，将每一帧视频实时输入到我们模型中。该模型由三部分组成，第一部分为人体检测，使用目标检测模型 YOLOv5^[12]；第二部分使用骨干网络为 HRNet^[13] 的 AlphaPose^[14] 模型处理人体检测框，对人体检测框进行优化；第三部分使用 Dark 对解码策略进行了优化。

人体关键点检测对于描述人体姿态，预测人体行为至关重要。传统的人体检测会遇到两大难题——检测框定位错误与检测框冗余。在对这两个难题的处理上，本文使用了如下策略。

- a) 数据增强 COCO 公开数据集^[15] 包含有 4000 张已标注数据，输出的人体姿态含

有 17 个骨骼关键点。在本文中，为了提高立定跳远项目智能评估的准确性以及提高动作的可辨识度，在原有 17 个关键点的基础上增加了脚跟、脚尖与下巴等关键点。此外，本次研究重新标注了 1 万张立定跳远项目中的人体骨骼关键点，通过提高标注图片数量及质量的方式，有效降低了检测框定位错误和检测框冗余两大问题。

b) 加入 AlphaPose AlphaPose 是一个多人姿态估计系统，具有极高的精准度，代码支持 Linux 和 Windows，系统采用 PyTorch 框架，在姿态估计的标准测试集 MSCOCO 上，达到 71mAP 的精度。同时，其速度达到 20FPS，相比其他主流模型，AlphaPose 在处理速度上有较大优势。

表 1 现有姿态估计系统在 COCO 数据集上的结果比较

人体姿态系统	检测准确率
Our PoseNet	80.5mAP
AlphaPose (ICCV 2017)	72.5mAP
Mask-RCNN	67mAP
OpenPose	61.8mAP

本文模型采用 RMPE 框架，该框架采用自顶向下的方法，即先检测人，再进行姿态估计。其包括三个主要组成部分，第一个组成部件是对称空间变换网络 (Symmetric STN)，第二个组成部件是由姿态引导的样本生成器 (Pose-guided Proposals Generator)，第三个组成部件是参数化的姿态非极大值抑制器 (Parametric Pose NMS)，模型框架图如图 10 所示。

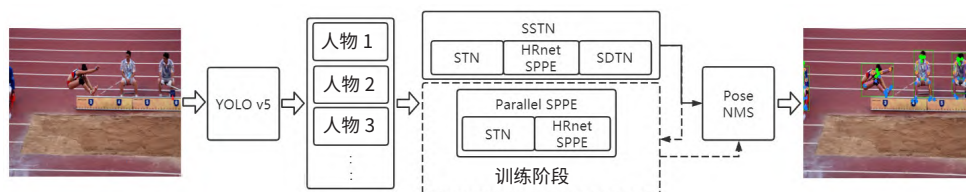


图 10 模型框架图

其中，STN 为空间变换网络，对于不准确的输入，得到准确的人体标识框，将其输入用于获取高质量的候选区域。SPPE 为进行姿态估计后的人体姿态。SDTN 为空间逆变换网络，将估计的姿态映射回原始的图像坐标。Pose-NMS 用于消除额外估计到的姿态。因此，AlphaPose 在很大程度上会解决检测框定位错误和检测框冗余这两大问题。

在模型训练中，通常会为了减少训练难度和资源占用率而降低输入图片的分辨率，在降采样后的分辨率上进行训练。为了让模型可以用热度图的方式进行训练，降采样的同时也需要将原图中的骨骼关键点坐标转化为降采样后分辨率下的关键点坐标，并用高斯模糊转换成热度图，这个过程为坐标编码。在进行热度图预测后，将降采样后的图片分辨率恢复到原始像素，并在原始坐标空间找到预测的关键点坐标，这一过程即为坐标解码。

