

文章编号: 2095-2163(2022)03-0133-06

中图分类号: TP319

文献标志码: A

融合机器视觉与惯性传感器的智能康复评估技术

孙洪明, 郑建立, 尹梓名

(上海理工大学 健康科学与工程学院, 上海 200093)

摘要:为缓解患有运动功能障碍的患者不方便频繁前往医院进行康复评估的问题,本文提出一种基于“互联网+”的居家康复评估技术,为康复医师提供参考,节约康复医师和患者的时间,缓解医疗资源不足。针对脑卒中患者上肢康复量表的评估问题,仅使用惯性传感器或者机器视觉建立的长短期记忆人工神经网络分类模型在测试样本中准确率分别为55.6%、92.6%。本文在基于机器视觉方式获取人体的3D坐标的同时,通过惯性传感器获取肢体的方向并融合两者的数据,进一步采用长短期记忆人工神经网络分类模型对数据进行分析,在测试样本中取得了98.1%的分类准确率,取得了良好的实验效果。

关键词:脑卒中; 康复评估; 惯性传感器; 机器视觉; 长短期记忆人工神经网络

Intelligent rehabilitation assessment technology integrating machine vision and inertial measurement sensors

SUN Hongming, ZHENG Jianli, YIN Ziming

(School of Health Science & Engineering, University of Shanghai for Science & Technology, Shanghai 200093, China)

[Abstract] In order to alleviate the problem that patients with motor dysfunction are inconvenient to frequently go to the hospital for rehabilitation evaluation, this paper proposes a home rehabilitation evaluation technology based on "Internet +", which provides reference for rehabilitation physicians, saves the time of rehabilitation physicians and patients, and alleviates insufficient medical resources. For the evaluation of the upper limb rehabilitation scale for stroke patients, the long-term and short-term memory artificial neural network classification model established only by inertial measurement sensors or machine vision has an accuracy rate of 55.6% and 92.6% in the test samples, respectively. At the same time as the 3D coordinates of the human body, the orientation of the limb is obtained through the inertial sensor and the data of the two are fused. The long short-term memory artificial neural network classification model is further used to analyze the data, and the classification accuracy rate of 98.1% in the test sample is obtained. good experimental results.

[Key words] cerebral apoplexy; rehabilitation assessment; inertial measurement sensors; machine vision; long short-term memory

0 引言

脑卒中(Cerebral Apoplexy, CA),又称中风、脑血管意外(Cerebral Vascular Accident, CVA),是导致人类死亡的主要疾病之一^[1],超过50%的脑卒中患者因为运动功能障碍导致生活质量降低^[2]。脑卒中患者在康复治疗的过程中伴随着运动功能的评估,这种评估在确定运动功能状态和制定康复训练方案方面起着重要的作用。

目前,人体姿态识别的实现方式主要有3种:

基于可穿戴式的人体姿态识别。Zhang等^[3]使用单个固定在患者手腕上方的惯性传感器(Inertial Measurement Unit, IMU),自动收集患者运动过程中

的加速度,实验结果表明在脑卒中术后偏瘫康复患者上肢运动功能评定量表(Brunnstrom)分期上具有82.1%的准确度;沈天毓等^[4]使用5个传感器获取受试者6个范式下的加速度、角速度信号,统计结果表明在H-Y分期和UPDRS III评分的预测准确率分别为72.8%和68.64%;金琦等^[5]研制人体姿态设备,采用星型+总线的混合拓扑结构,把采集单元佩戴在人体全身各个关键部位,可用于人体姿态的捕捉;周瑞文等^[6]基于九轴IMU的动作捕捉硬件设计系统框架,可以实现人体的姿态估计和关节角估计。

基于机器视觉的体姿态识别。Seunghee Lee和Ali Ozturk等^[7-8]使用Kinect获取地标位置、关节角度等信息对脑卒中术后偏瘫康复患者使用Fugl-

基金项目:国家重点研发计划子课题(2020YFC2005800,2020YFC2005801)。

作者简介:孙洪明(1995-),男,硕士研究生,主要研究方向:医学信息技术;郑建立(1965-),男,博士,副教授,主要研究方向:医学信息系统与集成技术、医学仪器嵌入式控制系统;尹梓名(1986-),男,博士,讲师,主要研究方向:临床决策支持、医学人工智能。

