

摘 要

全手行为跟踪是实现手势交互的重要技术之一，在人机交互的诸多场景中展现了应用价值。目前手部跟踪技术仍存在运动导致佩戴错位等问题，同时手势交互技术也面临 MidasTouch 等问题。在全手行为跟踪技术中，由于通过加速度计获取空间位置存在漂移问题，因此，本文面向基于姿态的全手行为跟踪与手势交互技术，开发手部约束特性以优化全手行为追踪技术，研究震颤信息区分交互意图以探究更自然的空中手势交互方法。

本文主要工作如下：

1) 研究了一种全手行为跟踪优化方法。首先，本文依据传感器姿态估计方法，构建全手行为跟踪算法。其次开发手部约束特性，将手部约束作为先验知识融合进全手行为跟踪算法中，作为全手行为跟踪优化算法。最后通过实验验证在对 PIP 姿态求解中优化算法相较于未优化算法的 $q_3(F_{1,33} = 396.4, p < 0.001, \eta^2 = 0.46)$ 和 $q_4(F_{1,33} = 396.3, p < 0.001, \eta^2 = 0.46)$ 误差存在显著差异，表明姿态优化方法结果更加贴近真实手部姿态。

2) 研究了一种基于姿态的手势集与实现方法，分析了基于震颤的交互意图识别方法。首先，基于以用户为中心的设计方法研究了一种基于姿态的手势集合；然后，分析了空中手势交互坐标映射方法与手势实现方法；最后，通过设计实验，提取用户震颤信息以识别用户交互意图行为，为空中手势交互中的 MidasTouch 问题提出一种解决方法。实验结果表明，通过震颤特征区分用户行为有无交互意图达到了 97% 的准确率。

3) 设计并实现了基于姿态的空中手势交互原型系统 iGestureGlove。在该交互系统中，实现了空中手势交互的五类基本功能：平移、选择、抓取、旋转和缩放，并且针对该空中手势交互系统设计了无线传输数据手套，为基于姿态的空中单手手势交互提供支持。

本研究面向基于姿态的全手行为跟踪与手势交互技术，通过开发手部约束特性以提高姿态跟踪的准确度，本研究所提方法可扩展到更多的约束维度，最终提高姿态估计方法的性能和可用性；本研究分析了行为震颤与交互意图的关系，本方法通过行为获取用户更多维度的信息，可以提高空中手势交互的自然性，这对于全手行为跟踪技术以及空中手势交互技术具有参考意义。

关键词： 人机交互；手势交互；手部姿态；行为跟踪；交互意图

Abstract

The whole-hand behavior tracking is one of the important technologies to realize gesture interaction, which shows its application value in many human-computer interaction scenes. At present, there are still some problems in hand tracking technology, such as misplacement caused by movement, and the hand gesture interaction technology also faces MidasTouch and other problems. Because of the serious drift problem in obtaining spatial position by acceleration integration, this paper develops hand constraint feature to optimize the whole hand behavior tracking technology for gesture-based hand tracking and gesture interaction, in order to explore a more natural method of air gesture interaction, vibration information is used to distinguish the intention of interaction.

The main work of this paper is as follows:

- 1) A whole-hand behavior tracking optimization method is studied. Firstly, the whole-hand behavior tracking algorithm is constructed based on the sensor attitude estimation method. Secondly, the hand constraint is fused into the whole-hand behavior tracking algorithm as a prior knowledge and used as the optimal algorithm of whole-hand behavior tracking. Finally, the experimental results show that there is a significant difference between the optimal algorithm and the non-optimal algorithm $q_3(F_{1,33} = 396.4, p < 0.001, \eta^2 = 0.46)$ and $q_4(F_{1,33} = 396.3, p < 0.001, \eta^2 = 0.46)$.
- 2) The user-centered method is used to design the gesture set based on the gesture, and the coordinate mapping method and the gesture implementation method in the air gesture interaction are proposed. Finally, the user experiment is designed to use the extracted tremor information to identify the user interaction intention, and a solution to the MidasTouch problem in air gesture interaction is proposed. The experimental results show that the accuracy of the extracted tremor features is 97%.
- 3) An air gesture interaction prototype system iGestureGlove based on attitude is designed and implemented. In this interactive system, five basic functions of air gesture interaction are realized: translation, selection, grasping, rotation and zooming, provides support for gesture-based air-to-air single-hand gesture interaction.

