基于卷积神经网络的 Leap Motion 运动数据优化网络

张欣天¹⁾, 谢文军²⁾, 李书杰¹⁾, 刘晓平^{1)*}

¹⁾ (合肥工业大学计算机与信息学院 合肥 230601)

²⁾ (合肥工业大学软件学院 合肥 230601)

(liu@hfut.edu.cn)

摘 要: 为提高 Leap Motion 设备的采集精准度, 解决自遮挡、采样频率不稳定等设备固有问题, 首先, 设计了使用 Leap Motion 和动作捕捉设备的手部多模态同步运动采集方案, 采集了日常动作数据集; 其次, 提出了基于卷积神经网络(convolutional neural network, CNN)的 Leap Motion 手部运动数据优化方法, 使用日常动作数据集训练 Leap Motion 数据到动作捕捉数据的映射网络; 最后, 提出手指平面约束, 确保网络输出数据保持稳定的手部骨骼结构. 通过 15 名志愿者采集了 6 类动作共 967 550 帧的同步运动数据集, 进行了手指平面约束有效性、动作一致性实验, 并与双向循环自编码器(bidirectional recurrent autoencoder, BRA)、双向编解码器(encoder-bidirectional-decoder, EBD)方法进行了精度对比. 结果表明, 文中方法支持使用 Leap Motion 获取固定采样频率且近似动捕设备精度的手部运动数据, 效果较 BRA 和 EBD 更加稳定平滑. 将文中方法应用于康复游戏中, 明显减少了交互动作识别的错误次数.

关键词: 运动数据优化; Leap Motion; 卷积神经网络; 多模态数据集; 自遮挡

中图法分类号: TP391.41

DOI: 10.3724/SP.J.1089.2021.18425

Convolutional Neural Networks Based Motion Data Optimization Networks for Leap Motion

Zhang Xintian¹⁾, Xie Wenjun²⁾, Li Shujie¹⁾, and Liu Xiaoping^{1)*}

¹⁾ (School of Computer Science and Information Technology, Hefei University of Technology, Hefei 230601)

²⁾ (School of Software, Hefei University of Technology, Hefei 230601)

Abstract: It is necessary to improve the capture accuracy and precision of Leap Motion equipment, and solve its inherent problems such as finger self-occlusion and instable sampling frequency, firstly, a multi-modal synchronous hand motion capture scheme is proposed based on Leap Motion and motion capture devices, and the dataset is captured correspondingly. Secondly, a convolutional-neural-network-based hand motion data optimization method for Leap Motion is presented. The proposed network is trained to learn the mapping from Leap Motion data domain to motion capture data domain with the synchronous dataset. Finally, a coplanar constraint for human fingers is proposed, which makes the outputs maintain stable hand skeleton structure. The 967 550 frames of synchronous hand motion data are captured in 6 categories from 15 volunteers. Experiments are designed for validating of the finger coplanar constraint, testing the consistency of the optimized motion data, and the comparison with BRA (bidirectional recurrent autoencoder) and EBD (encoder-bidirectional-decoder) methods. Experiments indicate that proposed method supports for capturing

收稿日期: 2020-06-05; 修回日期: 2020-11-12. 基金项目: 国家自然科学基金(61877016, 61972128, 61702155); 中央高校基本科研业务经费(JZ2018HGTA0215). 张欣天(1995—), 男, 硕士研究生, CCF 学生会员, 主要研究方向为计算机图形与可视化; 谢文军(1984—), 男, 博士, 实验师, CCF 会员, 主要研究方向为运动数据采集与交互; 李书杰(1983—), 女, 博士, 讲师, 硕士生导师, CCF 会员, 主要研究方向为运动数据处理、深度学习算法; 刘晓平(1964—), 男, 博士, 教授, 博士生导师, CCF 理事, 论文通讯作者, 主要研究方向为计算机图形学、协同计算.

hand motion data with fixed sample frequency and MoCap-like precision using a Leap Motion. Furthermore, the proposed method can obtain smoother and more stable results than BRA and EBD. The method is also applied in rehabilitation games, which significantly reduces the number of errors of hand interaction motion recognition.

Key words: motion data optimization; Leap Motion; convolutional neural networks; multi-modal synchronous dataset; self-occlusion

人手是人机交互过程中主要的交互手段^[1]. 手部运动数据在计算机领域拥有诸多用途, 如识别手势动作语义^[2]、分析手功能康复效果^[3]、支持虚拟现实自然交互^[4]等. 为了支撑这些应用, 便捷且尽可能精确地获取手部运动数据有着重要的现实意义. 现有最高精确度的方法是通过各类动作捕捉设备(下文简称动捕)采集手部运动, 但相关设备价格相对昂贵, 且需要贴身固定标记物或传感器, 使用户运动受到束缚, 影响用户体验. 近些年逐渐涌现出一批成本低、无标记、非接触的体感设备, 而面向手部的体感设备中最具有代表性的是 LEAP 公司出品的 Leap Motion, 此类非接触式的实时手部运动采集设备“为基于手势交互的人机交互系统带来变革”^[5].