通讯作者:郑建立 Email:zhengjianli163@163.com

收稿日期:2021-11-08

Meyer 量表进行上肢运动功能评定; Cao 等^[9]提出一种实时的方法(OpenPose),检测多人的 2D 姿态; Cheng^[10]等使用单个视频估计三维人体姿态,引入遮挡感知的深度学习框架,解决单目视频因为遮挡导致的三维人体姿态估计不准确的问题。

基于机器视觉与 IMU 的人体姿态识别。使用两者结合的方式,大都通过融合多视点视频和 IMU 传感器数据来估计三维人体姿态,不需要光学标记和复杂的硬件设置^[11-12]。Timo von Marcard 等^[13]提出了一种结合单个手持相机和附着在身体四肢的惯性测量单元来估计室外环境下 3D 人体姿态的方法;Zhang^[14]等使用多视角图像和附着在人体四肢的几个惯性测量单元来估计三维人体姿态。

针对脑卒中病人居家康复评估问题,本论文提出了一种应用于移动端的、融合机器视觉获取的单视频数据与固定在人体上肢的 IMU 数据的数据采集和智能分析方法,以弥补两种方式在单独康复评估过程中的劣势。从视频数据中获得无漂移的精准的 3D 人体坐标点信息;基于 IMU 提供完整的肢体

旋转、运动过程中的加速度和角速度的信息;旋转机器视觉的坐标系与 IMU 的坐标系对齐,保障数据采集的准确性,将采集的数据放到分类模型进行训练,以获得脑卒中患者的 Brunnstrom 分期量表的上肢智能康复评估结果。

1 系统框架和方法简介

1.1 系统框架

使用智能手机蓝牙获取绑定在患者上肢 IMU 传输的肢体方向等信息,即加速度、角速度、角度的变化,通过手机摄像头获取 BlazePose 算法捕捉的上肢运动轨迹,即肩关节、肘关节、腕关节的 3D 坐标点的变化。要采集患者的肢体方向和运动轨迹信息,需要配准 IMU 与机器视觉的坐标系,将 IMU 和机器视觉的数据融合,把融合后的数据输入长短记忆神经网络(LSTM),使用五折交叉验证提高所建立模型的可信度,最后得到 Brunnstrom 分期的结果。系统框架如图 1 所示。

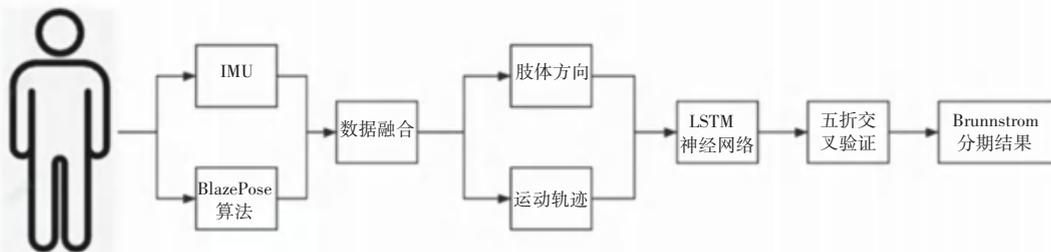


图 1 系统框架

Fig. 1 System framework

1.2 多节点可穿戴式惯性传感器方法

本文使用维特智能 BWT901BLECL 5.0 型号的九轴惯性传感器,该产品尺寸为 $36 \times 51.3 \times 15 \text{ mm}^3$,重量为 20 g,数据输出频率最高为 50 Hz,IMU 绑定的坐标系 X 轴指向上方、Y 轴水平向左、Z 轴垂直向外,如图 2(a)所示。通过蓝牙通信,可测量 3 轴的加速度、角速度和四元数信息,利用四元数计算相应传感器的姿态角,计算公式如式(1)~(3)所示。为满足患者居家康复评估的基本需要,尽可能降低康复成本,仅使用两个维特智能惯性传感器,用弹性绷带固定于患者手腕上方(距腕关节约 1/4 前臂长度距离处)和肩关节下方(距肩关节约 1/2 上臂长度距离处),采集患者上肢运动过程的数据,如图 2(b)所示。



(a) 维特智能传感器 (b) 传感器绑定位置

图 2 惯性传感器外形、坐标系定义及其绑定形式

Fig. 2 IMU shape, coordinate system definition and its binding form

$$\text{Pitch} = \tan^{-1} \frac{2(q_1 q_3 - q_2 q_4)}{1 - 2q_3^2 - 2q_2^2} \quad (1)$$

$$\text{Roll} = \sin^{-1} 2(q_3 q_4 + q_1 q_2) \quad (2)$$

$$Yaw = \tan 2^{-1} \frac{\varpi(q_2 q_3 - q_1 q_3) \ddot{\theta}}{\ddot{\theta} (2 q_1^2 + 2 q_2^2 - 1)} \quad (3)$$

