IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

审稿中：出版日期 xxxx 00, 0000，当前版本日期 xxxx 00, 0000。数字对象标识符
10.1109/ACCESS.2022.0122113

用于人体运动分析的真实工业任务与传统手工艺动作捕捉基准测试

BRENDAELIZABETHOLIVAS-PADILLA[1],ALINAGLUSHKOVA[1],以及 SOTIRIS MANITSARIS[1]机器人中心，MinesParis，PSL大学，法国巴黎 75006（电子邮箱：name.lastname@minesparis.psl.eu ）
通讯作者：Brenda Elizabeth Olivas-Padilla（电子邮箱：brenda.olivas@ minesparis.psl.eu）。

本研究取得的成果获得了 CARNOT 项目联合计划“负责任工厂”以及“地平线 2020”研究与创新计划的资助，资助协议编号为 820767（CoLLaboratE 项目）和 822336（Mingei 项目）。

arXiv:2304.03771v1 [cs.RO] 3 Apr 2023

摘要 人体运动分析是机器人学、生物力学和数据科学的关键研究领域，涵盖追踪、姿势估计和运动合成。尽管多种方法随着时间推移不断发展，但要确定当前的先进水平，仍需利用可验证的三维人体运动真实数据，对这些方法进行系统、定量的评估。本文介绍了七个采用惯性式动作捕捉记录的数据集。这些数据集包含工业操作人员和熟练工匠在真实现场条件下完成的专业手势。创建这些数据集的目的是将其用于人体运动建模、分析和生成研究。文中详细介绍了数据采集规程，并将对所采集数据的初步分析作为基准。手势操作模型是一种基于运动学描述符的混合随机—生物力学方法，用于对专家动作的动态特征进行建模，并构建其运动轨迹的数学表示，以便分析并量化其身体灵巧性。这些模型能够准确生成人体专业姿势，并直观地描述在执行任务过程中身体关节如何协同配合及随时间变化。

索引词 历史工艺、人体运动生成、工业任务、惯性传感器、动作捕捉数据集及真实场景。

I. 引言

以往对人体运动数据的研究帮助研究人员更好地理解人体动力学及其随机行为。采集不同情境下人体运动的原始数据，有助于量化并深入理解运动参数，以及影响运动表现的因素。通过分析这些数据，可以揭示隐藏参数，这些参数可用于体育运动、康复和艺术领域的运动评估，也可用于更专业的工业场景中的人体工学监测。与人体工学相关的职业疾病，例如肌肉骨骼疾病（MSDs），是一个重要问题，不仅会损害操作人员的健康，也会降低工厂/车间的生产效率。

在这种背景下，许多动作捕捉项目这些工作已在各个领域开展。其中大多数数据集都已公开，供可视化和分析使用，涵盖人类日常生活中的各种活动，从丰富的身体动作到细微的面部表情。这些数据集可根据所用技术（如基于标记点或无标记点的动作捕捉等）、记录的活动（日常活动、体育运动等）或用户数量（单用户或多用户交互）进行分类。例如，HumanEva 和 MoVi [1], [2]是两个现有数据集，包含单人进行日常活动（如行走和慢跑）及体育运动（拳击）时的视频和基于标记点的动作捕捉（MoCap）数据。HMDB51 [3]数据集还使用单目相机记录了一般身体动作，包括用户与物体交互、

人与人交互和面部表情。MHAD [4] 多模态人体动作数据库已发布，其中包括通过摄像机、加速度计和麦克风记录的肢体动作。另一项较新的工作也被纳入 CMU数据集 [5]„ 其中还包含食物准备活动的多模态信号。针对多人交互动作和场景，van derAa 等人 [6] 提出了UMPM基准。此外，KIT数据集记录了操作各种物体时人与人交互的活动 [7]。

在上述所有研究中，研究动作捕捉的学者都将兴趣集中在日常活动上。然而，在过去十年中，人们越来越感兴趣的是运用动作捕捉，并将数据分析方法应用于受专业场景启发的情境，在这些情境中，人类操作员执行各自的任务。下表列出了近期记录面向工业的人体运动数据的研究成果。建筑行业中有若干实例，因为它是受高强度体力活动影响最严重的行业之一。

为检测过度负重任务，Lee等人 [8]专注于创建一个使用非侵入式单个惯性测量单元的数据集。记录的数据已用于使用卷积神经网络自动预测负重重量和姿势。记录了14名受试者执行六种不同负重模式的情况。该数据集的分析包括对负重重量进行建模、分类和预测。另一个有趣的数据集是在AnDy欧盟项目[9]，框架下记录的，其中使用了各种传感器，例如包括用于手指运动的手套在内的全身IMU动作捕捉服、基于标记点的动作捕捉系统、手指压力传感器以及2台视频摄像机。受试者执行了受汽车制造启发的面向工业的活动。数据经过注释和标注，旨在供开发用于分类、预测或评估工业环境中人体运动的算法的研究人员使用。评估主要集中于标签可靠性，而不是运动分析本身。VTT-Conlot数据集包含受建筑行业启发、使用3个惯性测量单元记录的动作数据，共有13名受试者 [10]。该数据集的主要目标是用于活动识别和分类。其评估涉及传感器位置、所使用的模态以及提取的特征。然而，与前面所引用的示例不同，VTT-ConIot还使用真实建筑工地上真实工人的真实未标注数据对其数据进行了验证和比较（真实数据不包含在VTT-Conlot数据集中）。IKEAASM数据集是一个多视角家具组装视频数据集，其中包括深度信息、原子动作、物体分割和人体姿态 [11]。该数据集的特点之一是，它包含组装家具时所做出的不寻常人体姿势，但是其中不包含任何通过 IMU捕获的数据，主要旨在解决计算机视觉方面的挑战。WGD数据集提供了使用基于标记点的系统记录的数据，这些数据来自执行装配线作业活动的受试者 [12]。研究人员对这些数据进行了运动学评估，结果表明，该数据集可用于人体工效学评估。

上述所有研究都不再局限于记录日常活动，而是聚焦于专业任务／手势／姿势。然而，我们仍需要涵盖更多样化动作的动作捕捉数据，尤其是在真实场景中捕捉的专业手势。现有的大多数数据集都是在实验室内录制的，因此测量结果只是近似值，因为这些数据可能缺乏真实性，也并非来自真实的工作场所。因此，本文介绍了在专业环境和真实环境中创建的数据集，用于捕捉和研究操作人员与工匠在实际工作条件下的手势。

本文详细介绍了记录流程和处理方法。此外，本文还首次使用一种名为手势操作模型（GOM）的解析模型，对七个数据集进行了分析。该模型在先前的工作中提出 [13]。在这项分析中，基于 GOM 构建了可解释的运动表征，可用于人工生成人体动作，并解释模型所描述动作执行过程中各关节的协同作用。结果包括每个数据集的预测性能指标，以及对专业任务的灵巧度分析。灵巧度分析运用 GOM 的数学表征来描述专业手势的执行表现。灵巧度可定义为运用双手或其他身体部位完成给定动作或任务的技能。此外，本文还介绍了一种方法，用于确定在一组人体动作建模与识别中最重要的关节运动描述符。这些知识随后可用于确定人体动作识别问题的理想传感器配置。

II. 数据采集

本节首先介绍用于记录的动作捕捉系统，随后介绍各数据集中被采集的受试者和手势。

A. 动作捕捉技术

Nansense Inc.的BioMed套件动作捕捉系统[1] 被用于捕捉工业操作人员和工匠的手势。该系统由一套全身套装组成，配备了52个惯性测量单元，战略性地分布在躯干、四肢和手部。传感器以每秒90帧的速率测量身体节段的方向和加速度，在

[1]Baranger Studios，洛杉矶，加利福尼亚州，美国

表 1. 社区可用和采用的数据集。

数据集	所用技术	记录的活动	年份 出版物	# of 受试者	环境类型 数据已 采集于
DeTECLoad	单个惯性测量单元	建筑工人 负重任务	2020	14	受控条件 在实验室中
AnDy 项目	基于标记的动作 捕捉 系统 4 个压力传感器 2 个 RGB 相机 全身惯性测量单元 + 手套	面向工业的活动 （高/中/低位螺旋动作， 解开绳结等。）	2020	13	受控条件 在实验室中
VV-Conlot	RGB 单目 相机 3 个惯性测 量单元	面向建筑行业 （绘画、吸尘、 从楼梯上跳下等）	2021	13	受控条件 在实验室中
IKEA ASM	3 台 RGB 相机	家具组装 （拧螺丝等）	2020	48	受控条件 在实验室中
WGD	基于标记的动作 捕捉系统，配备 8 台摄像机	装配线作业手势 （敲击、拧螺丝等）	2021	8	受控条件 在实验室中

表 2. 生成的数据集概览。

数据集	记录的活动	地点	# of 受试者
电视机组装	在生产线上钻孔、连接组件 等。	土耳其，Arçelik 工厂	5
飞机浮筒组装	敲击铆钉，将 铆钉顶铁等放置到位。	罗马尼亚，Romaero 工厂	2
丝绸织造	使用提花织造手势 不同尺寸的织机	德国，克雷费尔德丝绸博物馆	2
玻璃吹制	塑造醒酒器，向.....吹气 吹管等。	法国，Cerfav 玻璃吹制工坊	1
乳香树脂采集	扫拂土壤，刺绣 树木 ，etc.	希腊 ，希俄斯岛 ，乳香博物馆田地（户外）和 在受控的室内条件下	2
根据姿势 EAWS 协议的姿势	弯腰、转动躯干等。	实验室中的受控条件	10

关节式脊柱链、肩部、上臂、腿部和指尖。录制完成后，通过NansenseStudio的逆运动学求解器自动计算 X、Y和Z轴上的欧拉局部关节角度，并以 Biovision Hierarchy 格式（BVH）存储。BVH文件是由两部分组成的文本文件。第一部分提供骨骼的层级描述，从根部（髋部）开始，延伸至每条肢体的末端。文件的第二部分包含录制中每一帧的骨骼根部绝对位置，以及 BVH 文件第一部分中定义的关节角度。

B. 招募的受试者

在创建每个数据集时，工业操作人员和熟练工匠均同意穿着 Nansense 服装，在实际工作场所接受录制，并遵循《通用数据保护条例》（GDPR）的原则。首先，研究人员录制了土耳其伊斯坦布尔一家电视机工厂和罗马尼亚布加勒斯特一家航空航天公司的工业操作人员执行专业任务的过程。电视机工厂的动作捕捉录制环节共有四名健康人士参与，其中三名男性和一名女性。他们的平均年龄为 31.5±6.2 岁，身高为 167.8±4.6 厘米，平均体重为 65.3±9.9 千克。航空航天公司的录制由两名男性受试者参与动作捕捉录制环节。他们的平均年龄为 50±5岁，身高为 170±2厘米，体重为77±1.4 千克。