Leap Motion 可采集范围为一个倒锥形, 采集范围距离设备中心点的垂直距离为 25~600 mm, 在采集数据的过程中不可避免地存在距离敏感的精度问题和自遮挡现象. 例如, 当用户手背面向红外相机做握拳动作时, 手指如图 1a 所示在弯曲时会被手背遮挡, 此时可能发生 Leap Motion 无法捕获到手, 出现如图 1b 所示手指变形的现象, 而如图 1c 所示动捕数据表现相对稳定. 一些研究者针对此问题提出了相应的精度优化策略, 如谷学静等^[6]利用层次化的校正方法提高 Leap Motion 的识

别准确性; 孙国道等^[7]针对单 Leap Motion 交互空间局限性带来的自遮挡问题, 提出使用双 Leap Motion 方案, 但仍存在校正方法单一、数据融合方法复杂等问题.

近年来, 深度学习方法在人体运动数据分析和优化领域展现出了出色的效果. 例如, Bütepage 等^[8]使用全连接生成网络实现运动数据缺损修复; Holden 等^[9]提出的自编码器可以在隐变量空间中对运动数据进行编辑和修正, 在相应数据集的支撑下, 其效率和效果皆得到了提升. 虽然手部运动数据也是全身运动数据的一部分, 但目前用神经网络进行针对手部数据优化的工作还较少, 其原因主要在于有标签的手部运动数据集较少. 现有主流人体运动数据集(如 CMU 运动数据集^[10], HDM05 数据集^[11]和 Berkeley MHAD 数据集^[12]等)均不包含手部数据. 针对手部建立的数据集通常会应用于机器视觉领域研究, 以 RGBD 图像或视频^[13-15]形式保存, 缺乏多模态的同步数据, 更缺乏适用于 Leap Motion 设备数据优化的专用数据集.

因此, 本文设计了面向 Leap Motion 和动捕设备的同步采集方案, 构建了多模态同步数据集; 搭建了基于卷积神经网络(convolutional neural networks, CNN)的深度神经网络, 对 Leap Motion 所采集的手部运动数据进行优化, 以期基于 Leap Motion 设备实现实时、非接触且有动捕精度级别运动数据采集. 本文的工作主要围绕以下 3 点开展.

(1) 针对使用深度学习方法进行手部运动数据优化的相应数据需求, 提出一种 Leap Motion 与动捕设备同步采集方案, 构建多模态手部运动同步数据集.

(2) 在同步数据集的基础上, 提出一种基于 CNN 的 Leap Motion 数据优化网络结构, 用于 Leap Motion 设备的实时数据优化.

(3) 针对神经网络输出结果的微抖动问题, 结合手部结构特征, 提出一种新的手指平面约束, 用于提高精度和运动平滑性.

图 1 遮挡状态下 Leap Motion 可能发生的采集错误和动捕采集的正确结果

1 相关工作

为探索 Leap Motion 在精度、稳定性等方面的具体表现, 研究者进行了一系列有针对性的实验分析. Weichert 等^[16]通过精确度较高的机械臂分析 Leap Motion 采集数据的精确度与鲁棒性, 结果显示对于静态位置的测量精度可达 0.2 mm, 而对于动态位置的测量精度要低得多. 这一结论也在随后的实验中得到证实, Guna 等^[5]利用模型手进行实际评测, 得出 Leap Motion 还无法应用于专业的跟踪系统中这一结论. Bachmann 等^[17]将鼠标与 Leap Motion 作为选择任务的输入设备进行对比, 得出 Leap Motion 设备在选择识别方面表现能力有限. 近 2 年, 一些研究团队对 Leap Motion 其他方面进行评估. Niechwiej-Szwedo 等^[18]将 Leap Motion 与标准的动作捕捉设备 Optotrak 进行对比, 在 15 名实验人员的参与下分别进行了 3 组实验, 在评估手的时空准确性、峰值速度误差方面与动捕设备仍有差距, 这可能会限制 Leap Motion 在某些场景下的使用. 文献[19]指出, Leap Motion 还存在采集频率不稳定的问题.