This research is aimed at gesture-based whole-hand behavior tracking and gesture interaction technology. By developing hand constraint features to improve the accuracy of gesture tracking, the proposed method can be extended to more constraint dimensions, finally, the performance and usability of the attitude estimation method are improved, and the relationship between

behavior tremor and interaction intention is analyzed, it can improve the naturalness of air gesture interaction, which has reference significance for the whole hand behavior tracking technology and air gesture interaction technology.

Key words: Human-Computer interaction; Gesture recognition; Hand attitude; Motion tracking; Interaction intention

目录

摘 要.....	I
Abstract	II
1 绪论.....	1
1.1 研究背景与意义.....	1
1.2 国内外研究现状.....	1
1.3 主要研究内容.....	5
1.4 论文结构.....	5
2 手部追踪与交互方法相关研究.....	7
2.1 手部追踪相关研究.....	7
2.1.1 基于磁感应的手部追踪.....	7
2.1.2 基于计算机视觉的手部追踪.....	8
2.1.3 基于惯性传感器的手部追踪.....	9
2.2 手势交互相关研究.....	10
2.2.1 基于原始数据的手势识别.....	10
2.2.2 基于手部姿态的手势识别.....	11
2.3 本章小结.....	12
3 基于姿态的全手行为跟踪.....	13
3.1 姿态表示与关节特性.....	13
3.1.1 姿态描述方法.....	13
3.1.2 手部关节特性.....	17
3.2 全手行为跟踪方法.....	19
3.2.1 IMU 姿态求解方法.....	19
3.2.2 全手行为跟踪方法.....	21
3.3 全手跟踪优化方法.....	22
3.3.1 全手行为追踪的误差分析.....	23
3.3.2 构建手部非线性约束模型.....	25
3.3.3 非线性约束下的姿态求解.....	27
3.3.4 实验分析.....	29
3.4 本章小结.....	33
4 基于姿态的手势交互技术.....	34
4.1 基于姿态的平面交互手势设计.....	34
4.1.1 平面交互基本功能.....	35
4.1.2 用户设计手势实验.....	36

4.2 平面空间下手势交互实现方法	41
4.2.1 手部姿态与平面空间坐标映射	42
4.2.2 基于手部姿态的交互指令实现	44
4.3 基于姿态的交互意图识别方法	45
4.3.1 手部行为震颤与交互意图	45
4.3.2 交互中手部震颤提取方法	47
4.3.3 基于震颤的交互意图识别	50
4.3.4 实验分析	54
4.4 本章小结	62
5 iGestureGlove 原型系统设计与实现	63
5.1 iGestureGlove 的总体设计方案	63
5.1.1 原型系统硬件设计	64
5.1.2 原型系统软件结构	65
5.2 iGestureGlove 原型系统实现与应用	66
5.3 本章小结	68
6 总结与展望	69
6.1 总结	69
6.2 展望	69
参考文献	70

1 绪论

1.1 研究背景与意义

随着信息技术的快速发展，人机交互方式日益丰富化和智能化。手势交互作为一种自然、直观的交互方式，越来越多地应用到日常生活中，与语音、键鼠等其他交互方式相比，手势交互具有操作空间大、表达能力强的特点，是实现人机自然交互的重要方式之一。作为人机交互领域中自然、灵活、高效的交互模式之一，手部交互能够提供人手行为的直接映射，降低交互过程中的操作复杂度，提高交互效率，从而进一步地提高用户的交互体验。

手势交互已经广泛应用于日常生活的各个领域，而随着这些应用交互行为复杂度的提高，对手部运动追踪的精确度要求也越来越高。例如，基于手势交互的远程遥控系统往往需要实现复杂的精细化控制，如远程手术中通过机械臂进行伤口缝合；医疗康复评估和治疗中，高精度的手部运动捕捉可以帮助医生更准确地判断患者的康复状况并制定后续方案，也可以使患者获得及时的运动反馈来纠正错误动作，避免二次伤害。手部追踪技术的不断进步使这些医疗评估和治疗活动量化和规范化，从而更加精准高效。总体来说，这些新兴应用对手势交互精确度提出了更高要求，如何继续提高手部行为捕捉和分析的准确性仍是该领域当前的关键研究方向。