十名健康人士同意在中性环境实验室中参与对具有潜在危险的人体工学姿势进行动作捕捉记录。受试者由三名女性和七名男性组成。平均年龄为 28.7±4.6 岁，平均身高为 172.9±9.2 厘米，平均体重为 70.5±12.9 千克。所有人均未遭受肌肉骨骼损伤，并在一小时内完成了所有试验。

熟练工匠从事三种不同工艺时的手势被记录下来。第一位是在德国克雷费尔德一家传统提花织造工坊中被记录的丝织工大师。这位专家身高168厘米，体重62千克。第二位工匠是一名玻璃吹制大师，他在进行玻璃吹制时被记录下来。

工坊。玻璃吹制工的身高为177厘米，体重为73千克。最后，在希腊希俄斯的一处乳香树脂采集地记录了两名乳香种植者。他们的平均年龄为30.5±5.5年，身高为178.8±8.5厘米，平均体重为69.3±8.0千克。

C. 专业任务的录制

下面概述每次录制所遵循的流程，并介绍每项捕捉到的任务。录制前先进行校准。受试者摆出不同的姿势，例如I字姿势或T字姿势，并完成不同的动作，例如行走或触碰指尖，每项持续10秒。为便于之后对数据进行标注和分割，仅要求操作人员和工匠在录制前讲解任务的各个组成部分。

1) 工业相关任务

已记录在两个工业环境中执行的手势，在操作人员执行工业任务时呈现自然动作。这些任务由实际操作人员在常规生产过程中于现场采集。

a：电视机制造

在一家电视机制造厂记录了与电视组装和包装相关的两项任务。每项任务所涉及的手势集合分别用缩写 TVA（组装）和TVP（包装）表示。图 1 展示了电视组装和包装过程中记录的一些手势。

电视组装任务包括将电子电路板安装到电视机机箱上，并使用电动工具将螺钉拧入电路板，使其牢固固定。针对这项任务，定义了以下手势词汇：

- 电视机A_1：单手伸到肩部以上的高处，从容器中取出一个组件（电路板）。
- 电视机A_2：用另一只空着的手向下伸至膝盖以下，从第二个容器中拿起第二个部件（电线）。
- 电视机A_3：连接各个部件，并将电路板放在底盘上以便拧紧。
- 电视机A_4：右手握持电钻在电路板上钻入四颗螺丝，左手放置螺丝。

最后一道工序是将装入纸箱的成品电视机堆叠在木托盘上，并用塑料薄膜包裹，以便运输（TVP）。本任务记录了以下一组手势：

- TVP_1：将八台电视机放在木托盘上（底层）。
- TVP_2：准备在底层包覆一层薄膜。
- TVP_3：正在包覆底层。
- TVP_4：在底层上方放置八台电视机（第二层）。
- TVP_5：用塑料薄膜包裹第二层。
- TVP_6：在第二层上方放置八台电视机（第三层）。
- TVP_7：用塑料薄膜包裹第三层。
- TVP_8：将八台电视机放置在第三层（第四层）顶部。
- TVP_9：用塑料膜包裹第四层。

箱子通过传送带送给操作员。操作员用双手一次将一个箱子放到托盘上。在同一层堆叠八个箱子后，他用双手抓住塑料膜，绕着箱子将其包裹起来。包裹妥当后，操作员继续在已包裹的上一层堆叠箱子，并重复这一过程。当托盘上堆叠了四层箱子时，任务即完成。

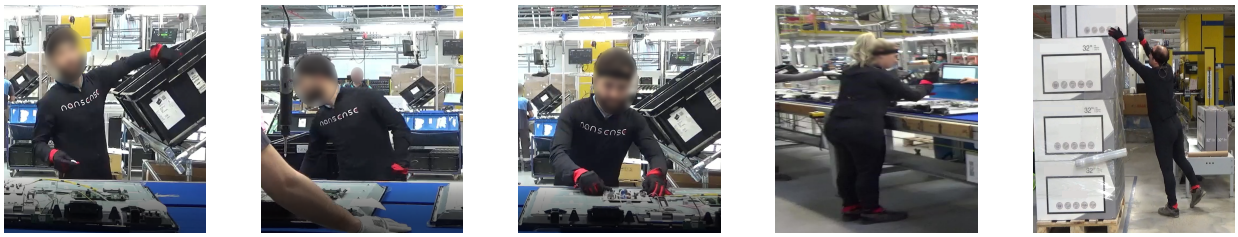
所有与电视组装相关的任务均在一个八小时班次期间进行记录：一名受试者在班次的前半段负责安装电路板，另一名受试者在后半段负责将电路板钻孔并安装到电视机机架上。另有三名受试者在一个班次期间分别执行包装任务并接受记录。

b：飞机浮筒组装

飞机浮筒的完整铆接任务是在一家航空航天公司中进行捕捉记录的。浮筒是飞机的一种部件，能够使飞机在水面降落时漂浮。该任务中记录的一组手势被 обознач 为 APA，相关内容也见图 2。两名操作人员之间的协作对完成这项活动至关重要。因此，他们的数据是按顺序采集的：一人先穿上动作捕捉服，在协作过程中记录其动作；随后将动作捕捉服交给第二人穿戴，并继续进行该活动。最终记录了以下手势，这些手势也在图 2 中进行了展示：

- APA_1：用气动锤铆接铆钉。
- APA_2：准备好气动锤并拿取铆钉。
- APA_3：将铆钉顶铁放在对应位置，以抵住即将打入的铆钉。

一次铆钉装配迭代包括第一名操作员将一枚铆钉放入一个孔中（图2a）。随后，来自浮筒另一侧的第二名操作员将铆钉顶铁放置在铆钉对面（图2c）。铆钉顶铁准确定位后，第二名操作员示意第一名操作员启动气动锤。第一名操作员通过触摸来确认已装配铆钉的位置正确



(a) 电视机$_1$ (b) 电视机$_2$ (c) 电视机$_3$ (d) 电视机$_4$ (e) TVP$_8$

图 1. 电视机制造中的专业手势。



(a) APA$_1$ (b) APA$_2$ (c) APA$_3$

图 2. 飞机组装手势示例。

完成后，操作员移至下一个孔位并重复这一过程。完成一排铆钉后，第一位操作员抓取更多铆钉，并为第二排准备气动锤（图 2b）。

由于操作员佩戴动作捕捉手套时无法真实地进行操作，因此未记录使用气动锤铆接过程中手指的动作。操作员需要用裸手触摸铆钉，以确认其是否正确就位。

c：不同人体工学风险等级的姿势

设计了一套记录方案，用于采集基于欧洲装配作业表（EAWS） [14]的28种不同人体工学风险等级的姿势。

每种姿势重复三次，因此每位受试者共进行了84次动作捕捉记录。记录的姿势为中性的，并未与特定活动相关，而仅用于展示若干不符合人体工学的姿势。这些姿势可分为三大类：站立时完成的姿势、坐在椅子上完成的姿势，以及跪姿时完成的姿势。这些姿势从舒适姿势逐渐过渡到越来越不舒适但始终不会造成危险的姿势。所有姿势均保持六秒，且未报告任何明显不适。这组具有不同人体工学风险等级的28种姿势称为ERGD。图3展示了受试者采取的三种姿势。起初，受试者站立，背部挺直。随后，受试者采取以下三种姿势：

- ERGD$_1$：受试者保持直立站立，上臂自然放松（I字姿势）。
- ERGD$_2$：受试者将躯干尽可能向左旋转，并保持六秒。
- ERGD$_3$：受试者将躯干向左侧弯曲，并保持六秒。

接下来的三种姿势中，躯干均略微向前弯曲：

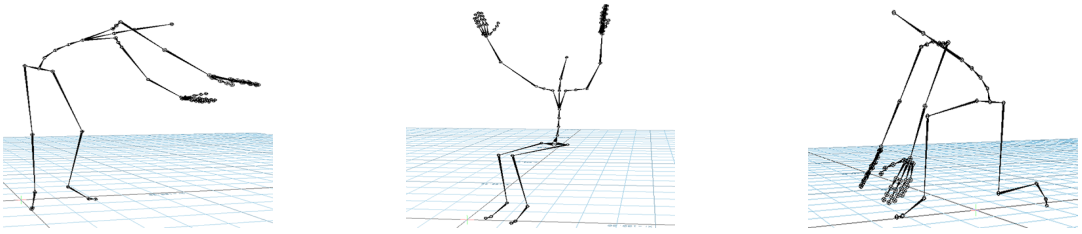
- ERGD$_4$：受试者保持弯曲姿势六秒。
- ERGD$_5$：受试者向前弯腰时，将躯干向左旋转并保持此姿势六秒钟。
- ERGD$_6$：当受试者向前弯腰并将躯干向左转动时，他们伸出手臂，仿佛试图够取地上的某样东西。

接下来的三种姿势中，躯干大角度前倾（$> 60°$）：

- ERGD$_7$：受试者保持弯曲姿势六秒钟。
- ERGD$_8$：受试者向前弯腰时，将躯干向左旋转并保持该姿势六秒。
- ERGD$_9$：当受试者向前弯腰并将躯干向左旋转时，他们伸出手臂，仿佛试图够到地上的某样东西。

在接下来的几种姿势中，上臂的位置会发生变化，而躯干姿势将重复出现：

- ERGD$_{10}$：受试者直立站立，前臂在 $90°$ 处弯曲，上臂抬至肩高，与地面垂直。



(a) ERGD$_7$ (b) ERGD$_{19}$ (c) ERGD$_{28}$

图 3。ERGD中姿势的示例。

- ERGD$_{11}$：上臂保持与 P10 相同的位置，受试者旋转躯干，并向左侧弯。
- ERGD$_{12}$：参与者将上臂抬至与地面垂直，前臂完全伸展。随后，参与者通过旋转并向左侧屈躯干。
- ERGD$_{13}$：受试者将上臂举过头顶，持续六秒。
- ERGD$_{14}$：双臂高举过头顶，受试者向左旋转并侧弯，持续六秒。

这些都是从站立姿势中采取的姿势。下一部分介绍将在人员坐在椅子上时记录的姿势。

- ERGD$_{15}$：此人坐在椅子上，上臂放松（中立位置）。
- ERGD$_{16}$：坐姿时，受试者向前弯曲的角度为 60° 或以上。
- ERGD$_{17}$：受试者向前弯曲60° 或更大角度，同时旋转躯干并向左侧弯。
- ERGD$_{18}$：受试者重复 P17，但将上臂伸向身前。
- ERGD$_{19}$：受试者将上臂举过头顶，并完全伸直。
- ERGD$_{20}$：双臂举过头顶时，受试者将躯干向左旋转并向左侧弯。