其中, q_1, q_2, q_3, q_4 代表四元数, $Pitch$ 、 $Roll$ 、 Yaw 分别代表俯仰角、翻滚角、偏航角。

1.3 机器视觉传感方法

2020 年, google 研究人员在跨平台的机器学习框架 MediaPipe 上, 推出了用于单人人体姿态估计的算法 BlazePose, 该算法是可以推断出 33 个人体关键点的轻量级卷积神经网络, 在移动设备上的适配性表现突出^[15]。

与基于热图的技术相比, 基于回归的方法对计算的要求更少, 可扩展性更高, 但是其在预测关节的平均坐标值时, 常常存在无法解决的潜在的模糊性问题。Newell 等人^[16] 使用堆叠沙漏网络对人体关节点进行预测, 应用了更少的参数, 也取得了较好的预测效果, 但是该网络较大, 不利于在移动设备上运行。BlazePose 算法受到堆叠沙漏网络的启发, 采用热图、偏移量和回归相结合的方法, 但其只在训练阶段使用热图和偏移损失, 并在运行推理之前从模型中删除相应的输出层, 使其足够轻, 可以在移动端运行, 网络体系结构如图 3 所示^[15]。

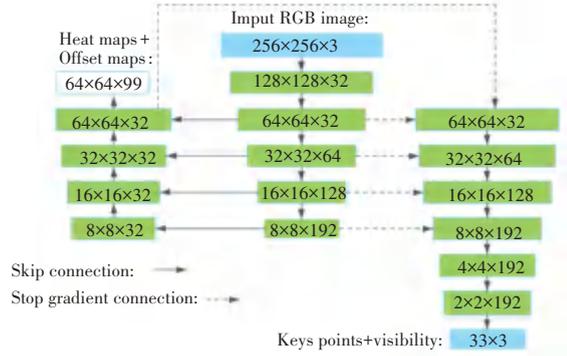


图 3 网络体系结构

Fig. 3 Network architecture

通过智能手机拍摄视频数据, 使用 BlazePose 算法对视频中的人物进行姿态识别, 获取人体的 3D 坐标点, 坐标原点位于视频图像的左上角, X 轴正方向水平指向视频图像的右方, Y 轴正方向垂直指向视频图像的下方, Z 轴垂直于视频图像向里。为减小手机摄像头与受试者之间距离不同, 导致所获取 3D 坐标点的差异, 将坐标原点换算到肩关节坐标与髋关节坐标的交点, 如图 4 所示。

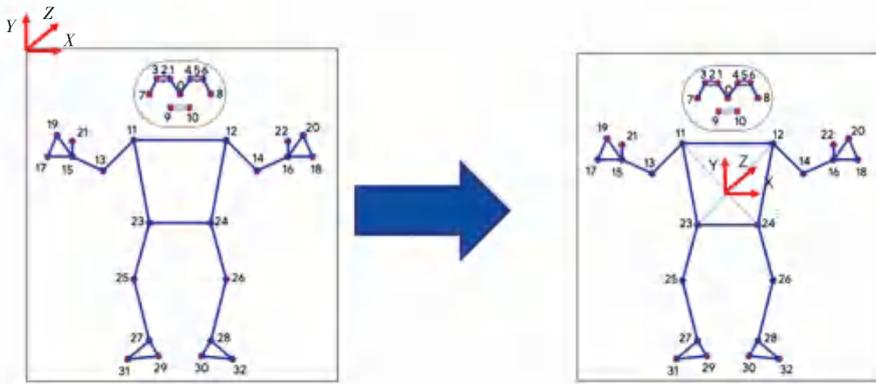


图 4 坐标原点平移转换

Fig. 4 Coordinate origin translation conversion

1.4 关联机器视觉与惯性传感器的数据

为提高采集数据的准确性, 将机器视觉与惯性传感器的数据进行关联, 旋转 BlazePose 的坐标系使其与惯性传感器的坐标系对齐。脑卒中患者固定好 IMU 后, 可在外力的辅助下, 肩关节外展 90° , 肘关节屈曲 90° , 通过将机器视觉所测量的二维平面上肘关节的角度与惯性传感器所测量 X 轴的关节角度做对比, 调整手机摄像头的倾斜角度与 IMU 的绑定位置, 进一步校准两者的坐标系。以右臂为例, 按式(4)使用患者手腕关节坐标点 (x_{16}, y_{16}, z_{16}) 和肘关节的坐标点 (x_{14}, y_{14}, z_{14}) 计算前臂