在现有的人体姿态评估模型中，很少有研究关注编码与解码的过程，但其对最终结果的影响具有很大作用。在模型训练结束后，需要对图像进行上采样。此时，大多数模型会将最终预测得到的热度图中的最大点坐标作为关键点位置。但由于降采样和上采样的存在，会存在较大的量化误差，即热度图最大的激活位置并非为关键点位置的精确定位，其存在一定的偏差。现有的解码策略如式 (1) 所示。

$$p = m + 0.25 \frac{s - m}{\|s - m\|_2} \quad (1)$$

m 表示最大激活位置, s 表示第二大激活位置, s 的目的就是补偿降采样导致的量化误差,但这种策略缺乏理论依据。因此,本文采用了另一种策略——Dark。这种策略充分利用了热度图的分布统计信息,增加了最大激活位置的准确度,并且Dark策略只需要计算每个热度图的一阶和二阶导数,因此并没有增加太多计算复杂度,没有任何计算成本障碍。其中,Dark策略如下所示。

$$\varphi(x;u,L) = \frac{1}{(2\pi)^{\frac{1}{2}}L^2} \exp\left(-\frac{1}{2}(x-u)^T L^{-1}(x-u)\right) \quad (2)$$

x 为预测热度图中的一个像素, u 为对应于要估计的关节位置的高斯平均值,协方差 L 是一个对角矩阵。使用Dark策略,可以提高模型准确率,输出最大的激活位置。通过本模型输出的立定跳远人体姿态如图11所示。



图 11 立定跳远中的人体姿态

2.3.2 立定跳远关键帧匹配

本文中用 10 个关键帧衡量立定跳远动作的准确性。如何进行匹配?其选取的依据就是根据人体姿态各部位之间的角度、在连续帧中的位置,以及骨骼点相对于图片的 (X, Y) 坐标等,按照一定的比例系数计算其关键帧相似度,公式如下所示。

$$\Delta = \theta_1 \cdot X_1 + \theta_2 \cdot X_2 + \theta_3 \cdot X_3 + \dots \quad (3)$$

其中, θ_1 、 θ_2 、 θ_3 为各项特征系数, X_1 、 X_2 、 X_3 为人体姿态各部位之间的角度、在连续帧中的位置,以及骨骼点相对于图片的坐标 (X, Y) 等特征。

计算每一帧的关键帧相似度,按照关键帧相似度,抽取最佳的人体姿态关键帧,设立一个误差阈值,相似度在误差阈值内都可以被选取。再按照一定的预设规则匹配“自定义-关键帧”和“专家-关键帧”共 10 个动作过程。规则如下。

- P1-P7 过程按照先后顺序进行匹配
- 身体角度总数最接近优先匹配
- 腿部骨骼角度特征优先匹配
- 是否离地特征优先匹配
- 相似度大优先匹配

2.3.3 姿态评估及对比

立定跳远运动的智能评估不仅受其运动场景的影响,还与待评估人员的身高、体形,以及与摄像头的相对位置有关。因此,直接使用骨骼点数据不能建立普适的评估模型,需对获得的人体骨骼点坐标进行坐标转换并进行标准化,以此为基础开始立定跳远动作的评估工作^[16]。根据本文模型得到的主要人体骨骼关键点如图 12 所示,包括头

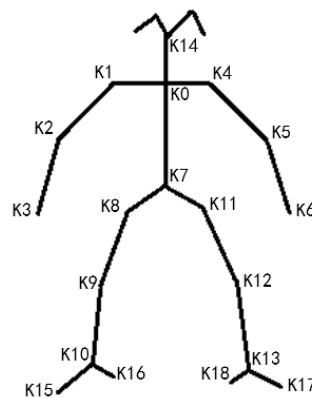


图 12 人体关节点示意图

部(包含鼻、眼、耳)、脊柱中心、上下肢、脚尖和脚后跟等。

根据立定跳远运动的特点,四肢在整个运动中具有较大的波动,而脊柱中心(K0与K7分别为肩部及髋部中点)则相对稳定。因此,本文以髋部中心点K7作为基准点,评估分析四肢在整个运动过程中的变化。首先,按照标准动作中人员的髋部中心点坐标,将待评估人员对应部位移动到相同位置。以动作开始前人员站立姿态为标准,计算标准姿态与待评估姿态各关节点坐标X值和Y值之间的比率。以左上肢为例,公式分别如下。