最后，受试者将以右膝跪姿完成其余姿势。这些姿势在人体工学上最令人不适。此外，上半身的动作选项与之前相同：

- ERGD$_{21}$：受试者保持直立。
- ERGD$_{22}$：受试者尽可能向左旋转躯干，并保持该姿势六秒。
- ERGD$_{23}$：受试者将躯干向左侧弯曲。
- ERGD$_{24}$：受试者向前弯曲的角度大于 60°。
- ERGD$_{25}$：躯干弯曲角度大于 60° 时，受试者将躯干向左旋转并侧弯。
- ERGD$_{26}$：重复 P25 姿势，但这次受试者伸出上臂，仿佛要从地面捡起什么东西。
- ERGD$_{27}$：受试者抬起上臂，使其与地面垂直。
- ERGD$_{28}$：受试者抬起上臂，同时旋转躯干并向左侧弯曲。

录制完成后，ERGD 收录了从最舒适的姿势到根据 EAWS 定义的风险因素评定的最不符合人体工学的姿势示例。尽管这些姿势并非针对任何特定目标，但可作为基准，用于测试不同的人体工学评估方法。

2）传统手工艺任务

我们记录了工艺大师和乳香种植者在各自真实的工作场所执行专业任务的过程。此外，还进行了一次动作捕捉，以记录不使用任何材料或工具模拟乳香栽培任务的过程。

a：丝绸织造

在德国克雷费尔德的一家提花织机工坊中，我们记录了一位熟练丝织工的手势。这组记录下来的手势以 SLW 为代称，其中一些示例见图 4。在为期三天的记录中，专家完成了以下与丝绸织造相关的任务：

1) SLW$_1$：制作穿孔卡片。2) SLW$_2$：缠绕织轴。3) SLW$_3$：准备织轴。4) SLW$_{4,1:3}$：使用不同尺寸（小、中、大）的织机进行提花织造。

第一天，丝织工被记录下来连续进行 SLW$_1$, SLW$_2$, 和 SLW$_3$ 。穿孔卡片的制作过程被记录了一个小时。由于任务的复杂性和持续时间较长，丝织经轴的缠绕和准备过程只记录了一次，记录时长约为四小时。接下来的两天则是对该专家进行连续记录



(a) SLW_1 (b) SLW_3 (c) $SLW_{4,3}$ (d) $SLW_{4,1}$

图4. 记录的提花织造手势示例。

使用三种不同尺寸的织机进行织造。只有当织工换用另一台织机时，记录才会停止。使用织机织造的任务可分为三个主要手势（$SLW_{4,1}$，$SLW_{4,2}$，和 $SLW_{4,3}$）。首先，专家用右腿踩下踏板，同时用左手拨开纱线（织工的初始姿势如图 4c 和 4d 所示）。然后，他用右手控制让纱线水平穿过的梭子，以快速拉动的手势将梭子送到另一侧。最后，他用左手拉回纱线，同时用右腿松开踏板。这个过程会重复进行，直到织造完成。

b: 玻璃吹制

玻璃醒酒器的制作过程在法国瓦讷勒沙泰尔的一家欧洲玻璃工艺研究与培训中心记录了四次。由于整个过程中必须保持玻璃的温度，每次试验都连续记录，手势之间没有暂停。因此，每次尝试都生成了一个动作文件，从采集熔融玻璃开始，到将醒酒器放置冷却时结束。组成醒酒器制作过程的一组手势记为 GLB。图 5 展示了醒酒器制作过程中记录的一些手势。玻璃醒酒器分三个阶段制作。首先，将醒酒器主体（容器）内的熔融玻璃吹胀并塑形。接着制作底座，然后制作手柄。之后，专家在整个任务过程中不断滚动并塑造醒酒器，以防止玻璃因重力而变形。最后，专家塑造醒酒器主体时，需要一名助手向玻璃内吹气。

为了塑造熔融玻璃，玻璃吹制工不断用左手转动吹管，同时用右手塑造玻璃。他用右手操作各种工具，包括木坯（图 5b）、夹钳（图 5c）、soffietta工具、剪钳和金属笔。这些工具用于将玻璃塑造成醒酒器的形状，并添加更多装饰细节。木坯用于保持玻璃的圆形。夹钳用于搭配。用来塑造醒酒器的颈部。剪钳用于切割玻璃并形成醒酒器的尖端。soffietta工具用于塑造醒酒器的顶部。随后用金属笔添加手柄和额外的玻璃细节（颈部环绕装饰线），并制作醒酒器的底座。操作工具需要持续活动右肩、右臂和右前臂。同时，玻璃吹制工坐着，用左手在金属结构上来回转动吹管。在金属结构上移动吹管时，需要稍微弯腰，以保持对吹管的握持。放置手柄或用夹钳塑造颈部有时需要玻璃吹制工站起来，但他仍继续用左手移动吹管。

成型过程中，工匠频繁地将吹管上的玻璃放入熔炉中（图 5d）。他还一边持续通过吹管吹气，一边在肩高处用双臂水平握住吹管，以保持醒酒器的圆形（图 5b）。完成后，将其转移到托杆上冷却。

c：乳香树脂采集

乳香树脂的种植过程在希腊希俄斯历时三天进行记录。第一天和第二天的录像在室外、乳香树前拍摄。最后一天的录像则在室内模拟拍摄。由于种植过程的特点，每项任务都分开录制，因此每个过程环节都有单独的动作捕捉文件。总体而言，乳香树脂的种植过程是以写实方式记录的。然而，现实中有些任务相隔数天或数周才进行，或需要数小时才能完成。因此，要求专家在保持写实的同时，简要演示这些手势。从这一种植过程中记录的手势标记为 MSC。图 6 展示了从乳香树种植者身上捕捉到的一些手势。整个过程始于清理树木下方的土壤。为了便于收集滴落的乳香树脂，需要清理乳香树周围的泥土，并将树干周围的地面整平。记录了种植者使用两种不同工具刮土。



(a) GLB_3 (b) GLB_4 (c) GLB_8 (d) GLB_9

图 5. 在玻璃吹制工坊中捕捉到的手势示例。



(a) MSC_3 (b) MSC_4 (c) MSC_5 (d) MSC_8

图 6. 乳香树脂种植过程中捕捉到的手势示例。

第一件是一种古老的农具（Amia），带有金属头和木制手柄，类似于泥刀。农民用这件工具时跪在地上，用右手握住工具刮土。第二件工具是一把铲子，农民可以站着用它刮土。随后，农民用一把短扫帚清扫地面（图 6a ）。准备好土壤后，农民将碳酸钙（ $CaCO_3$ ）均匀撒在地面上，以形成平整的表面。执行这项任务时，农民跪在地上，右手撒白色粉末，左手拿着容器（图 6b）。

随后，为了获取乳香树脂，需要对树进行切割。树上有三种不同的刻划工具。第一种是末端带有尖锐凸起的小工具（Kenditiri），第二种是另一种名为 Timitiri 的小工具，第三种是一把小斧头。农夫使用每种工具时都处于站立姿势，但需要俯身在树上刻划。他用右手握持工具。接下来记录的步骤是采集和收获从树伤口处渗出的乳香树脂。农夫用小篮子和镊子捡起掉落的乳香树脂（图 6d），然后用剃刀从树上刮取更多树脂（图 6c）。这两种手势都需要农夫弯腰，并用右手操控工具。

农民用刷子将土壤刷到金属网上收集起来。为了去除乳香树脂上的灰尘，他们不断移动（或挪动）金属网。记录了两种金属网的使用方式。所有情况下，农民都跪着用双手移动金属网。最后，还记录了第三种去除乳香树脂灰尘的方法：将乳香树脂和灰尘抛向迎风站立。

III. 数据处理与分割

动作捕捉的处理分为两个步骤。首先应用低通滤波器，然后校正因电磁干扰或传感器漂移而导致的错误姿势；当录制时间过长时，还需要进行校准。对原始动作捕捉数据应用低通巴特沃斯滤波器，以消除高频噪声。为避免数据过度平滑，根据信号的功率谱密度选择截止频率。

动作捕捉系统的传感器在录制过程中可能会发生漂移，或受到周围金属物体产生的磁干扰影响。因此，在原本精准的动作捕捉过程中，偶尔会记录到错误的关节角度。我们使用 3D角色动画软件[2]对录制内容进行了调整，以纠正这一错误。该软件根据常识和视频反馈调整了不自然的动作。在调整动作捕捉数据并去除噪声后，我们按手势对其进行了分段。首先，录制内容按任务采集，每段录制内容代表一个完整任务；不过，之后我们按手势对这些录制内容进行了分段。图 7 展示了电视组装任务的分段示例，其中提取了手势 TVA_1，TVA_2，and TVA_3。ERGD 的七个数据集中的所有任务重复动作，均按手势或姿势进行了分段。一个任务可能包含一个手势，该手势会被执行

[2]MotionBuilder，Autodesk公司，圣拉斐尔，加利福尼亚州，美国

多次，或者其中可能包含在整个任务过程中反复出现的其他手势。

电视组装和包装的分割基于第 II-C1 节所述手势的重复次数。从录制内容中分割出的重复动作如表 3 所示。对于铆接任务，第一个手势包括铆接并完成整排铆钉。第二个手势是为下一排铆钉架设气动锤。最后一个手势则是为整排铆钉放置铆钉顶铁。表 4 展示了最终的分割结果。不同人体工学风险等级姿势的录制内容被分割为重复动作。假设录制了十名受试者，每人以三次重复的方式完成 28 种姿势，分割后共生成 840 个文件，每个文件包含一种姿势的一次重复动作。

传统手工艺任务的录制内容按单个手势进行分割（因为存在重复动作）。分割结果如表 5、表 6 和表 7 所示。

仅为便于训练后续章节所述的模型，我们手动减少了部分动作捕捉文件中欧拉关节角的不连续性。这些不连续性是仅出现在某些局部关节角度中的值 $180°$ 与 $-180°$ 之间的剧烈跳变。通过检查每个动作捕捉文件，我们确定将存在不连续性的时间序列转换为范围为 $[-250°, 250°]$ 的数据。请注意，对于使用惯性测量单元记录的新动作，这种转换可能并不适用。尽管如此，对于消除本文所介绍数据集中的大多数不连续性而言，这种方法已足够。我们对每项转换都进行了记录，以便将转换后的数据还原为欧拉角。以下是这些转换的一个示例：

表 3。 电视组装任务的分割。

Task	手势	重复次数
电视组装	TVA_1	107
	TVA_2	107
	TVA_3	108
	TVA_4	157
包装	TVP_1	8
	TVP_2	2
	TVP_3	7
	TVP_4	5
	TVP_5	12
	TVP_6	7
	TVP_7	7
	TVP_8	4
	TVP_9	2