为了弥补上述 Leap Motion 问题, 研究者提出了各类数据优化方法. 例如, Ponraj 等^[20]利用隐马尔可夫模型进行运动数据的合成, 其对于手势识别有所改善. 但为了克服遮挡问题, 通过双 Leap Motion 同时进行数据采集是更加常见的方法. Fok 等^[21]为了处理用户手在实际交互过程中存在自遮挡的问题, 引入了额外的 Leap Motion 进行同步采集, 提高对美国手语(American sign language, ASL)的识别精度. Jin 等^[22]提出了一种基于双 Leap Motion 的手势识别系统, 用于机器人系统中的远程控制, 在通过增加传感器的同时, 自定义当前 Leap Motion 状态; 在实际应用中选择 2 个 Leap Motion 中“稳定”的作为系统的输入, 在一定程度上解决了自遮挡问题. 孙国道等^[7]综合分析、比较了 2 个 Leap Motion 摆放位置, 提出了以垂直方式摆放为最佳方式, 不过在数据融合方面只是简单地对 2 个 Leap Motion 采集的数据进行平均, 在某些情况下仍然无法彻底解决自遮挡问题. 多台 Leap Motion 的引入, 势必进行一系列的数据融合操作, 增加了计算复杂程度. 此外, 2 个设备发出的红外光会相互影响引起测量误差^[23]. Marin 等^[24]将 Leap Motion 与 Kinect 设备结合, 以此提高手势识别的准确率, 但该方法增加了系统的复杂度, 不利于系统移植. Cai 等^[25]提出了一种基于 Leap Motion 和 Metaio 的

交互式增强现实系统方法, 增强了系统的交互体验, 但是没有解决数据的精确度问题.

目前, 在人体运动数据处理方面, 研究者提出了多种可行的网络模型. Fragkiadaki 等^[26]采用了一种双向编解码器结构(encoder bidirectional decoder, EBD)模型, 但该模型本身并非用于数据去噪且未考虑运动学约束, 训练效果不佳. Holden 等^[27]提出基于 CNN 的数据自编码器处理运动数据, 但在实际处理过程中, 网络输出的数据会出现不同程度的抖动和骨骼变形现象. Mall 等^[28]提出了双向编解码滤波器模型(encoder-bidirectional filter, EBF), 该模型无法处理数据输入与输出对应的拓扑结构不同的情况, 即无法让 Leap Motion 数据精确度优化到动捕级别. Li 等^[29]提出双向循环自编码器(bidirectional recurrent autoencoder, BRA), 用于对带有多种噪声的骨骼数据进行优化, 得到的运动数据更加符合运动学特征, 具有更好的视觉效果. 受上述文献的启发, 本文提出利用动捕设备与 Leap Motion 同步采集的方式进行数据采集, 利用 CNN 对 Leap Motion 采集的数据进行优化. 整体框架如图 2 所示.

图2 系统框架示意图

2 同步数据采集方案和数据集构建

提高深度神经网络模型的准确度需要大量数据的支撑, 调研发现, 目前并没有适用于 Leap Motion 等手部体感设备的公开同步手部运动数据集. 因此, 本文首先利用 Perception Neuron 和 Leap Motion 设备建立多模态手部运动同步数据采集方案.

2.1 采集方案设计

采集系统设计如图 3 所示. Leap Motion 采用桌面工作模式, 平置于用户前方的桌面. 用户手部穿戴 Perception Neuron 运动捕捉设备, 手掌中心水平位置与 Leap Motion 中心重合, 垂直距离为 15~20 cm. 为了减少由于超出 Leap Motion 有效采集范围而形成长时间错误数据, 在桌面绘制参考量程边界. 其中上下边界距离设备 20 cm, 左右边界距离设备 27 cm.

图 3 采集系统组成

协助采集本实验数据集的志愿者共有 15 名, 年龄分布在 20~27 岁. 结合日常生活中常见手动作, 采集的动作类型有翻转、握拳、对指和食指指向, 采集手均为左手, 每人每种动作采集 10 次以上, 共形成 6 组数据, 共计训练集帧数 797 325 帧, 测试集帧数为 170 225 帧.

动捕设备、Leap Motion 对应的左手拓扑结构如图 4 所示, 采集的每一帧运动数据由所有节点三维坐标构成.

图 4 设备采集的手拓扑结构

由于 Leap Motion 只可对手部进行识别^[17], 直接佩戴动捕设备的传感器绑带会破坏手部特征, 导致 Leap Motion 识别困难. 本文实验通过佩戴棉手套将动捕传感器包裹, 减少佩戴传感器造成的手指表面跳变, 同时维持与皮肤接近的红外光反射率, 以降低引入动捕设备对 Leap Motion 设备采集的影响.