$$r_x = (K4_x_2 - K6_x_2) / (K4_x_1 - K6_x_1) \quad (4)$$

$$r_y = (K4_y_2 - K6_y_2) / (K4_y_1 - K6_y_1) \quad (5)$$

其中, $K4_x_2$, K_x_2 分别为待评估人员的左肩及左手腕的X值, $K4_x_1$, $K6_x_1$ 分别为标准姿态人员的左肩及左手腕的X值, Y值同理。据此,即可消除待评估人员在骨骼关节点坐标不统一及体形差距上的影响,更加方便地进行姿态评估。

在进行待测人员骨骼关节点数据标准化后,本文将各帧中重点关注的人体关键点坐标组成坐标点序列,分别通过采用皮尔逊积矩相关系数(Pearson Product-Moment Correlation Coefficient, PPMCC)度量待评估人员关键点和标准动作关键点之间的相关性,之后根据立定跳远运动的特点按照一定的比例系数计算关键帧之间的总体相似度,作为衡量立定跳远动作准确性的依据之一。

皮尔逊积矩相关系数是用于度量两个序列X和Y之间相关关系的方法,其值介于-1与1之间。若其为正数,则表示两个序列正相关;反之为负相关。其公式定义为两序列之间的协方差和标准差的商,如式(6)所示。

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \quad (6)$$

按照一定比例系数计算其关键帧总体相关系数,公式如下。

$$\delta = \theta_1 \cdot \rho_1 + \theta_2 \cdot \rho_2 + \theta_3 \cdot \rho_3 + \dots \quad (7)$$

其中, θ_1 、 θ_2 、 θ_3 为各个关键点对立定跳远评估的影响系数, ρ_1 、 ρ_2 、 ρ_3 为待评估姿态与标准姿态的皮尔逊积矩相关系数。

之后确立了以计算标准动作与待测动作对应关键帧中各骨骼关键点组成关键特征姿态之间角度的方式进行姿态评估及差异化对比的补充。关键特征姿态可由两个或者多个骨骼关键点连线组成,可根据研究项目的特征有针对性地选取与组合不同的骨骼关键点进行分析。以本文所研究的立定跳远为例,针对其特征,我们选取如下骨骼关键点组成关键特征姿态进行分析。

如表2所示,关键特征姿态就是由各个骨骼关键点组合成的向量构成,因此本文采用两种方法结

表2 人体骨骼点构成的关键特征

关键特征姿态	骨骼关键点
左上肢	$V_1 = (K_1, K_2, K_3)$
右上肢	$V_2 = (K_4, K_5, K_6)$
左下肢	$V_3 = (K_8, K_9, K_{10})$
右下肢	$V_4 = (K_{11}, K_{12}, K_{13})$
躯体主干	$V_5 = (K_0, K_7, K_8, K_9)$
左脚	$V_6 = (K_{12}, K_{13}, K_{17})$
右脚	$V_7 = (K_9, K_{10}, K_{15})$

合的方式进行立定跳远的姿态差异化对比。第一，采用余弦相似度来对待测姿态和标准姿态中组成特征姿态的各个向量进行相似度度量；第二，分别计算标准姿态与待测姿态中关键特征姿态所构成的角度，并进行对比，给出改进意见。