表 4. 铆接任务的分割。

Task	手势	重复次数
铆接	APA_1	6
	APA_2	5
	APA_3	8

表 5. 丝绸织造任务的分割。

Task	手势	重复次数
制作卡片	SLW_1	110
织轴准备	$SLW_{2,1}$	3
	$SLW_{2,2}$	2
	$SLW_{2,3}$	4
	$SLW_{2,4}$	1
	$SLW_{2,5}$	1
卷绕经轴	SLW_3	2
使用小型织机织造	$SLW_{4,1,1}$	11
	$SLW_{4,1,2}$	11
	$SLW_{4,1,3}$	11
使用中型织机	$SLW_{4,2,1}$	35
	$SLW_{4,2,2}$	35
	$SLW_{4,2,3}$	35
织造大型织机	$SLW_{4,3,1}$	16
	$SLW_{4,3,2}$	16
	$SLW_{4,3,3}$	15

表 6. 玻璃吹制任务的分割。

Task	手势	重复次数
鸟喙修剪	GLB_1	11
	GLB_2	6
吹制与成型	GLB_3	5
	GLB_4	8
	GLB_5	15
	GLB_6	7
	GLB_7	35
瓶颈精整	GLB_8	6
帘线铺设	GLB_9	2
	GLB_{10}	8
	GLB_{11}	4
收尾细节	GLB_{12}	5
手柄铺设	GLB_{13}	4
	GLB_{14}	5
	GLB_{15}	4
转移至托杆	GLB_{16}	4
腿部和底足铺设	GLB_{17}	6
	GLB_{18}	7

表 7。 乳香栽培任务的分割。

Task	手势	重复次数
刮削（新工具）	MSC_1	3
刮削（旧工具）	MSC_2	9
扫拂	MSC_3	9
除尘	MSC_4	9
刺绣 A	MSC_5	9
刺绣 B	MSC_6	3
用斧头刺绣	MSC_7	3
采集	MSC_8	8
收获	MSC_9	7
擦拭	MSC_{10}	6
移动 A	MSC_{11}	6
转移 B	MSC_{12}	3
随风清洁	MSC_{13}	3

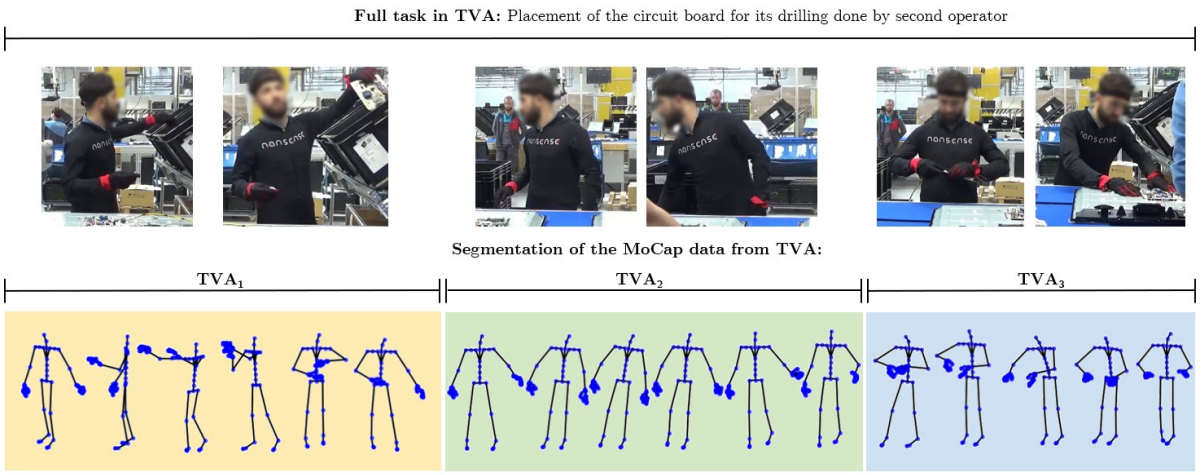


图 7。 电视组装任务中一次重复动作的手势分割。

mations在图 8 中表示。该图展示了修改前后的动作捕捉数据，以及重建骨架。

上臂、前臂以及髋部的一个角度主要是带有不连续性的局部角度。髋部在Y轴上的角度（指向上方，用于测量躯干旋转）问题最为严重，也最容易发生漂移。这可能与传感器的位置有关。如果动作捕捉服过于宽松，传感器可能会产生不准确读数。另一个因素是，在动作捕捉服打开并连接到计算机进行记录后，受试者必须移动整个身体来“唤醒”传感器。在执行校准时，该传感器很可能仍处于空闲状态。因漂移或校准不良而产生失真的任何动作捕捉文件都已从数据集中移除。以下章节中使用的七个数据集总大小为5GB。总计163,4776帧，即5小时2分钟，构成了分割手势，其中测量了156个局部关节角度。七个数据集的所有BVH文件均可在Zenodo[3]中访问。

IV. 使用解析模型分析数据集

身体节段的任何自主运动都是通过肌肉骨骼系统完成的。肌肉骨骼系统是一个复杂的结构，由骨骼、肌肉、韧带和肌腱组成。因此，对如此复杂的结构进行建模并非易事。然而，尽管肌肉骨骼系统是人体运动复杂性的主要来源，但使用解析模型来表示人体动作仍是可以接受的，只要这些模型纳入了有关身体关节关联及其时间依赖关系的相关假设。人体动作

随后，使用基于手势操作模型的解析模型对这七个数据集中的内容进行分析。GOM 可根据其假设的学习得到的参数量化人体灵活性。本文分析所用的代码可在GitHub[4] 上获取。

GOM使用一组数学方程来表示人体动作，这些方程纳入了关于人体运动随机性以及身体关节相互调节作用的假设。这些假设能够利用训练好的模型恰当地模拟人体动作，并解释人体运动描述符随时间的演变，从而支持对这些信息的主动应用。例如，在以人为中心的人工智能技术中，人的物理具身性是核心关注点（人机协作、风险监测或灵巧度分析）。理解并捕捉不同关节运动之间的依赖关系，不仅对于创建更真实的人体动作模拟至关重要，也有助于研究多样而复杂的全身人体动作是如何完成的。通过这些模型，可以获得对复杂灵巧度和运动学习背后神经生理机制的认识。最终，使用此类分析模型可能有助于建立跨学科框架，以研究在工业或手工业领域执行专业任务时的学习过程和技能习得。此外，这些模型还可能促进人体工学领域对导致肌肉骨骼疾病的关键因素的研究。

A. 手势操作模型

GOM 是一种用于表示全身人体运动的数学方法，会考虑身体关节的时空动态特性。该数学表示由一组模型组成，每个模型

[3]基准网站：https://doi.org/10.5281/zenodo.5356992

[4]代码库：https://github.com/olivas-bre/GOM.git

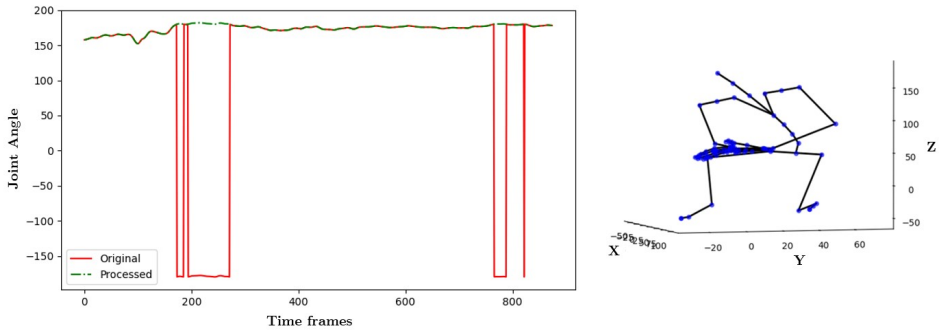


图 8. 消除数据不连续性，以便进行后续分析。记录的动作对应于姿势 ERGD 的左小腿绕 X 轴的欧拉角（如图 9 所示为 ***LL***）[25]。

都通过结合卡尔曼滤波器的单次训练，对一种独特的关节运动描述符进行建模 [15]。GOM 方程系统中的模型数量等于 GOM 所定义的身体关节数量，乘以每个关节运动描述符（例如角度或位置）所分解的维度数（例如 X、Y 和 Z）。在本研究中，建模时仅使用了 19 个惯性测量单元的运动描述符（数据集中共有 52 个可用单元）来训练 GOM。为简化人体运动表示，舍弃了手指和脚部的动作捕捉数据。图 9 展示了传感器的放置位置、标记和方向。为生成具有不同体型受试者的姿势，人体姿态以三维欧拉关节角表示。与关节位置不同，欧拉关节角不受个体特有体型的影响。此外，欧拉角可在解析模型中直观地解读，并能更清晰地展示人体动作的执行方式。

因此，本文使用的 GOM 由 57 个模型组成，用于分析每个数据集中的全身运动。这 57 个模型通过状态空间建模创建，其中每个运动描述符的二阶模型都包含内生数据和外生数据。

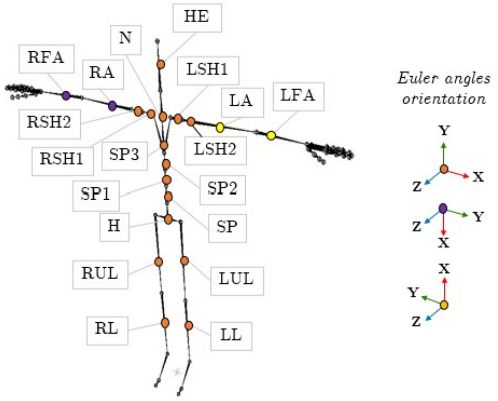


图 9. 提供 ***XYZ*** 关节角度的传感器在 GOM 中的位置和欧拉角方向。

例如，在对身体关节 P_t 在 X 轴上的角度轨迹（Px_t）进行建模时，其运动被分解到 XYZ 轴（Px_t、Py_t 和 Pz_t）上，并与 j 个身体部位相关联。如式 1 所示，前两个值被整合到转移模型中，其中 s_t 对应状态变量。然后，如式 2 所示，将对应于潜在关节内关联（H2）、肢体间协同（H3）或肢体内中介作用（H4）的外生数据（u_t）纳入观测模型。

$$s_t = As_{t-1} = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} \begin{bmatrix} Px_{1,t-1} \\ -Px_{1,t-2} \end{bmatrix} \tag{1}$$

$$\begin{aligned} Px_{1,t} = \begin{bmatrix} 1 & 1 \end{bmatrix} s_t + Bu_t = \\ \begin{bmatrix} 1 & 1 \end{bmatrix} s_t + \beta_1 Py_{1,t-1} + \beta_2 Pz_{1,t-1} + \\ \beta_3 Px_{2,t-1} + \cdots + \beta_n Px_{j,t-1} \end{aligned} \tag{2}$$