$IMU(WT1(x, y, z))$ 的位置、按式(5) 肘关节坐标点 (x_{14}, y_{14}, z_{14}) 和肩关节的坐标点 (x_{12}, y_{12}, z_{12}) 计算大臂 $IMU(WT2(x, y, z))$ 的位置, 降低因为不同患者上肢长度不同而导致数据采集的误差。

$$WT_1(x, y, z) = \frac{1}{4}(x_{14} - x_{16}, y_{14} - y_{16}, z_{14} - z_{16}) \quad (4)$$

$$WT_2(x, y, z) = \frac{1}{2}(x_{12} - x_{14}, y_{12} - y_{14}, z_{12} - z_{14}) \quad (5)$$

1.5 神经网络模型

长短期记忆网络 LSTM (Long Short - Term Memory) 是一种时间循环神经网络, 可以解决循环

神经网络(Recurrent Neural Network, RNN)在长序列训练过程中的梯度消失和梯度爆炸问题。因为RNN在某时刻的隐藏层信息只与当前时刻的输入和上一时刻的隐藏层信息有关,没有记忆功能,所以在处理序列较长的数据时,序列后面部分的梯度很难反向传播到前面的序列,产生梯度消失的问题。为了解决RNN的长期依赖问题,LSTM应运而生,引入了输入门、遗忘门、输出门,通过输入门和遗忘门的信息,可以控制LSTM长期记忆某个时间步(timestep)细胞状态的值,LSTM网络结构如图5所示。与RNN相比,LSTM能够在较长的时间序列中有更好的表现。

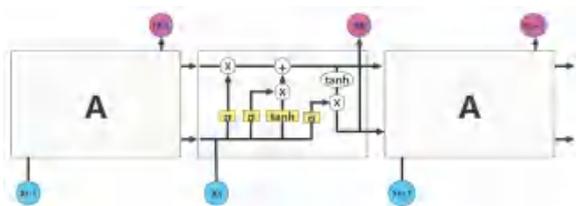


图5 LSTM神经网络结构

Fig. 5 LSTM neural network structure

2 实验验证

2.1 数据采集

由于在临床环境下对脑卒中患者康复评估涉及到患者隐私问题,且患者所处的Brunnstrom分期分布不均匀,导致很难收集到脑卒中患者真实的运动数据进行康复评估模型的训练,所以采取招募志愿者培训后模拟脑卒中患者各个Brunnstrom分期的方式进行实验数据的采集。招募的45名志愿者都需完成一次I~VI期Brunnstrom分期上肢运动过程的模拟。分析Brunnstrom量表可知,让脑卒中患者偏瘫侧做摸肩动作,其运动过程可以反应出脑卒中患者相应的Brunnstrom

分期^[17]。具体数据采集过程如下:

(1)初始位置:被试者上身挺直,采取坐立位,面向前方,颈部伸直,待测上肢自然下垂于身体一侧,掌心朝向前方,另一侧上肢自然放在同侧下肢之上,下肢放松屈膝90°;

(2)偏瘫侧的手臂慢慢抬高,使肩关节外展,外展至90°时,手掌呈中立位;

(3)若肩关节无法外展至90°,则屈肘90°,前臂旋前旋后;

(4)屈曲肘关节,手掌缓缓旋外,用手指触摸偏瘫侧的肩膀;

(5)肘关节伸展,手掌慢慢旋内,回到(2)终止时的位姿,上肢自然下垂回到初始位姿。

2.2 数据分析

Brunnstrom不同分期的主要差异是由肘关节、肩关节的活动度决定的。在康复评定过程中,机器视觉可获取上肢的运动轨迹,IMU可检测前臂和上臂的加速度、角速度、角度变化。角度变化能够直观地显示出脑卒中患者上肢关节的活动范围;V期患者在肘关节伸直的情况下,肩可外展90°;VI期患者运动协调程度近于正常人,可以完成摸肩动作。V和VI期患者在康复评定过程中前臂和上臂的活动范围和旋转情况如图6和图7所示,实线和无线条的一系列点分别代表前臂和上臂角度变化情况,蓝色线表示上肢与水平面(地面)的夹角,当与水平面的夹角超过180°以后,该夹角会变成-180°,绿色线和红色线分别表示上肢沿着冠状轴和矢状轴的旋转角度。此外,加速度、角速度的变化也可反应出患者在康复评定过程的运动细节,所以选取两个九轴传感器的加速度、角速度、角度和机器视觉获取的上肢关节坐标点共40个指标作为LSTM神经网络的特征值。