首先，我们采取第一种方法，利用余弦相似度进行评估，其公式如下所示。

$$\cos \theta_i = \frac{V_i^t \cdot V_i^s}{|V_i^t| \times |V_i^s|} \quad (8)$$

其中 V_i^t 和 V_i^s 分别代表待测姿态和标准姿态中第 i 个特征向量的相似度，取值范围为 $[-1,1]$ 。1 代表构成的向量方向相同，完全相似；-1 则代表其方向相反。由于 AlphaPose 是一种二维人体姿态估计算法，得到的骨骼点为一组二维坐标，因此特征向量的模可由式 (9) 计算得出。

$$|V_i^t| = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (9)$$

根据立定跳远项目的特征，结合上文提到的优先度匹配规则，给予优先度高的特征更大的特征系数，最后按特征系数求取平均值作为评估结果，从而得到待测姿态和标准姿态骨骼关键点向量之间的总体相似度 Sim。

然后，我们采用第二种方法，分别定义标准姿态与待测姿态中关键特征姿态向量组成的角度。本文定义了 8 组角度用于组成计算每帧图片的得分，在原有的骨骼点中加入了水平线作为参考，这 8 组角度会根据测试者的外侧骨骼进行区分，若右侧骨骼在外侧，则计算右侧相关角度，反之使用左侧骨骼进行角度计算。定义的角度如表 3 所示。

表 3 关键特征姿态组成角度

组数	相关骨骼或参照
X1	(手腕, 手肘, 肩部)
X2	(手肘, 肩部, 臀部)
X3	(肩部, 臀部, 膝盖)
X4	(臀部, 膝盖, 脚踝)
X5	(膝盖, 脚踝, 脚尖)
X6	(脚尖, 脚后跟, 水平线)
X7	(肩部, 臀部, 水平线)
X8	(膝盖, 脚踝, 水平线)

分别求取标准姿态与待测姿态中关键特征姿态向量组成的角度，并进行对比分析，发现其差异。求取公式如下。

$$\theta_i = \arccos\left(\frac{A_i \cdot B_i}{|A_i| |B_i|}\right) \quad (10)$$

假设 θ_i 为左下肢的 $V_3 = (K_8, K_9, K_{10})$ 的夹角，由关键帧 P1 可知，标准姿态其夹角为 100 度。若待测人员姿态的角度大于 100 度，则需要在起跳前适当下压臀部以用来蓄力；若其角度小于 100 度，则需要提臀使身体重心保持平衡向前冲出，以取得更好的成绩。

分别计算待评估帧与标准帧中 8 组角度的具体数值，再根据下述公式进行组合，求得最后计算角度得分部分。

$$\rho = (\log_a \sum_{i=1}^8 |\Delta x_i| + 1)^{-1} \quad (11)$$

最后，我们综合上面三种评估方法，对上述 10 个关键帧分别计算评分，单帧的评分公式如下。

$$Score_i = \omega_1 \cdot \cos \theta_i + \omega_2 \cdot \rho_i \quad (12)$$

其中， $\cos \theta_i$ 和 ρ_i 分别代表第 i 帧特征姿态向量的相似度和各关键角度对比的得分。

之后结合皮尔逊相关系数评估结果，给出最终得分公式。

$$Score_{jump} = \omega_1 \cdot Sim + \omega_2 \cdot \rho + \omega_3 \cdot \delta_X + \omega_4 \cdot \delta_Y \quad (13)$$

3 实验结果与分析

本文试验采用的环境如表 4 所示。

表 4 实验环境介绍

	名称	型号
硬件平台	CPU 处理器	Intel Xeon W-2245
	GPU 图形卡	Quadro RTX 4000
	内存	32GB
	操作系统	Ubuntu 18.04.1 LTS
软件环境	开发环境	Anaconda 3+Python 3.8+CUDA 11.0+cuDNN+PyTorch 1.7
	第三方库	OpenCV 4.1+Numpy 1.2 等

在本次实验中，由 2 位待评估人员模仿标准姿态进行立定跳远评估，根据上文所示方法提取 10 个关键帧，将 10 帧中重点关注的骨骼关键点与标准姿态的相应点组成坐标序列，分别计算它们在 X 值和 Y 值的相关性，如表 5 所示。