最后，通过合并方程 1 和 2，得到该运动描述符的状态空间表示：

$$\begin{aligned} Px_{1,t} = \underbrace{\alpha_1 Px_{1,t-1} - \alpha_2 Px_{1,t-2}}_{\text{H1}} + \underbrace{\beta_1 Py_{1,t-1} + \beta_2 Pz_{1,t-1}}_{\text{H2}} + \\ \underbrace{\beta_3 Px_{2,t-1} + \cdots + \beta_n Px_{j,t-1}}_{\text{H3 or H4}} \end{aligned} \tag{3}$$

[13]中进一步详细介绍了模型的假设和结构。方程组的常数系数 A 和 B 通过卡尔曼滤波使用最大似然估计（MLE）进行估计。GOM 使用每个类别的一个参考手势进行训练，该参考手势通过动态时间规整（DTW）算法 [16]确定。该算法用于测量两个时间序列之间的相似度。随后，选择与同类其他所有手势重复动作最接近的手势重复动作，使用 Intel Core i7-8750H CPU 进行单次训练。

接下来，IV-B 节讨论了训练后的 GOM 在七个数据集的每种手势上的仿真性能。附录中提供了指标和生成姿势的示例。这些指标旨在作为数据集的初步基准，用于

比较其他使用所提供数据集的方法的仿真或生成性能。之后，IV-C 节介绍了使用训练后的 GOM 表示进行专业手势的灵巧度分析，并说明如何基于这些模型识别最重要的运动描述符，以便对专业任务中的手势进行建模和识别。

B. 全身运动生成

本节介绍了 GOM 生成专业人体姿势的结果。训练后的 GOM 可以通过求解其方程组来生成专业人体姿势，其中每个 GOM 模型在每次迭代中预测一个时间步。

为了衡量模型模拟所学习的专业手势的能力，使用各自训练好的 GOM 对所有手势重复动作进行了模拟。随后，计算每次模拟的均方根误差（RMSE）和平均绝对误差（MAE）：

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{T}}\sum_{\mathrm{t}=1}^{\mathrm{T}}\left(\mathrm{P_t} - \hat{\mathrm{P}}_\mathrm{t}\right)^2} \tag{4}$$

$$\mathrm{MAE} = \frac{1}{\mathrm{T}}\sum_{\mathrm{t}=1}^{\mathrm{T}}\left|\mathrm{P_t} - \hat{\mathrm{P}}_\mathrm{t}\right| \tag{5}$$

ˆ 真实的全身姿态对应于 $\mathrm{P_t}$，$\mathrm{P_t}$ 是使用训练后的 GOM 模拟的动作。指标中还包括 Theil 不平等系数（$\mathrm{U_1}$）的平均值，该系数的计算方法如下：

$$\mathrm{U_1} = \frac{\sqrt{\frac{1}{\mathrm{T}}\sum_{\mathrm{t}=1}^{\mathrm{T}}\left(\mathrm{P_t} - \hat{\mathrm{P}}_\mathrm{t}\right)^2}}{\sqrt{\frac{1}{\mathrm{T}}\sum_{\mathrm{t}=1}^{\mathrm{T}}\mathrm{P_t^2}} + \sqrt{\frac{1}{\mathrm{T}}\sum_{\mathrm{t}=1}^{\mathrm{T}}\hat{\mathrm{P}}_\mathrm{t}^2}} \tag{6}$$

对于 $\mathrm{U_1}$,，其越接近零，预测质量越高。附录中的表 9 至表 13 展示了每项任务的手势的平均指标。此外，图 13 至图 15 展示了某个手势生成的姿势以及真实姿势序列的示例。

1）模型仿真讨论

结果表明，通过求解构成 GOM 的联立方程，可以将欧拉关节角作为运动描述符，生成各种人体姿态。GOM 能够容忍人体运动中的细微变化，以及由记录条件不同（受试者不同或记录日期不同）导致的同类运动之间的偏差。然而，如果从准确预测全身运动的能力来评估其性能，那么由于某些专业手势存在类内差异，关节角度的均值会增加。

错误。原因可能是用于模型单次训练的参考手势与用于模拟测试的手势之间存在差异。

对于 TVA数据集而言，最难以准确模拟的手势是 $\mathrm{TVA_1}$ 和$\mathrm{TVA_2}$„，它们对应的是操作员可以更自由地使用左手或右手抓取电路板或电缆的手势。另一方面，$\mathrm{TVA_3}$ 和 $\mathrm{TVA_4}$ 对应的是操作员在每次迭代中都更容易复现的手势，因为在每次记录的迭代中，电路板操作和钻孔的执行方式都相似。

GOM为APA、TVP 和 ERGD 提供了最佳的仿真性能。这些手势和姿势的类内变异性最低，因为它们是在更受控的环境中执行的。例如，在 ERGD中，受试者在实验室里执行各种姿势，同时不断接受有关如何执行这些姿势的指导。然而，随着姿势变得更加复杂，例如在跪姿下进行躯干和上臂运动，误差也随之增加，因为受试者在执行指定姿势时的动作表现出更大的差异（底座和膝盖的位置，或上臂的最终位置）。在 APA中，操作人员被记录下来组装一个飞机浮筒，连续数小时反复执行相同的任务；唯一较大的差异出现在 $\mathrm{APA_2}$ 中，即操作人员抓取铆钉并准备气动锤时。在TVP中，操作人员只有在包装电视机时，所执行的相同手势才表现出较大的差异。

在工业环境中记录的手势更容易模拟，因为这些手势主要涉及用手操纵物体；相比之下，例如工匠和农民在工作时必须调动全身，才能正确完成任务。

参考手势和模拟手势是在不同的织机上完成的，这可能导致了 SLW 中的误差（参考动作是在大型织机上完成的，而模拟动作是在中型织机上完成的）。因此，踏板高度和位置的变化可能导致动作模拟出现更大的误差。同样，在 GLB 的动作模拟中，即使是重复进行同一项活动，技艺娴熟的玻璃吹制工也会逐步调整姿势，以便恰当地塑造熔融玻璃。因此，GLB 数据集中每个类别的训练手势未能充分代表该类别中的所有手势（类内方差较高），导致模拟准确率下降。

最难模拟的手势涉及乳香树脂采集。这可能是因为 MSC 中包含农民跪着移动时所做的手势。在其他六个数据集中，受试者完成大多数任务时都处于站立状态。农民在做相同手势时，腿部位置并不固定；他重新调整了自己……

跪姿时调整双腿以改善平衡，从而够到树木或物体。

C. 基于GOM的专家动作灵巧性分析

对学习得到的GOM 表征进行统计分析，以确定模型假设相对于专业手势的重要性。随后，利用显著的假设（运动描述符）及其学习得到的系数，描述各关节为完成该手势所进行的协同运动。此外，通过分析每个假设的 p值，找出用于对专业任务中的人体动作进行建模和识别的最重要运动描述符。在许多人体运动分析应用中，使用全身动作捕捉服既不可行也不切实际。因此，为了支持采用智能手机和智能手表等侵入性较低的技术，本文还详细介绍了一种使用 GOM 寻找需要测量的最小运动描述符集合的方法。

1）模型的统计分析与解读

接下来将介绍三种经过训练的动作表征的统计分析。为便于直观呈现所建模的手势，每个示例都配有姿势序列图，并以颜色标注突出显示方程的假设。GOM 的表征旨在涵盖四项假设：时间依赖转移、关节内关联、肢体间协同，以及串行和非串行肢体内中介。每项假设都由一组特定的参数化变量（在此为关节角度）构成，用以描述身体关节之间的特定关系或时间依赖性。其核心思路是利用这些参数化假设来描述身体灵巧性。通过考察计算得到的系数及其统计显著性（p 值小于 0.05 即为显著），可以了解这些假设与所建模的手势及预测的关节角度之间的相关性。

第一个示例展示了执行手势 TVA_1 时的关节角序列方程 $\mathrm{RAy_t}$ （右臂绕 Y轴运动）：（从容器中拿取一块电路板，如图10所示）：

$$\begin{aligned}\mathrm{RAy_t} = &\underbrace{(1.010)\mathrm{RAy_{t-1}}}_{p=0.001} + \underbrace{(-0.076)\mathrm{RAy_{t-2}}}_{p=0.188} \\ &+ \underbrace{(0.720)\mathrm{RAx_{t-1}}}_{p=0.003} + \underbrace{(1.214)\mathrm{RAz_{t-1}}}_{p<0.001} \\ &+ \underbrace{(-0.324)\mathrm{LAy_{t-1}}}_{p<0.001} + \underbrace{(6.123)\mathrm{RSH1y_{t-1}}}_{p<0.001} \\ &+ \cdots + \underbrace{(0.555)\mathrm{RFAy_{t-1}}}_{p=0.009} \end{aligned} \quad (7)$$

p值 < 0.05 表明，因变量的先前值与当前值之间存在依赖关系，但提前两个时间步的值与当前值之间不存在依赖关系。这可能意味着手势的变化速度适中。如果两个先前值都具有显著性，则表明动作速度较慢；如果两者都不显著，则动作速度较快。关节 RA 的运动在 X、Y 和 Z 轴上表现出关节内关联。与 LAy（左臂）的肢体间协同表明，执行手势时，LAy 会与 RAy 协同运动。RSH1y（右肩）和 RFAy（右前臂）的运动产生了肢体内串行中介。这一结果合乎情理，因为此类手臂运动主要取决于肩部动作（抬臂）。此外，参见图1a，操作员必须抬起肩膀并弯曲前臂，才能从容器中取出电路板。前臂弯曲可能解释了 RFAy 的统计显著性。

第二个示例是在执行 APA_3（握住铆钉顶铁，如图 11 所示）时，颈部绕 X 轴的关节角度（$\mathrm{Nx_t}$）的计算公式：

$$\begin{aligned}\mathrm{Nx_t} = &\underbrace{(1.020)\,\mathrm{Nx_{t-1}}}_{p<0.001} + \underbrace{(0.355)\,\mathrm{Nx_{t-2}}}_{p<0.024} + \\ &\underbrace{(-1.220)\,\mathrm{Ny_{t-1}}}_{p<0.001} + \underbrace{(-0.470)\,\mathrm{Nz_{t-1}}}_{p<0.001} \\ &+ \underbrace{(-0.018)\,\mathrm{SP3x_{t-1}}}_{p<0.001} + \underbrace{(-0.010)\,\mathrm{SP2x_{t-1}}}_{p=0.002} \\ &+ \cdots + \underbrace{(0.010)\,\mathrm{Hx_{t-1}}}_{p=0.84} \end{aligned} \quad (8)$$