2.2 数据预处理与时空同步

2.2.1 空间同步

对原始数据进行归一化处理, 统一空间尺度. 对于动捕数据, 将图 4a 中节点 0 作为根节点; 相应地, 对于 Leap Motion 数据, 将图 4b 相对位置较近的节点 21 作为根节点, 将其他节点坐标转换为相对坐标. 同时, 依据所采集的骨骼长度大小计算, Leap Motion 与动捕数据的单位比例为 1000 : 1, 通过对 Leap Motion 数据进行缩放, 达到尺度归一化.

数据归一化后, 本文以动捕数据位姿为基准,

将 Leap Motion 数据进行水平旋转 $R(\mathbf{d}_{\text{Leap}}, \mathbf{d}_{\text{Mocap}})$, 其中, $\mathbf{d}_{\text{Leap}}$ 和 $\mathbf{d}_{\text{Mocap}}$ 分别为 Leap Motion 和动捕数据中根节点到中指指尖的向量.

对 Leap Motion 中 x, z 坐标进行矩阵操作

$$\begin{pmatrix} x' & y & z' \end{pmatrix}^T = \begin{pmatrix} x & y & z \end{pmatrix}^T \times R(\mathbf{d}_{\text{Leap}}, \mathbf{d}_{\text{Mocap}}) \quad (1)$$

将 Leap Motion 数据与动捕数据空间对齐.

2.2.2 时间同步

本方案使用的动捕设备采样频率为稳定 60 Hz, 而 Leap Motion 采样率并不稳定, 均值约为 115 Hz. 因此, 以动捕设备的采样率和时间域为基准进行时间同步. 在数据采集时要求动捕设备最后打开、最先关闭, 以保证其采集时间域中一定存在对应的 Leap Motion 数据. 假设 T_{Mocap1} 和 T_{Mocap2} 分别代表动捕设备开始和结束采集时刻; T_{Leap1} 和 T_{Leap2} 代表 Leap Motion 设备开始和结束采集时刻. 对于时域 T_{Mocap1} 和 T_{Mocap2} 中每一帧数据 $\mathbf{d}_{\text{Mocap}_i}$, 获取其时间值 t_{Mocap_i} ; 通过三次样条差值, 如图 5 所示从 Leap Motion 的原始数据中采样数据

$$\mathbf{d}_{\text{Leap}_i} = f_{\text{resample}}(\mathbf{d}_{\text{Mocap}_i}, t_{\text{Mocap}_i}),$$

最终形成时域同步的 Leap Motion 数据.

图 5 帧数据插值示意图

3 数据优化网络

3.1 数据初始化

对同步后的原始数据, 采用滑动窗口的方式分别对 Leap Motion 和动捕数据进行分割. 设 T 为滑动窗口的长度, S 为滑动步长, 本文取 $T = 60$, $S = 30$, 将数据初始化为数据段输入深度神经网络.

然后, 将所有数据随机分为训练集和测试集 2 个部分, 其占比分别为 80% 和 20%. 为了确保训练集和测试集中的样本在语义上同分布, 对于每个运动语义数据的划分均保持该比例.

3.2 网络结构设计

本文提出的网络结构由编码器和解码器 2 个部分组成, 如图 6 所示. Leap Motion 源数据输入维度为 $m \times 60 \times 69$. 其中, m 为同时输入到网络中数据文件个数; 每个数据文件包含 60 帧数据; 69 代表 Leap Motion 23 个关节, 每个关节对应 X, Y, Z 这 3 个通道. 编码器由卷积层、激活层和池化层组成. 其中, 卷积层的维度为 $25 \times 69 \times 256$, 分别代表卷积核的宽度、输入数据列数和卷积核的数量; 卷积层的输出数据经过激活函数 ReLU 输入到池化层. 本文采用的池化方法是最大池化. 在解码器部分, 包含反池化层和反卷积层 2 部分. 与卷积层对应, 反卷积层的维度为 $25 \times 256 \times 60$; 其中, 60 表示为输出数据特征数(20 个节点, 每个节点对应 3 个通道).

图6 面向多模态手部同步数据的优化网络结构

在网络训练过程中, 本文引入了文献[27]对于全身运动数据优化时使用的空间位置约束、骨骼长度约束和平滑性约束, 并根据手指运动学特征, 提出了手指平面约束.