表 5 各骨骼点 X 值 /Y 值相关性

骨骼点	测试者 1		测试者 2	
	X 值相关性	Y 值相关性	X 值相关性	Y 值相关性
K6 (左手腕)	0.8440	0.7496	-0.0799	0.8391
K5 (左手肘)	0.4561	0.5922	0.6476	0.1827
K4 (左肩)	0.6480	0.4941	0.6928	-0.0236
K11 (左臀)	0.9302	0.2286	0.8518	0.6400
K12 (左膝)	0.5684	0.8122	0.3345	0.8118
K13 (左脚踝)	0.9871	0.3696	0.9209	0.5469
K18 (左脚后跟)	0.9755	0.6718	0.9665	0.5625
K17 (左脚尖)	0.9794	-0.2009	0.9079	0.9256

然后根据式 (7) 计算待评估姿态与标准姿态总体相关系数。基于此计算每一帧中关键特征向量的相似度及各角度差值，并计算得分，将各帧得分情况绘制如图 13 所示。



图 13 立定跳远关键帧得分图

由此可以看出，测试者 1 相比测试者 2 在立定跳远起跳飞跃阶段得分更高，而在起跳前准备动作及落地时缓冲动作得分更低。

根据对关键特征向量的角度差及相似度进行分析可以得出，在第六帧中，测试者 1 和测试者 2 其特征向量相似度较为相近，其手腕、手肘和肩部组成的角度与标准动作角度差分别为 14 和 35，可知测试者 2 在第 6 帧动作中手肘未打直，且整体动作与标准动作差距过大，远不如测试者 1 的动作规范，因此第 6 帧其得分低于测试者 1；在第 4 帧

中,测试者1与测试者2各组成关键姿态的向量相似度分别为0.95和0.97,但其各关键特征姿态角度均较为相近,因此它们最终得分也较为相近。观察图14和图15可以发现,其结果较为准确。最后,参考式

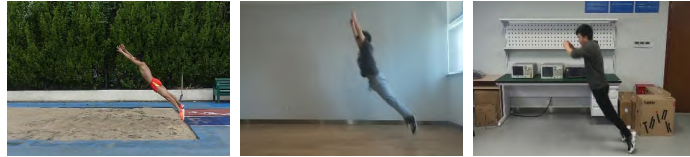


图14 第6帧对比图

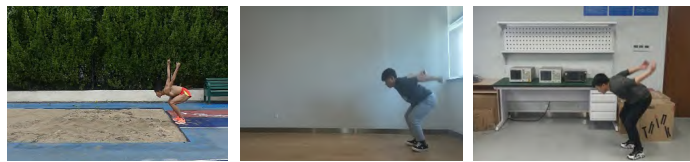


图15 第4帧对比图

(13),通过有效结合相关系数、特征向量相似度,以及关键姿态角度的影响,得到测试者1和测试者2的成绩分别为0.70和0.68,综合得分较为相近,测试者1比测试者2略好。经如上验证可知,本系统可以综合评估立定跳远待测人员的姿态,对其进行量化,并给出改进意见。

4 结束语

本文提出了一种基于人体姿态识别技术的立定跳远动作智能评估系统,该系统通过运用关键点数据预处理、立定跳远动作标准库的构建,以及关键帧匹配等方法进行姿态评估。在姿态评估方面,根据立定跳远运动的特征,将皮尔逊相关系数、特征向量相似度,以及关键特征姿态角度进行结合,从而对待评估姿态进行评估,输出其得分。就以上研究而言,本文还存在如下不足。

第一,本文考虑的影响因素过多,在计算时消耗资源过大,虽可避免因只考虑一种指标带来的评估不准确等问题,但仍可能存在两种评判方式相互抵消的问题,影响最后得分,应寻找更加高效合理的动作评估方法。