在（8）中揭示了与Ny和 Nz 的关节内关联，以及与SP3（上部脊柱）的串行肢体内中介作用。SP2（中部脊柱）表现出非串行的肢体内中介作用，但H（髋部）没有。手持铆钉顶铁以抵住铆钉，需要向前弯曲并使上部躯干稍微扭转（如图11和图2c 所示），沿脊柱的 X轴和 Y轴移动。这一运动反映在（8）中，因为 SP2 和 SP3 在 X轴和 Y轴上的关节角度在统计上显著，并且与 Nx的运动相关。然而，与 H缺乏中介作用可能表明，操作员试图保持髋部静止，这很可能是为了在弯曲时保持平衡。此外，受试者必须转动颈部，以观察应将铆钉顶铁放置在哪里；因此，这与 Ny和 Nz的 p值所表明的关节内关联一致。最后，由于两个转换假设均具有显著性，该手势以较慢的速度完成。

最后一个示例是通过手势 GLB_4 （如图 12所示，用木坯塑造醒酒器曲线）学习得到的方程，表示左肩（$\mathrm{LSH2x_t}$）在 X 轴上的关节角度。更准确地说，



图10。 在 TVA 中执行的手势示意图1，其中操作员从容器中抓取一块电路板。颜色标注基于 (7) 中着色的假设。着色的关节表示模型中纳入的运动描述符之间可能存在的依赖关系：橙色表示 ***RAy*** 的转换假设；绿色表示与 ***RAx*** 和 ***RAz*** 的关节内关联；蓝色突出显示与***LAy*** 的肢体间协同；紫色表示与 ***RSH2y*** 的肢体内串行中介，红色则表示与 ***RFAy*** 的非串行肢体内调节。录制画面也可见图1a。

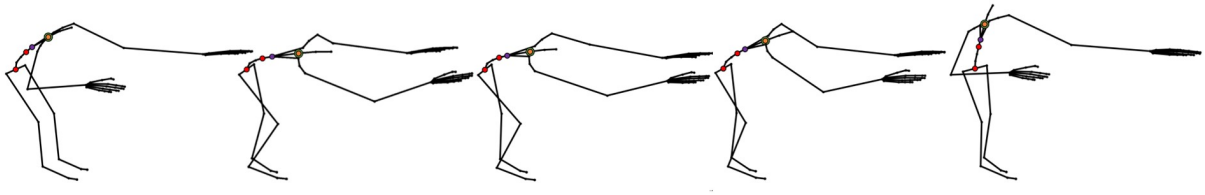


图11。APA 中所执行手势的示意图3，其中，操作员放置铆钉顶铁，以抵消 incoming 铆钉的冲击。颜色标注基于 (8)中着色的假设。着色的关节表示与模型中纳入的其他运动描述符之间的潜在依赖关系：橙色表示 ***Nx*** 的过渡假设；绿色反映与 ***Ny*** 和 ***Nz*** 的关节内关联；紫色表示与 ***SP3x*** 的肢体内串行中介，红色表示与 ***SP2x*** 和 ***Hx*** 的非串行肢体内中介。录制的画面也可以在图 2c 中查看。

该方程模拟左锁骨的运动：

$$\begin{aligned}\mathrm{LSH2x_t} = &\underbrace{(1.877)\mathrm{LSH2x_{t-1}}}_{p<0.001} + \underbrace{(-0.913)\mathrm{LSH2x_{t-2}}}_{p<0.001} \\ &+ \underbrace{(0.292)\mathrm{LSH2y_{t-1}}}_{p=0.002} + \underbrace{(0.252)\mathrm{LSH2z_{t-1}}}_{p=0.004} \\ &+ \underbrace{(0.145)\mathrm{RSH2x_{t-1}}}_{p=0.014} + \underbrace{(0.36)\mathrm{LAx_{t-1}}}_{0.004} \\ &+ \cdots + \underbrace{(0.016)\mathrm{LFAx_{t-1}}}_{p=0.030} + \underbrace{(-0.543)\ \mathrm{SP3x_{t-1}}}_{p=0.049} \quad (9)\end{aligned}$$

对（9）的统计分析揭示了时间依赖性（缓慢运动）；关节内关联（LSH2y 和 LSH2z）；与右肩的肢体间协同；与左臂（LAx）的肢体内串行中介，以及与左前臂（LFAx）的非串行中介。SP3 被认为具有边缘显著性，因为本研究采用 0.05 作为确定显著性的 p值阈值。

为了正确塑造醒酒器，在执行这一手势时，双臂必须协同工作。这一点可由（9）中存在肢体间协同看出。因此，右肩的关节角度会影响左肩的反应，因为玻璃吹制工用右臂塑造醒酒器的曲线，同时用左手转动吹管。此外，如图 13 和图 5a 所示，专家在执行这一手势时大多保持躯干挺直。然而，当他向前旋转吹管时，躯干会略微倾斜，以保持对吹管的握持；这可能表明 SP3 的 p 值较高但还没有高到使左肩运动达到不显著的程度。

如前述示例所示，GOM 可以提供无法直接观察到的量化信息，说明专家如何执行所建模的手势，并帮助解读各关节如何协同运动，通过特定的关节运动轨迹完成预期任务。对这些假设进行显著性分析，突显了手势中起关键作用的关节及其对其他关节运动的影响。这些信息之后可用于检验技能习得策略。例如，新手可以通过尽量减少自身运动表征与专业工匠或操作人员的运动表征之间的差异，学会做出精准的手势。此外，在人体工学方面，所提出的运动表征能够帮助分析人员理解进行具有人体工学危害的动作与安全动作时全身如何运动，并据此设计不易导致受伤或不适的工作环境和任务。一个研究方向是确定执行高风险动作时影响最大的关节，并对其进行监测，以降低专业任务中的人体工学风险。

2）为每个数据集选择最重要的运动描述符

为了选择用于每项专业任务手势识别的必要惯性传感器，运动描述符（假设）的统计次数为

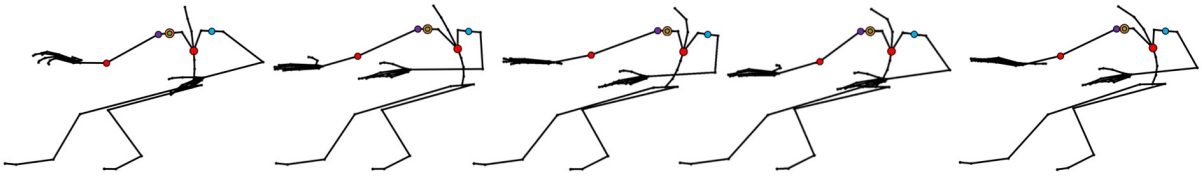


图 12. 在 GLB 中执行的手势示意图$_4$，专家玻璃吹制工使用木坯塑造醒酒器的曲线，同时前后旋转吹管。颜色标注基于 (9) 中标出的假设。着色的关节表示模型中纳入的运动描述符之间可能存在的依赖关系：橙色表示 ***LSH2x*** 的过渡假设；绿色表示与 ***LSH2y*** 和 ***LSH2z*** 的关节内关联；蓝色突出显示与 ***RSH2x*** 的肢体间协同；紫色表示与 ***LAx*** 的肢体内串行中介，红色表示与 ***LFAx*** 和 ***SP3x*** 的肢体内非串行中介。录制画面见图 5a。

对构成GOM的所有方程具有统计显著性的内容都会被计数。随后，采用在测量手臂、脊柱和腿部时最常被认为具有显著性的不同描述符组合，以全镜头方法进行训练。由于单个惯性传感器可提供三个关节角度，因此，如果传感器的关节角度中至少有一个属于数据集中所有手势中更常具有显著性的关节角度，则识别时会使用该传感器的全部关节角度。

为了利用不同的传感器组合识别人类手势，我们使用 10 折交叉验证训练了隐马尔可夫模型（HMM）。为了正确训练 HMM，我们为每个数据集指定了一个手势词汇表，其中包含重复次数最多的手势。TVA、APA 和 ERGD 的手势类别总数分别为 4、3 和 28。TVP、GLB 和 MSC 的手势词汇表仅包含至少重复 7 次的手势。因此，它们各自的手势词汇表分别包含 5、7 和 6 类手势。对于 SLW，手势词汇表仅由 3 类织机丝绸织造手势组成。尽管织机尺寸不同，但在小型、中型和大型织机上进行织造的手势相似。因此，在手势识别问题中，这些手势被合并为 3 类。

对每个数据集中定义的手势词汇，评估了遍历型和从左到右型 HMM 拓扑，以及不同数量的隐状态，以确定最佳设置。采用的性能指标为准确率和 F1 分数，后者是精确率与召回率的调和平均数。

从左到右型 HMM 拓扑在所有识别任务中都取得了最佳结果。隐状态数量方面，TVA 和 ERGD 的 HMM 设为 7 个状态，TVP 设为 6 个状态，APA、GLB 和 MSW 设为 8 个状态，SLW 设为 3 个状态。表 8 列出了各数据集中手势识别效果最佳的传感器配置，以及所达到的最高识别准确率和 F1 分数。

在表 8 中可以观察到，仅使用来自所选传感器的动作捕捉数据所取得的性能，that 其性能与使用 52 个惯性传感器采集的全部动作捕捉数据相当，或更好。对于 TVA 和 TVP，根据 GOM 训练的模型仅选择了三个传感器，占采集的动作捕捉数据的 5.76%，且仅造成准确率和 F1 分数轻微下降。对于 APA 和 ERGD，分别仅选择并使用了 13.46% 和 9.61 % 的动作捕捉数据，其识别性能高于另外两种传感器配置。

APA 手势最难识别，需要使用更多传感器才能有效识别。这可能是因为该词汇中的手势更复杂、持续时间更长。最难建模和识别的手势是 APA_2, , 这符合预期，因为在三个类别中，该手势的执行方式变化最大（类内方差高）。操作员每次重复动作时准备材料的方式并不完全相同。有些重复动作中，操作员比平时更慢，因为他需要更多时间调整气动锤或准备额外的铆钉。此外，由于该数据集只建造了一架飞机结构，因此类内方差较大。没有任何一次重复动作中，气动锤被放置在同一位置超过一次。

关于识别传统手工艺中的手势，对于 GLB 和 MSC，所选传感器仅使用了 7.69% 的动作捕捉数据，但取得了与使用所有传感器数据相当的结果。对于 SLW，在识别任务中选用了 9.61% 的动作捕捉数据；与使用全部 52 个传感器的配置相比，两项指标的表现均下降了约 0.05。双传感器配置在 SLW 上表现不佳，可能是因为难以区分与肩部（投梭）和腿部相关的动作，仅依靠髋部和左前臂的动作数据并不足够。