3.2.1 手指平面约束

人手骨骼可近似看做刚体结构, 尤其是各个手指关节. 在手运动过程中, 指尖、指端关节、指间关节和指掌关节构成的向量 v_1, v_2, v_3 始终处于同一平面内, 如图 7 所示. 现以食指为例, 设某次迭代过程中, 网络输出为 $P_O = \{P_{O0}, P_{O1}, \dots, P_{O20}\}$. 其中, $P_{O4}, P_{O5}, P_{O6}, P_{O7}$ 分别为食指上 4 个节点坐标, 则向量 v_1, v_2, v_3 计算公式为

$$v_1 = P_{O4} - P_{O5} \quad (2)$$

$$v_2 = P_{O6} - P_{O5} \quad (3)$$

$$v_3 = P_{O7} - P_{O5} \quad (4)$$

将得到的向量单位化, 得到其对应的单位向量 e_1, e_2, e_3 . 向量对应的混合积为

$$F_{\text{index}} = e_1 \cdot (e_2 \times e_3) \quad (5)$$

当向量 e_1, e_2, e_3 共面时, 其对应的混合积应为零; 混合积越小, 表明越符合平面约束. 因此, 引入关于手指平面约束的损失函数定义为

$$L_f = \frac{1}{4} \times (\|F_{\text{index}}\|^2 + \|F_{\text{middle}}\|^2 + \|F_{\text{ring}}\|^2 + \|F_{\text{little}}\|^2) \quad (6)$$

其中, $F_{\text{index}}, F_{\text{middle}}, F_{\text{ring}}$ 和 F_{little} 分别为食指、中

指、无名指和小拇指对应的混合积. 由于拇指结构特殊性, 相较于其他 4 指缺少指间关节, 因此本文未考虑对拇指的平面约束.

图7 用于手指平面约束计算的向量

3.2.2 空间位置约束

本实验中位置的输入为相对坐标, 为保证网络输出结果尽可能与原始位置相近, 为方便后续将相对坐标转换为绝对值坐标, 本文引入了空间位置约束. 空间位置约束相关的损失函数定义为

$$L_p = \frac{1}{2} \times \frac{1}{m} \times \sum_{j=0}^{19} \|P_{\text{Out}_j} - P_{\text{Leap}_j}\|^2 \quad (7)$$

其中, m 的值为 20, 代表输出的 20 个关节; P_{Out_j} 和 P_{Leap_j} 分别对应网络输出和源数据中编号为 j 的坐标值.

3.2.3 骨骼长度约束

为保证数据优化前后手部骨骼长度具有一致性, 在训练过程中添加骨骼长度约束. 设某次迭代过程中网络输出集合 $P_{\text{Out}} = \{P_{\text{Out}_0}, P_{\text{Out}_1}, \dots, P_{\text{Out}_{19}}\}$, 则拇指远指端骨骼对应节点为 P_{Out_2} 和 P_{Out_3} , 如图 4 所示, 相应的长度 $D_{\text{Out}_{2_3}}$ 可表示为

$$D_{\text{Out}_{2_3}} = \|P_{\text{Out}_3} - P_{\text{Out}_2}\| \quad (8)$$

计算其余骨骼节点长度方法类似. 集合 D_{Out} 保存最终网络输出的手部骨骼长度, D_{Leap} 表示 Leap Motion 捕获的手部骨骼长度, 最终得到关于骨骼长度的损失函数为

$$L_b = \sum_{i=1}^{60} \sum_{j=0}^{19} \|D_{\text{Out}_i} - D_{\text{Leap}_i}\| \quad (9)$$

其中, D_{Out_i} 和 D_{Leap_i} 分别为第 i 帧对应的手部骨骼长度集合.

3.2.4 平滑性约束

为确保相邻帧间的数据连续性, 对网络输出后

的数据进行帧间平滑性约束. 其损失函数^[29]定义为

$$L_s = \frac{1}{j} \|\mathbf{OY}'\|_2 \quad (10)$$

其中,

$$\mathbf{O} = \begin{pmatrix} -1 & 1 & 0 & & 0 & 0 & 0 \\ 1 & -2 & 1 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -2 & & 0 & 0 & 0 \\ & \vdots & & \ddots & & & \\ 0 & 0 & 0 & & -2 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 1 & -2 & 1 \\ 0 & 0 & 0 & & 0 & 1 & -1 \end{pmatrix}_{(m+2) \times (m+2)} \quad (11)$$

表示平滑矩阵; m 代表帧数; j 为关节的个数; $\mathbf{OY}'$ 是通过在网络输出的基础上重复第 1 帧数据和最后 1 帧数据得到.

网络最终采用的损失函数为

$$L = a \times L_p + b \times L_s + c \times L_b + d \times L_r \quad (12)$$

为确保 L_p , L_s 与 L_b 在同一数量级, 方便网络训练, 本文中设 $a = d = 1$, $b = c = 0.000\ 2$.