第二,人体姿态评估模型在某些特殊情况下存在骨骼关键点跳动等不稳定因素,未来可考虑改善人体姿态识别算法提高对比准确率。

参考文献

- [1] 新华社.关于全面加强和改进新时代学校体育工作的意见[R/OL].(2020-10-15)[2022-03-23].http://www.gov.cn/zhengce/2020-10/15/content_5551609.htm.
- [2] Piwek L, Ellis D A, Andrews S, et al. The Rise of Consumer Health Wearables: Promises and Barriers[J]. *PLoS Medicine*, 2016, 13(2): e1001953.
- [3] 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].中国科学院大学(中国科学院西安光学精密机械研究所), 2020.
- [4] 罗大为, 郑颖, 张姗.基于视觉的人体姿态评估方法研究与进展[C]// 2018年第二十二届网络新技术与应用年会.苏州:中国计算机用户协会网络应用分会, 2018.
- [5] Wang J, Qiu K, Peng H, et al. AI Coach: Deep Human Pose Estimation and Analysis for Personalized Athletic Training Assistance[C]// Proceedings of the 27th ACM International Conference on Multimedia. Nice ACM, 2019.
- [6] Kim T T, Zohdy M A, Barker M P. Applying Pose Estimation to Predict Amateur Golf Swing Performance using Edge Processing[J]. *IEEE Access*, 2020, 8: 143769-143776.
- [7] 胡建朗, 罗亚荣, 郭迟.一种用于运动姿态评估的视觉伺服机器人[J]. *测绘通报*, 2020, 52(8): 36-41.
- [8] 王莞.基于深度神经网络的人体运动姿态估计与识别[D].电子科技大学, 2020.
- [9] 杨君, 张素君, 张创豪, 黄晶晶.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): 11-14.
- [10] 陈学梅.基于人体三维姿态的动作评价系统[D].浙江大学, 2018.
- [11] Abraham. Savitzky, M. J. E. Golay. Smoothing and differentiation of data by simplified least square procedure[J]. *Analytical Chemistry*, 1964, 36(8): 1627-1639. DOI: 10.1021/ac60214a047.
- [12] glenn-jocher. Borda.yolov5[EB/OL].(2020-06-04)[2022-03-23].<https://github.com/ultralytics/yolov5/blob/master/LICENSE>.
- [13] Sun K, Xiao B, Liu D, et al. Deep High-Resolution Representation Learning for Human Pose Estimation[C]// 2019 IEEE/CVF

- Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, 15-20 June 2019, [S.1.]:IEEE, 2020: 5686-5696.
[14] Fang H S, Xie S, Tai Y W, et al. RMPE: Regional Multi-person Pose Estimation[C]// 2017 IEEE International Conference on Computer Vision (ICCV), Venice, 22-29 Oct. 2017, [S.1.]:IEEE, 2017: 2353-2362.
[15] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: Common Objects in Context[C]//European Conference on Computer Vision (ECCV 2014). Cham: Springer International Publishing, 2014.
[16] 解辉. 基于视觉的人体动作矫正系统[D]. 山东师范大学, 2020.



李 晟

中国移动（成都）产业研究院产品经理。硕士。研究方向为计算机数据与科学。



宋可儿

佐治亚大学计算机专业兼认知科学专业本科在读。研究方向为计算机视觉。



欧阳柏强

中国移动（成都）产业研究院算法工程师。硕士。研究方向为计算机视觉。



张博中

中国移动（成都）产业研究院算法工程师。硕士。研究方向为计算机视觉。



刘 宠

中国移动（成都）产业研究院算法工程师。硕士。研究方向为计算机视觉。



刘 朋

中国移动（成都）产业研究院项目经理。硕士。研究方向为计算机数据与科学。



刘 睿

中国移动（成都）产业研究院产品经理。学士。研究方向为计算机数据与科学。



在线阅读（节选）