在手势交互的应用领域，往往在不同的交互环境中，会有着独立的手势交互指令与手势集，这意味着用户在使用不同的交互设备时，需要时间熟悉对应的指令系统。例如在手机的智慧交互领域，通过单手的旋转和抓握实现了翻页、截屏等操作；在智慧驾驶领域，通过在空中做出手势实现接听电话、切换歌曲、控制音量、快捷导航等操作，在熟悉操作手势之后，会使驾驶过程中的操作更加自然便捷，也使得驾驶员的驾驶体验更加舒适愉悦，但是复杂且难记的操作指令和手势往往会适得其反；在智慧课堂领域中，通过适配且符合场景的手势控制呈现互动式、情景式教学课堂，不仅可以极大地激发学生的积极性，更能够加深学生对知识的理解和记忆。此外，针对不同交互任务和场景，手势功能范围存在差异。因此如何根据交互任务和环境特征，定义和选择合适的手势集合是手势交互系统中需要考虑的环节，同时提高交互过程中的流畅度，通过交互在真实环境与虚拟环境中流畅的完成交互目的，也是手势交互应用领域中所面临的挑战。

综上所述，日益增长的应用场景和需求，以及提升用户体验的要求，都使手部跟踪技术在跟踪效果和交互体验等方面面临新的挑战。因此研究更高精度的全手行为跟踪技术和自然的手势交互技术，具有重要的研究和应用价值。

1.2 国内外研究现状

目前，手部交互方式可以分为手势识别和手部跟踪两类。与手势识别相比，手部跟

踪可以获取手部运动中的更多细节，能够提供更加灵活和自主的交互，在一些注重手部交互细节的场景中具有更多优势。由于手部拥有多自由度，其空间姿态信息相对复杂，手部跟踪具有非常重要的研究价值。

基于视觉的手部跟踪，通常使用相机捕捉手部图像，然后利用机器学习和优化算法识别手部特征，达到手部跟踪的目的。单目、双目和深度(RGB-D)摄像机是基于视觉的姿态检测系统中使用的三种主要摄像机类型。基于视觉的方法进行手部跟踪具有无接触、非侵入性，使用方便；计算效率高，许多算法已经实现了实时追踪和识别等优点，现阶段已经有许多商业化设备提供姿态估计，微软的第一代 Kinect 是一个深度相机，于 2010 年 6 月 14 日首次亮相，结合了 SDK 库来监测人类关节的骨骼，作为手势识别研究的基础。Hisham 等人【1】为 Kinect 引入了一个动态的阿拉伯手语识别系统。将决策树与贝叶斯分类器相结合用于手势识别，并采用 AdaBoost 方法对系统进行改进，识别率达到 93.7% 左右。Xi 等人提出一种基于 Kinect 的实时手部跟踪方法【2】，在手部区域提取过程中，使用一种级联结构和递归连接组件算法，提高计算效率并保留三维空间中的连接关系。Frank Weichert 等人对 Leap Motion 的准确度进行评估【3】，在静止情况下 3D 位置的误差小于 0.22mm，在运动情况下误差为 1.2mm。



图 1 Leap Motion 与 Kinect 设备

Li 等【4】提出了一种基于静态 RGB-D 图像进行手势联合定位和识别的高效深度注意技术。该技术根据软注意力策略训练了 CNN 模型，实现了手部的自动定位和手势识别。2019 年，Pinto 等【5】提出了一种基于 CNN 的高效手势识别方法。这些步骤包括分割、多边形逼近、轮廓生成和形态学滤波器的应用，有利于增强特征提取。通过比较所开发的体系结构和其他机制，使用不同的 cnn 进行测试和培训。Hammadi 等【6】根据图卷积网络(GCN)提出了一种有效的手语识别结构。该系统由多个可分离的 3dgcnn 层组成，并在空间注意力策略的支持下进行了改进。该模型利用了约束层，避免了深度学习网络中的基本过平滑问题。此外，注意机制增强了手势的空间上下文表征。廖等人【7】提出了一种使用深度学习机制的多模态动态手语识别技术。最初，手部对象被定位在视频帧上，以最小化网络的空间和时间复杂度。然后，对视频序列进行时空特征提取，利用深度模式和适当确定的手语对视频序列进行分类。基于视觉的手部跟踪研究，都是依靠外部环境的视觉设备进行手部跟踪，这种方式限制了交互主体的交互空间范围，因