4 实验与分析

本文的网络模型训练和应用均在个人计算机上进行, 其中, CPU 为 Intel i7-7700K@4.20 GHz, 内存为 16 GB, 显卡为 NVIDIA GeForce 1080 GTX, 显存为 8 GB; 操作系统为 Windows 10, 64 位, 深度学习环境为 TensorFlow 1.7.0, 应用展示平台为 Unity3D 引擎.

4.1 手指平面约束有效性

为验证第 3 节提出的手指平面约束对于手部运动特征的有效性, 本节分别选取拇指对食指、拇指对中指与拇指对无名指 3 种手部动作序列作为输入, 记 CNN 在添加了空间位置约束、骨骼长度约束和平滑性约束后的网络输出为 CNN_{nf} , 记 CNN 同时添加第 3.2 节中介绍的 4 种约束后的输出为 CNN_{f} , 各选取运动片段中对应的 3 帧数据作为结果展示, 如图 8 所示.

图 8 运动片段效果对比

从图 8 可以看出, CNN_{nf} 输出的结果中某些骨骼关节会扭曲变形; 在拇指对食指运动过程中, CNN_{f} 无法较好地还原拇指动作; 此外, CNN_{nf} 对应的手部运动数据实际播放会有明显的抖动现象; 从而可以验证本文提出的手指平面约束的有效性.

4.2 数据优化后与实际动作一致性实验

为验证本文网络在自遮挡现象发生时的处理能力, 分别选取掌心背向 Leap Motion 做握拳动作 (见图 1) 和手翻转这 2 种动作进行实验. 第 1 种动作之前已进行分析, 在手翻转过程中, 当掌心方向与 Leap Motion 垂直时, 靠近 Leap Motion 位置的手指会不同程度地遮挡住较远位置的手指, 使数据精确度受到影响. 在握拳过程中, 选择食指指间关节角度和指掌关节角度作为对比, 在运动过程中比较原始数据值、优化后数据和动捕数据之间大小, 实验结果如图 9 所示. 在手翻转和握拳过程中, 选择中指 (11 号节点) 空间运动轨迹作为对比, 其对应的实验结果如图 10 所示.

由于在握拳的过程中, 手掌遮挡住 Leap Motion 探测器的视野, 发生自遮挡现象. 图 9 中原始数据关节角度变化幅度远没有达到实际中达到的角度范围, 优化后数据对应的角度变化趋势与动捕数据近似, 表明数据优化后可以较好地恢复原始的手部动作.

由图 10 可知, 不论在原始数据噪声较大 (图 10a) 或相对平稳 (图 10b) 的情况下, 优化后数据的空间节点位置仍能够接近动捕原始轨迹.

图 9 握拳动作中手指的主要关节角度结果对比

图10 中指指尖运动轨迹对比图(单位: 1/10 mm)

4.3 对比实验

手部动作帧之间的平滑性高低、位置误差大小

图11 不同方法在3种测试动作下对应误差对比

等因素是衡量网络结构输出的重要指标, 因此, 本文选择数据集中握拳、拇指对中指、食指指向3种手部数据作为测试集, 比较CNN结构^[27], BRA结构^[29], EBD结构^[28]和本文提出的优化方法各自对应的3种类型误差: 手骨骼长度误差和帧间误差、空间位置误差, 以此进行对比实验, 结果如图11所示。

经过实验可得出如下结论。

(1) 在空间位置误差方面, 本文方法与CNN结构相近, EBD和BRA方法的位置误差相对略高。

(2) 对于选择的3种手部动作, 本文方法得到的手骨骼长度误差均为最低值, 更为接近实际动捕数据值。

(3) 所有动作对应的平滑度误差中, 本文方法误差最低, 总体稳定在0.2左右, 视觉效果更佳。

综上所述, 本文提出的基于CNN的Leap Motion运动数据优化网络总体表现优于其他网络模型, 且手指平面约束的引入使之能够针对性地提高优化结果的平滑性和视觉效果。

4.4 数据优化方法有效性分析

为了验证本文所提的方法在应用场景中的实际效果, 本文选取了抓放目标物体的简单交互场景进行相关实验, 并邀请了15名年龄分布为20~27岁的志愿者来进行交互系统的任务测试。他们中既有熟悉Leap Motion设备的研究人员, 也有了解或未接触Leap Motion设备的一般人员。

具体实验内容如下: 虚拟场景上方区域会随机刷新 4 个不同物体, 在 60s 规定时间内, 参与者通过捏合抓取、捏住移动、释放等一系列动作, 将物体放入下方盒子内。每放置成功一次, 场景内会随机出现新的物体, 如图 12 所示。