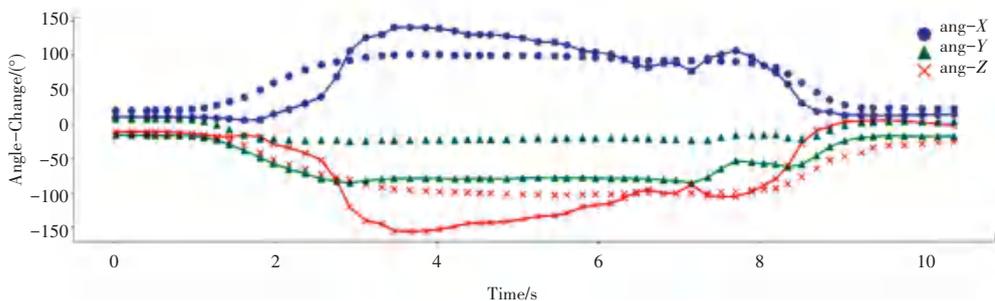


图6 Brunnstrom V期患者上肢肢体方向变化与运动轨迹

Fig. 6 The change of direction and movement trajectory of upper limbs in Brunnstrom stage V patient

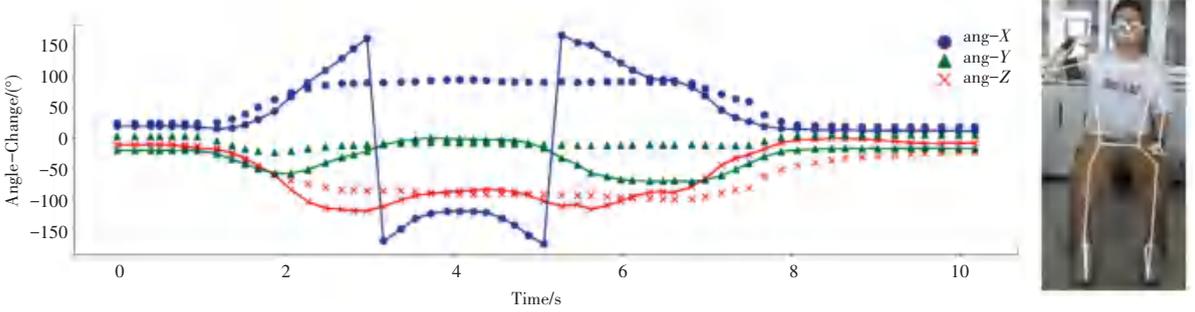


图 7 Brunnstrom VI期患者上肢肢体方向变化与运动轨迹

Fig. 7 The change of direction and movement trajectory of upper limbs in Brunnstrom stage VI patients

2.3 建立预测模型

本文使用 LSTM 神经网络分别对 IMU 采集数据、机器视觉采集的上肢关键点数据、两者结合的数据进行学习, 分别建立相应的脑卒中患者的 Brunnstrom 分期预测模型。45 个志愿者 I ~ VI 期的数据共 270 组, 根据所模拟的分期情况, I ~ VI 期的数据各 45 个。为减小数据集划分对分期模型精准度的影响, 通过五折交叉验证的方式, 将某期受试者按照 6.4:1.6:2 的比例划分为训练集、验证集和测试集, 并利用不同的数据集划分来对模型做 5 组不同的训练与验证, 提高 Brunnstrom 预测分期结果的可信度。

2.4 实验结果与分析

测试集样本总数为 54 个, 使用相应的数据对该数据训练的 Brunnstrom 分期模型进行验证, 不同数据源训练模型效果如图 8 所示。仅使用 BlazePose 算法获取的上肢关键点训练出的模型在测试集样本的准确率为 55.6%, 此模型对 III ~ VI 期识别效果极差, Brunnstrom 分期预测真实值与预测值的对比, 如图 9 所示。准确率较低的原因: 该模型不能检测到脑卒中患者在 Brunnstrom 量表上肢肢体的旋转和细微震颤的情况。只使用 IMU 数据训练的模型在测试集样本的准确率为 92.6%, 相对于前者, 准确率有所提高, 但是与两者结合后的数据训练的模型效