V. 结论

本文介绍了七个数据集：TVA、TVP、APA、ERGD、SLW、GLB 和 MSC。大多数公开可用的数据集包含模拟动作，这些动作由

表 8. 各传感器配置下的识别性能。

手势词汇	类别数	传感器	准确率（%）	F1-分数（%）
TVA	4	全部 52 个传感器	0.967	0.966
		两个传感器：RFA ，H LA, SP1, RUL	0.891 0.910	0.871 0.902
TVP	5	全部 52 个传感器	0.950	0.916
		两个传感器：RFA ，H RSH1, LFA, SP2	0.888 0.901	0.850 0.843
APA	3	全部 52 个传感器	0.850	0.833
		两个传感器：RFA ，H RA, LSH1, LSH2,SP3, SP2,LUL,RUL	0.750 0.915	0.701 0.901
ERGD	28	全部 52 个传感器	0.916	0.902
		两个传感器：RFA ，H LA, RSH1, LFA,SP2, RUL	0.738 0.927	0.735 0.917
SLW	3	全部 52 个传感器	0.954	0.943
		两个传感器：RFA ，H RSH1, LSH1, HE,LUL, RL	0.620 0.909	0.610 0.892
GLB	7	全部 52 个传感器	0.917	0.816
		两个传感器：RFA ，H LSH2, RFA, H, SP3	0.810 0.842	0.801 0.850
MSC	6	全部 52 个传感器	0.866	0.866
		两个传感器：RFA ，H LSH1, SP3, LUL, LL	0.799 0.866	0.750 0.866

在实验室中，并且与日常活动或体育运动相关。因此，创建了新的数据集，其中包含在工业或手工艺作坊的专业任务中执行的手势。这些手势由实际操作人员和专家在真实工作场景中使用配备 52 个传感器的惯性全身套装进行记录。其目的是使用这些复杂手势测试人体运动模型，并提取与灵巧度、技能以及正确使用材料和工具等有形要素相关的专门知识。每项专业任务均按重复次数进行分段，并减少了不连续性，以改进本基准分析中手势的建模。

所 प्रस्तुत的人体运动分析包括使用 GOM 对记录的专业任务进行模拟，以及基于训练得到的运动表征开展灵巧度分析。其目的是利用训练得到的运动模型，观察并量化工业操作人员和资深手工艺人所展现的技能。训练模型的参数提供了有关人员如何运动以实现特定目标的信息，例如组装一台电视机或制作一件特定的玻璃制品。未来，可以构建多学科框架，通过考察专家和初学者训练得到的解析模型，研究人们如何学习并提升完成工业或手工艺任务的能力。此外，还可以利用 GOM，通过比较安全动作与危险动作的运动表征，研究导致工作相关肌肉骨骼疾病的生物力学风险因素。

最后，介绍了捕捉并准确识别每项已记录专业任务中手势所需的最少惯性传感器数量及其位置，并进行了说明。如前所述，在许多人体运动分析应用中，使用全身动作捕捉服既不可行也不实用。确定需要测量的最少运动描述符，有助于采用智能手机和智能手表等侵入性较低的技术，这些设备也能够测量这些运动描述符。

致谢

我们谨向 Jean-Pierre Mateus、Pireaus基金会、Haus der Seidenkultur博物馆，以及 Romaero 和 Arçelik 工厂表示感谢，感谢他们为数据集的创建作出贡献。

附录 A 各数据集的预测性能指标

表 9. 数据集 TVA、TVP 和 APA 的仿真性能。

数据集	手势	RMSE	MAE	平均 U_1
TVA	TVA_1	52.637	25.610	0.508
	TVA_2	65.893	37.600	0.825
	TVA_3	6.836	2.742	0.116
	TVA_4	4.385	1.368	0.088
TVP	TVP_1	10.241	2.050	0.046
	TVP_2	15.746	4.779	0.078
	TVP_3	10.164	5.545	0.099
	TVP_4	9.657	2.267	0.060
	TVP_5	13.669	12.305	0.232
	TVP_6	18.192	15.816	0.202
	TVP_7	21.746	16.452	0.219
	TVP_8	16.898	4.132	0.082
	TVP_9	27.492	8.178	0.105
APA	APA_1	8.311	1.550	0.110
	APA_2	42.558	7.286	0.480
	APA_3	2.575	1.100	0.042

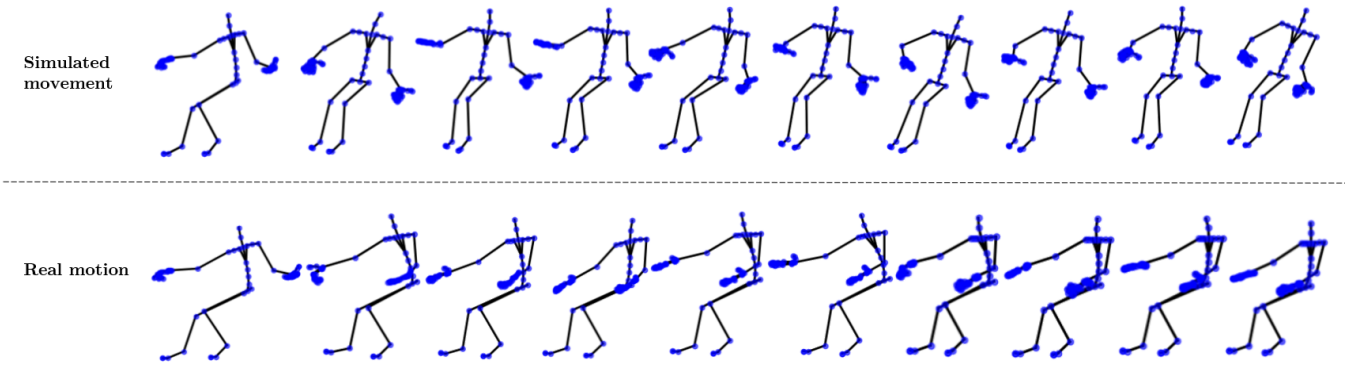


图 13. 生成的姿势序列与真实姿势序列的对比。GLB_4 的生成姿势序列及其真实值。玻璃吹制工用左手转动吹管，同时用右手塑造玻璃（玻璃吹制工的录制画面见图 5a）。

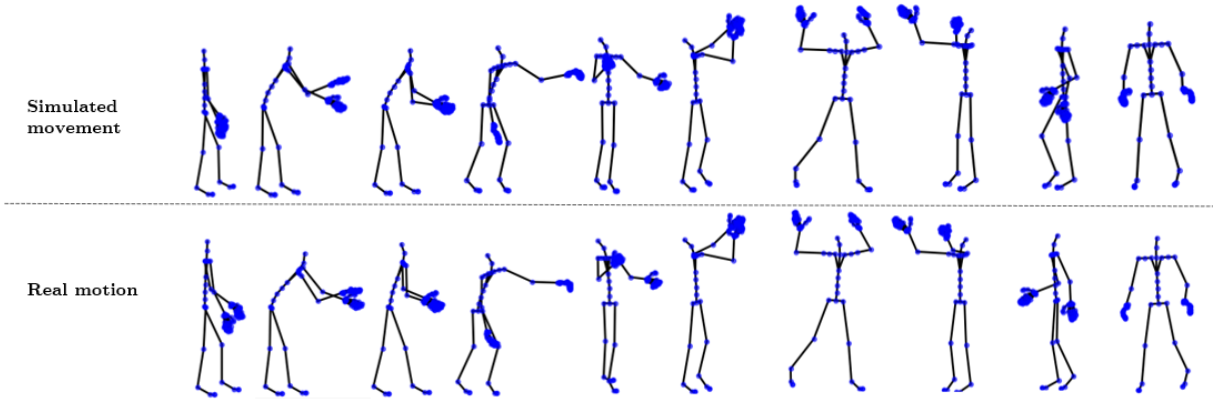


图 14. TVP 生成姿势序列与其真实值的对比$_8$。操作员将一台电视机放在托盘的第三层上（录制画面见图 1e）。

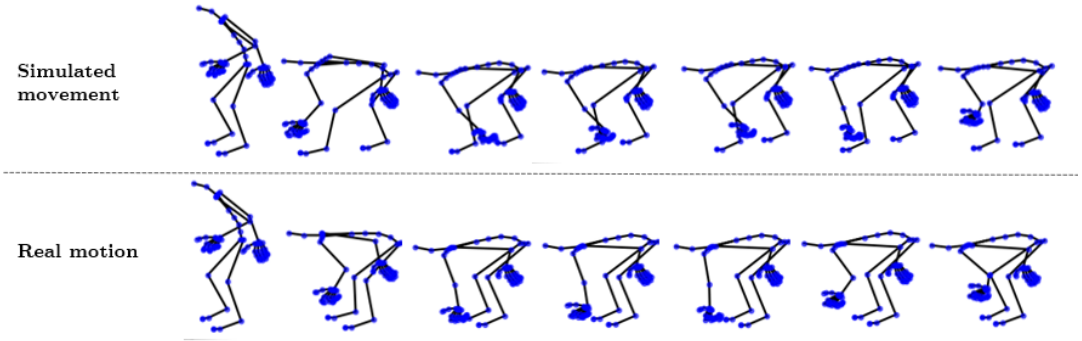


图 15. MSC 生成的姿势序列与$_5$ 真实姿势的视觉对比。乳香树种植者用一把小刀切割乳香树的根（录制画面见图 6c）。

表 10. ERGD 数据集的仿真性能。

数据集	手势	RMSE	MAE	平均 U_1
ERGD	$ERGD_1$	7.347	6.516	0.323
	$ERGD_2$	9.793	4.378	0.165
	$ERGD_3$	4.120	2.633	0.136
	$ERGD_4$	5.031	3.332	0.162
	$ERGD_5$	5.066	2.759	0.137
	$ERGD_6$	7.596	4.750	0.137
	$ERGD_7$	6.159	3.657	0.165
	$ERGD_8$	4.780	3.090	0.108
	$ERGD_9$	4.914	3.063	0.092
	$ERGD_{10}$	22.163	18.234	0.247
	$ERGD_{11}$	28.283	22.209	0.264
	$ERGD_{12}$	11.698	6.586	0.177
	$ERGD_{13}$	11.316	4.952	0.172
	$ERGD_{14}$	35.360	16.796	0.225
	$ERGD_{15}$	32.047	17.635	0.224
	$ERGD_{16}$	16.538	8.923	0.196
	$ERGD_{17}$	18.340	10.190	0.215
	$ERGD_{18}$	24.824	18.354	0.211
	$ERGD_{19}$	25.478	22.967	0.306
	$ERGD_{20}$	34.291	18.294	0.221
	$ERGD_{21}$	30.535	15.333	0.307
	$ERGD_{22}$	24.023	16.639	0.351
	$ERGD_{23}$	22.024	15.077	0.205
	$ERGD_{24}$	39.893	23.038	0.348
	$ERGD_{25}$	44.179	27.285	0.363
	$ERGD_{26}$	44.331	29.911	0.473
	$ERGD_{27}$	43.128	23.421	0.401
	$ERGD_{28}$	45.235	26.899	0.459