图 12 抓放物体场景

在实验过程中, 记录并保存了参与者手部全程运动数据, 并将数据通过本文方法进行优化, 随后将优化后的数据重新应用在场景中。分别记录数据优化前后每位参与者抓取失败次数和最终成绩并计算出平均抓取失败次数。

参与者的抓取失败次数统计结果如图 13 所示。数据优化前后的平均抓取失败次数分别为 4.2 和 1.4, 可见随着数据准确度的提升, 由误判导致的抓取失败次数下降, 使判定结果更加符合真实情况。

图 13 各参与者抓取失败次数

实验前后各参与者的最终成绩结果如图 14 所示。由于运动数据优化后, 在抓取过程中手指动作更符合预期, 使整体最终得分提高了 12.45%。

实验结果表明, 尤其对未接触过 Leap Motion 设备的一般人员(P_3 , P_6 , P_9 , P_{12} , P_{15})抓取的数据, 本文数据优化效果提升明显, 更符合用户的交互需求。

图 14 各参与者最终成绩结果

5 结 语

目前对于 Leap Motion 设备局限性导致的数据精度下降问题, 利用深度神经网络结构对其数据进行优化的相关研究仍较少。针对缺乏适用于 Leap Motion 设备数据优化的数据集, 本文设计出了同步采集方案, 建立了面向 Leap Motion 和动捕设备的同步数据集。基于此, 本文提出了基于 CNN 的 Leap Motion 运动数据优化网络, 较好地解决了 Leap Motion 数据精度问题, 有效地提高了交互过程中识别准确性。此外, 本文根据手部特征, 在网络训练过程中创新性地引入手指平面约束, 很好地解决了网络输出抖动问题。

本文方法仍有较大的扩展空间。首先, 通过设计和收集更加完备的同步数据采集方案, 能够使其进一步扩展到更多的应用场合; 其次, 当前的输入设备只采用单一 Leap Motion 设备, 未来将扩展到多个 Leap Motion 输入的情况, 从而扩大系统采集范围。

参考文献(References):

- [1] Zhang Fengjun, Dai Guozhong, Peng Xiaolan. A survey on human-computer interaction in virtual reality[J]. Scientia Sinica: Informationis, 2016, 46(12): 1711-1736(in Chinese) (张凤军, 戴国忠, 彭晓兰. 虚拟现实的人机交互综述[J]. 中国科学: 信息科学, 2016, 46(12): 1711-1736)
- [2] Ye Sufen, Lai Jizhou, Lyu Pin, et al. An intelligent perception and gesture recognition technology for wearable piano-playing glove[J]. Chinese Journal of Scientific Instrument, 2019, 40(5): 187-194(in Chinese) (叶素芬, 赖际舟, 吕品, 等. 穿戴式钢琴弹奏手套智能感知与手势识别技术[J]. 仪器仪表学报, 2019, 40(5): 187-194)
- [3] Yi Rongwu, Wang Aimin. Design of remote rehabilitation training robotic system for fingers[J]. Electronic Measurement Technology, 2016, 39(9): 128-132+140(in Chinese) (易荣武, 王爱民. 手指远程康复训练机器人系统设计[J]. 电子测量技术, 2016, 39(9): 128-132+140)
- [4] Poupayev I, Ichikawa T, Weghorst S, et al. Egocentric object