果相比还是有所欠缺。

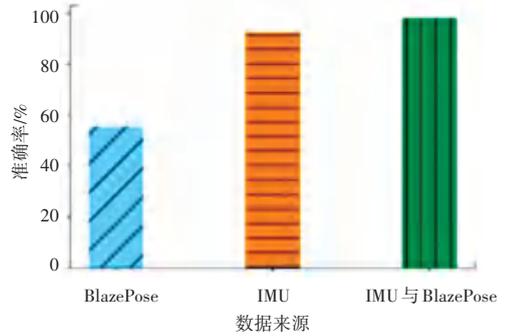


图 8 不同数据源训练效果对比

Fig. 8 Comparison of training effects of different data sources

IMU 与机器视觉结合的数据训练出的 Brunnstrom 分期模型在测试集的预测结果如图 9 所示, 仅仅一个 I 期脑卒中患者被预测为 II 期, 预测模型对 I 期患者预测的准确率为 88.9%, 其余分期预测的准确率为 100%, 模型综合预测的准确率为 98.1%。I 期患者预测出现偏差的原因: 首先, 脑卒中 I 期患者上肢虽然没有明显运动, 但是脑卒中患者在发起运动的过程中可能会发生震颤, IMU 捕获到上肢震颤的信号, 在模型训练的过程中, 导致与 II 期仅出现轻微联合反应运动的患者情况混淆; 其次, 训练模型的数据量不足, 也对预测的准确性造成影响。测试集的分类结果表明, 本文方法能够较好地实现 Brunnstrom 分期预测。

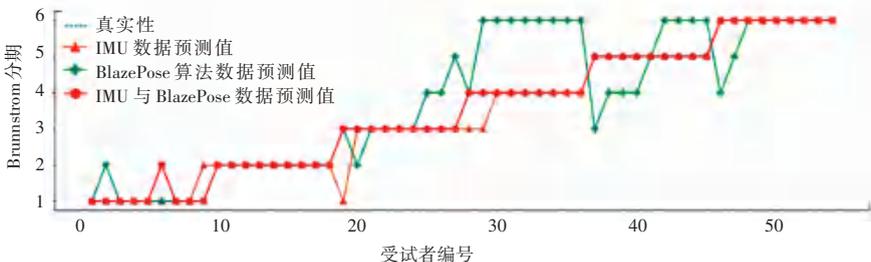


图 9 Brunnstrom 分期预测真实值与预测值的对比

Fig. 9 Comparison of Brunnstrom's staged forecasts between real and predicted values

3 结束语

为了实现脑卒中患者远程的居家康复评估,本文利用智能手机和惯性传感器来对脑卒中患者进行远程居家Brunnstrom分期预测,避免脑卒中患者频繁前往康复医院进行康复评估,有利于康复医师了解到患者的康复情况,及时制定合适患者的康复计划。在未来研究中,会将该系统推广到临床中,建立更加真实、全面的脑卒中患者的康复过程的数据库,提高预测模型的适用性和准确率。此外,陆续开展对手部功能的预测评估研究,期望建立一个更完善的基于“互联网+”的远程康复评估系统。

参考文献

- [1] FEIGIN V L, FOROUZANFAR M H, KRISHNAMURTHI R, et al. Global and regional burden of stroke during 1990–2010: findings from the Global Burden of Disease Study 2010[J]. *The Lancet*, 2014, 383(9913): 245–255.
- [2] WRITING GROUP MEMBERS, LLOYD-JONES D, ADAMS R J, et al. Heart Disease and Stroke Statistics—2010 Update: A Report From the American Heart Association [J]. *Circulation*, 2010, 121(7): 46–215.
- [3] ZHANG Z, FANG Q, GU X. Objective Assessment of Upper-Limb Mobility for Poststroke Rehabilitation[J]. *IEEE Transactions on Biomedical Engineering*, 2016, 63(4): 859–868.
- [4] 沈天毓, 王计平, 郭立泉, 等. 利用可穿戴设备对帕金森病患者运动功能进行量化评估[J]. *生物医学工程学杂志*, 2018, 35(2): 206–213.
- [5] 金琦. 基于IMU的人体姿态实时跟踪与可穿戴式设备研究[D]. 杭州: 浙江大学, 2020.
- [6] 周瑞文. 基于惯性测量单元的三维动作捕捉系统关键技术研究[D]. 哈尔滨: 哈尔滨工业大学, 2020.
- [7] LEE S, LEE Y, KIM J. Automated Evaluation of Upper-Limb Motor Function Impairment Using Fugl-Meyer Assessment [J].

- IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2018, 26(1): 125–134.
- [8] OZTURK A, TARTAR A, ERSOZ HUSEYINSINOGLU B, et al. A clinically feasible kinematic assessment method of upper extremity motor function impairment after stroke [J]. *Measurement*, 2016, 80: 207–216.
- [9] CAO Z, HIDALGO G, SIMON T, et al. Open pose Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields [J]. *arXiv:1812.08008 [cs]*, 2019.
- [10] Cheng Y, Yang B, Wang B, et al. Occlusion-Aware Networks for 3D Human Pose Estimation in Video [C]// 2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2019: 723–732.
- [11] GILBERT A, TRUMBLE M, MALLESON C, et al. Fusing Visual and Inertial Sensors with Semantics for 3D Human Pose Estimation [J]. *International Journal of Computer Vision*, 2019, 127(4): 381–397.
- [12] MALLESON C, GILBERT A, TRUMBLE M, et al. Real-Time Full-Body Motion Capture from Video and IMUs [C]// 2017 International Conference on 3D Vision (3DV). Qingdao: IEEE, 2017: 449–457.
- [13] VON MARCARD T, HENSCHER R, BLACK M J, et al. Recovering Accurate 3D Human Pose in the Wild Using IMUs and a Moving Camera [C]// ECCV 2018. Cham: Springer International Publishing, 2018: 614–631.
- [14] ZHANG Z, WANG C, QIN W, et al. Fusing Wearable IMUs With Multi-View Images for Human Pose Estimation: A Geometric Approach [C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020: 2197–2206.
- [15] BAZAREVSKY V, GRISHCHENKO I, RAVEENDRAN K, et al. BlazePose: On-device Real-time Body Pose tracking [J]. *arXiv:2006.10204 [cs]*, 2020.
- [16] NEWELL A, YANG K, DENG J. Stacked Hourglass Networks for Human Pose Estimation [J]. *arXiv:1603.06937 [cs]*, 2016.
- [17] 杨年峰, 王人成, 金德闻, 等. 基于Fitts定律的人体上肢运动功能评定方法[J]. *中国康复医学杂志*, 2001(6): 13–16.

(上接第132页)

参考文献

- [1] 中国气象局. 中华人民共和国气象行业标准: 能见度等级和预报: QX/T 114-2010[S]. 北京: 气象出版社, 2010: 1–3.
- [2] IOAN S, RAZVAN-CATALIN M, FLORIN A. System for visibility distance estimation in fog conditions based on light sources and visual acuity [C]// 2016 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR), 2016: 1–6.
- [3] CHENG X G, YANG B, LIU G Q, et al. A variational approach to atmospheric visibility estimation in the weather of fog and haze [J]. *Sustainable Cities and Society*, 2018: 215–224.
- [4] NNOLIM U A. Adaptive multi-scale entropy fusion de-hazing based on fractional order [J]. *Journal of Imaging*, 2018, 4(9): 1–22.
- [5] KAREVAN Z, SUYKENS J A K. Transductive LSTM for time-

- series prediction: An application to weather forecasting [J]. *Neural Networks*, 2020, 125: 1–9.
- [6] KASHIWAO T, NAKAYAMA K, ANDO S, et al. A neural network-based local rainfall prediction system using meteorological data on the internet: A case study using data from the Japan Meteorological Agency [J]. *Applied Soft Computing*, 2017, 56: 317–330.
- [7] SELVARAJ P, MARUDAPPA P. Prediction of rainfall using intensified LSTM based recurrent neural network with Weighted Linear Units [J]. *Atmosphere*, 2019, 10(11): 668.
- [8] AKMAL JON P, CHO Y I. VisNet: Deep convolutional neural networks for forecasting atmospheric visibility [J]. *Sensors*, 2019, 19(6): 1–34.
- [9] ZHANG Z H, YUE S G, ZHANG G P. Fly visual system inspired artificial neural network for collision detection [J]. *Neurocomputing*, 2015, 153: 221–234.