表11。 数据集SLW的仿真性能。

数据集	手势	RMSE	MAE	平均 U_1
SLW	SLW_1	16.518	8.841	0.063
	$SLW_{2,1}$	14.563	7.411	0.079
	$SLW_{2,2}$	14.441	7.104	0.008
	$SLW_{2,3}$	20.840	10.217	0.016
	$SLW_{2,4}$	16.551	3.097	0.005
	$SLW_{2,5}$	6.805	3.120	0.024
	SLW_3	21.907	10.666	0.082
	$SLW_{4,1,1}$	28.765	15.427	0.393
	$SLW_{4,1,2}$	27.136	13.489	0.180
	$SLW_{4,1,3}$	52.251	30.356	0.532
	$SLW_{4,2,1}$	31.695	15.635	0.234
	$SLW_{4,2,2}$	50.182	25.050	0.318
	$SLW_{4,2,3}$	47.670	18.471	0.337
	$SLW_{4,3,1}$	30.179	14.747	0.265
	$SLW_{4,3,2}$	38.361	18.740	0.293
	$SLW_{4,3,3}$	42.731	25.895	0.383

表12. 数据集GLB的仿真性能。

数据集	手势	RMSE	MAE	平均 U_1
GLB	GLB_1	68.644	36.340	0.421
	GLB_2	7.921	2.929	0.125
	GLB_3	43.928	21.410	0.294
	GLB_4	23.980	11.502	0.156
	GLB_5	49.547	24.251	0.314
	GLB_6	41.230	17.904	0.351
	GLB_7	10.292	5.213	0.146
	GLB_8	65.316	28.045	0.394
	GLB_9	62.949	21.055	0.297
	GLB_{10}	59.623	25.070	0.290
	GLB_{11}	51.223	32.240	0.534
	GLB_{12}	16.418	7.859	0.176
	GLB_{13}	38.091	18.714	0.160
	GLB_{14}	88.048	33.678	0.519
	GLB_{15}	20.144	9.203	0.198
	GLB_{16}	11.893	7.300	0.176
	GLB_{17}	46.920	15.606	0.230
	GLB_{18}	66.354	27.487	0.475

表13. 数据集 MSC 的仿真性能。

数据集	手势	RMSE	MAE	平均 U_1
MSC	MSC_1	4.101	2.143	0.069
	MSC_2	39.067	12.283	0.254
	MSC_3	69.123	34.171	0.468
	MSC_4	50.594	24.116	0.307
	MSC_5	52.428	25.538	0.370
	MSC_6	65.605	31.718	0.587
	MSC_7	23.665	2.197	0.030
	MSC_8	47.459	20.070	0.270
	MSC_9	55.631	26.040	0.313
	MSC_{10}	68.613	47.524	0.796
	MSC_{11}	79.191	48.210	0.841
	MSC_{12}	45.385	5.860	0.094
	MSC_{13}	68.059	23.202	0.290

参考文献

[1] L. Sigal、A. O. Balan和 M. J.Black，“HumanEva：用于评估人体关节运动的同步视频和动作捕捉数据集及基线算法”，《国际计算机视觉期刊》，第87卷，第 1-2 期，第4–27 页，2010年。[2] S. Ghorbani、K.Mahdaviani、A. Thaler、K.Kording、D. J. Cook、G. Blohm 和 N. F. Troje，“MoVi：大型多用途人体动作与视频数据集”，《PLoSONE》，第 16 卷，第 6 期，2021 年 6 月。 [在线]。可访问：http://dx.doi.org/10 1371 0253157 。/journal.pone。[3] H. Kuehne、H.Jhuang、E. Garrote、T.Poggio 和 T.Serre，“HMDB：用于人体动作识别的大型视频数据库”，载于《2011年国际计算机视觉大会》。IEEE，2011年 11 月，第 2556–2563 页。 [在线]。可访问：http://ieeexplore.ieee.org/document/6126543/ [4] F. Ofli、R. C haudhry、G.Kurillo、R. Vidal 和 R. Bajcsy，“伯克利多模态人体动作数据库”，载于《2013年IEEE计算机视觉应用研讨会（WACV）》。IEEE，2013年1月，第53–60页。[在线]。可访问：https://ieeexplore.ieee.org/document/ 6474999[5] “CMU图形实验室动作捕捉数据库。” [在线]。可访问：http://mocap.cs.cmu.edu/ [6] N. P. Van Der Aa、X. Luo、G. J. Giezeman、R. T. Tan和R. C. Veltkamp，“UMPM基准：一个带有同步视频和动作捕捉数据的多人数据集，用于评估人体关节运动和交互”，《IEEE国际计算机视觉大会论文集》，第1264–1269页，2011年。[7] C. Mandery、O. Terlemez、M. Do、N. Vahrenkamp和T. Asfour，“KIT全身人体运动数据库”，载于《2015年国际先进机器人会议（ICAR）》。IEEE，2015年7月，第329–336页。 [在线]。可访问：http://ieeexplore.ieee.org/document/7251476/ [8] H. Lee、K. Yang、N. Kim和C. R. Ahn，“使用带有格拉米角场的深度学习网络检测过度负重任务”，《建筑自动化》，第120卷，第103390页，2020年12月。 [在线]。可访问：https: //linkinghub.elsevier.com/retrieve/pii/S 0926580520309705 [9] P. Maurice、A. Malaisé、C. Amiot、N. Paris、G.-J. Richard、O. Rochel和S. Ivaldi，“人体运动与人体工学：一个面向工业的协作机器人数据集”，《国际机器人研究杂志》，第38卷，第14期，第1529–1537页，2019年12月。 [在线]。可访问：http: // journals.sagepub.com/doi/10.1177/0278364919882089

[10] S .-M. Liikkanen、J. Lylykangas、J. Pakkanen、J. Järvinen、R. Raisamo、K. Väänänen、S. Pakarinen、T. Keränen、J. Suhonen、S. Saarinen、S. Tuomisto、J. Järvinen 和 M. Bordallo López，“介绍 VTT-ConIot：一个真实的

基于 IMU 设备的建筑工人活动识别数据集”，《Sustainability》，第 14 卷，第 1 期，第 220 页，2021 年 12 月。 [在线]。链接：https://www.mdpi.com/ 2071 1050 14 1 220 - / / / Y.
[11] Ben-Shabat、X. Yu、F. S. Saleh、D. Campbell、C. Rodriguez-Opazo、H. Li 和 S. Gould，“IKEA ASM数据集：通过动作、物体和姿态理解人们组装家具的过程”，2020 年 7 月。 [在线]。链接：http://arxiv.org/abs/ 2007 00394 . C. Tamantini、F.
[12] Cordella、C. Lauretti 和 L. Zollo，“WGD——用于人体工学分析和预防工伤的装配线工作手势数据集”，《Sensors》，第 21 卷，第 22 期，第 7600 页，2021 年 11 月。 [在线]。链接：https://www.mdpi.com/ 1424 8220 21 22 7600 - / / /
[13] S. Manitsaris、G. Senteri、D. Makrygiannis 和 A. Glushkova，“用于识别专业手势并预测其轨迹的多变量时间序列人体运动表示”，《Frontiers in Robotics and AI》，第 7 卷，第 1–20 页，2020 年。[14] K. Schaub、G. Caragnano、B. Britzke 和 R. Bruder，“欧洲装配工作表”，《Theoretical Issues in Ergonomics Science》，第 14 卷，第 6 期，第 616–639 页，2013 年 11 月。[15] R. E. Kalman 等人，“线性滤波与预测问题的新方法”，《Journal of Basic Engineering》，第 82 卷，第 1 期，第 35–45 页，1960 年。[16] X. Wang、H. Ding、G. Trajcevski、P. Scheuermann 和 E. Keogh，“时间序列数据表示方法与距离度量的实验比较”，2010 年。 [在线]。链接：http://arxiv.org/abs/1012.2789



BRENDA ELIZABETH OLIVAS-PADILLA 于 2016 年获得墨西哥蒙特雷理工学院机电一体化工程学士学位，2018 年获得墨西哥国立理工学院电子工程硕士学位，并于 2023 年获得法国 PSL 大学实时计算机科学、机器人学、系统与控制博士学位。自 2018 年起

2019年之前，她曾在法国巴黎高等矿业学院 PSL 大学机器人中心担任研究工程师；2019年至2022年，她是博士生；自2023年起，她在同一机构担任博士后研究员。她的研究重点是可穿戴传感和运动建模，用于分析资深手工艺人和制造业操作人员的人体运动。此外，她还参与了两个 H2020 项目，为捕捉、处理和分析技术动作以促进人类学习和人体工学作出了贡献。



ALINA GLUSHKOVA 于 2012 年获得法国巴黎第八大学信息与传播学学士学位，以及巴黎第二先贤祠-阿萨斯大学和巴黎第九多菲纳大学的两个管理学硕士学位。她于 2016 年获得希腊马其顿大学应用信息学博士学位。自 2020 年起，她在……中心担任研究员

巴黎矿业学院机器人中心，她在那里共同负责硕士后项目 AIMove——“机器人与交互系统中的人工智能与运动”。她的研究重点是人机交互，更具体地说，是开发能够改进机器感知能力的系统和算法，使机器能够分析和评估人体运动，从而提供反馈，指导学习者的手势和姿势表现。她的研究成果已应用于专业手势的人体工学分析与纠正，以及专家手势的学习。



SOTIRIS MANITSARIS 是巴黎矿业学院机器人中心（巴黎文理研究大学）的副主任。他拥有希腊塞萨洛尼基亚里士多德大学应用数学学士学位、法国布莱兹·帕斯卡大学与希腊塞萨利大学工程学院联合授予的地方发展双硕士学位，以及该大学人工智能博士学位

希腊马其顿大学。此外，他还在生物医学工程、人机协作和基于动作的交互系统领域开展了三项博士后研究。他曾担任多个欧洲公共项目和工业项目的首席研究员，并领导硕士后项目 AIMove——“机器人与交互系统中的人工智能与动作”。2020 年，他加入欧洲议会科技选择评估委员会（STOA）的国际顾问委员会；该委员会通过新成立的人工智能中心，特别关注人工智能领域。他的研究重点是以人为中心的人工智能，包括机器学习和模式识别的概念与方法；这些概念与方法应用于从人体采集的信号，并将其作为与智能机器协作的模态。