- manipulation in virtual environments: empirical evaluation of interaction techniques[J]. *Computer Graphics Forum*, 1998, 17(3): 41-52
- [5] Guna J, Jakus G, Pogačnik M, *et al.* An analysis of the precision and reliability of the leap motion sensor and its suitability for static and dynamic tracking[J]. *Sensors*, 2014, 14(2): 3702-3720
- [6] Gu Xuejing, Wang Xu, Kan Kuo. Dynamic gesture interaction system based on Leap Motion[J]. *Computer Engineering and Applications*, 2020, 56(10): 220-225(in Chinese)
(谷学静, 王旭, 阚阔. Leap Motion 手势交互层次化校正方法研究[J]. *计算机工程与应用*, 2020, 56(10): 220-225)
- [7] Sun Guodao, Huang Puyong, Liu Yipeng, *et al.* Interactive 3D visualization with dual Leap Motions[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2018, 30(7): 1268-1275(in Chinese)
(孙国道, 黄普永, 刘义鹏, 等. 基于双 Leap Motion 的三维可视化交互方法研究[J]. *计算机辅助设计与图形学学报*, 2018, 30(7): 1268-1275)
- [8] Bütepage J, Black M J, Kragic D, *et al.* Deep representation learning for human motion prediction and classification[C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2017: 6158-6166
- [9] Holden D, Saito J, Komura T. A deep learning framework for character motion synthesis and editing[J]. *ACM Transactions on Graphics*, 2016, 35(4): Article No.138
- [10] Carnegie Mellon University. CMU graphics lab motion capture database[OL]. [2020-06-05]. <http://mocap.cs.cmu.edu/>
- [11] Müller M, Röder T, Clausen M, *et al.* Documentation mocap database HDM05[R]. Bonn: Universität Bonn, Insitut für Informatik II, 2007
- [12] Ofli F, Chaudhry R, Kurillo G, *et al.* Berkeley MHAD: a comprehensive multimodal human action database[C] // *Proceedings of the IEEE Workshop on Applications of Computer Vision*. Los Alamitos: IEEE Computer Society Press, 2013: 53-60
- [13] Cao Z, Hidalgo G, Simon T, *et al.* OpenPose: realtime multi-person 2D pose estimation using part affinity fields[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(1): 172-186
- [14] Athitsos V, Neidle C, Sclaroff S, *et al.* The American sign language lexicon video dataset[C] // *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. Los Alamitos: IEEE Computer Society Press, 2008: 1-8
- [15] de Smedt Q, Wannous H, Vandeborje J P, *et al.* 3D hand gesture recognition using a depth and skeletal dataset[C] // *Proceedings of the 10th Eurographics Workshop on 3D Object Retrieval*. Heidelberg: Springer, 2017: 1-6
- [16] Weichert F, Bachmann D, Rudak B, *et al.* Analysis of the accuracy and robustness of the leap motion controller[J]. *Sensors*, 2013, 13(5): 6380-6393
- [17] Bachmann D, Weichert F, Rinkenauer G. Evaluation of the leap motion controller as a new contact-free pointing device[J]. *Sensors*, 2015, 15(1): 214-233
- [18] Niechwiej-Szwedo E, Gonzalez D, Nouredanesh M, *et al.* Evaluation of the Leap Motion controller during the performance of visually-guided upper limb movements[J]. *PLoS One*, 2018, 13(3): e0193639
- [19] Vosinakis S, Koutsabasis P. Evaluation of visual feedback techniques for virtual grasping with bare hands using Leap Motion and Oculus Rift[J]. *Virtual Reality*, 2018, 22(1): 47-62
- [20] Ponraj G, Ren H L. Sensor fusion of leap motion controller and flex sensors using Kalman filter for human finger tracking[J]. *IEEE Sensors Journal*, 2018, 18(5): 2042-2049
- [21] Fok K Y, Ganganath N, Cheng C T, *et al.* A real-time ASL recognition system using leap motion sensors[C] // *Proceedings of International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*. Los Alamitos: IEEE Computer Society Press, 2015: 411-414
- [22] Jin H Y, Chen Q, Chen Z X, *et al.* Multi-LeapMotion sensor based demonstration for robotic refine tabletop object manipulation task[J]. *CAAI Transactions on Intelligence Technology*, 2016, 1(1): 104-113
- [23] Placidi G, Cinque L, Petracca A, *et al.* A virtual glove system for the hand rehabilitation based on two orthogonal LEAP motion controllers[C] // *Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods*. Heidelberg: Springer, 2017: 184-192
- [24] Marin G, Dominio F, Zanuttigh P. Hand gesture recognition with jointly calibrated Leap Motion and depth sensor[J]. *Multimedia Tools and Applications*, 2016, 75(22): 14991-15015
- [25] Cai X Q, Tu Y X, He X. An interactive augmented reality system based on LeapMotion and Metaio[C] // *Proceedings of the 11th International Conference on Multimedia and Ubiquitous Engineering and the 12th International Conference on Future Information Technology*. Heidelberg: Springer, 2018: 237-244
- [26] Fragkiadaki K, Levine S, Felsen P, *et al.* Recurrent network models for human dynamics[C] // *Proceedings of the IEEE International Conference on Computer Vision*. Los Alamitos: IEEE Computer Society Press, 2015: 4346-4354
- [27] Holden D, Saito J, Komura T, *et al.* Learning motion manifolds with convolutional autoencoders[C] // *Proceedings of SIGGRAPH Asia 2015 Technical Briefs*. New York: ACM Press, 2015: 1-4
- [28] Mall U, Lal G R, Chaudhuri S, *et al.* A deep recurrent framework for cleaning motion capture data[OL]. [2020-06-05]. <https://arxiv.org/abs/1712.03380>
- [29] Li S J, Zhou Y, Zhu H S, *et al.* Bidirectional recurrent autoencoder for 3D skeleton motion data refinement[J]. *Computers & Graphics*, 2019, 81: 92-103