此,也有一些学者尝试将视觉设备从环境剥离,实现基于可穿戴视觉设备的手部跟踪。Pail 等人【8】提出了一个称为 HOOV 的手部动作跟踪系统,系统使用佩戴在手腕的惯性传感器,可以持续估计用户手部在 3D 空间中的位置,实现屏幕外交互,平均跟踪误差为 7.7 厘米。使用视觉设备进行手部跟踪的方式尽管已经取得了很高的准确度,但仍存在多种局限性,比如容易受到环境光、遮挡、图像失真等影响,并且用户对摄像头的隐私保护方面表示担忧。

基于无线信号的手部跟踪,通过无线信号的发射和接收时间差测量人手各部分的距离信息,从而实现手部跟踪。Wi-Fi 与雷达由于 Wi-Fi 设备在室内的应用日益普及,基于 Wi-Fi 信号的手势识别技术日益受到关注。Wi-Fi 设备现在已经出现在许多不同的室内环境中。2000 年, Bahl 等人【14】首次将 Wi-Fi 信号用于人体感知,他们提出了一种基于这种信号的接收信号强度的室内定位系统。Pu 等【15】提出了一种无线捕捉系统,使用多普勒频移作为识别 9 种手势的特征,对于识别预设手势的准确率为 94%。Mudra 【16】通过利用不同位置的天线之间的信号差异开发了一种基于 Wi-Fi 的毫米级手势检测方法,准确率为 96%。WiFall 【17】应用随机森林方法和 SVM 分类器对各种人类活动进行分类并实现跌倒检测,平均误报率为 18%,检测准确率为 87%。为了解决信号穿墙传播的问题,Wu 等【18】提出了一种基于 Wi-Fi 信号的被动式人体活动识别系统,不需要任何额外的设备。结合上述研究内容,可以发现,基于无线信号的手部跟踪,侧重于对手指指尖的位置跟踪,在此基础上,可利用手部的正向运动学约束,实现对手部姿态的跟踪。另外,值得注意的是,目前基于无线信号的手部跟踪研究,大多数是针对手部单一手指指尖的跟踪,对于手部多手指或者多部位的跟踪,仍然是一个具有挑战性的研究方向。

基于惯性传感器的手部跟踪,利用了惯性传感器集成的加速度计、陀螺仪和磁力计的测量信息,融合计算传感器的姿态,从而计算手部姿态信息,进而实现手部跟踪。Kortier 等人【19】提出使用安装多个惯性传感器的数据手套获取手部的运动信息,在运动过程中测量所有手指和关节的姿态,从而实现手部的完整三维重建,但是该设备使用 USB 接口传输数据,容易受到环境的限制。之后,清华大学的 Bin Fang 等人【20】深入研究了基于数据手套的手部跟踪和交互技术,利用惯性传感器的测量数据,使用传感器融合算法估计关节的姿态,并利用正向运动学从关节姿态推算指尖位置,从而实现对手部的位姿跟踪。BorShing Lin 等人【21】【23】则基于模块化的数据手套展开了深入的研究,研究了模块化的数据手套系统,利用数据手套上的惯性传感器获取手部各关节的运动信息,再通过融合算法测量用户手部各关节的运动范围,从而跟踪用户的手部关节运动,其模块化的设计使得数据手套的灵活性和可维护性得到提高。HsienTing Chang 等【24】提出一种传感器融合算法,在融合数据手套的加速度、磁场强度和角速度数据基础上,利用反馈思想不断修正手部跟踪的方向,平均测量误差小于 3° 。总的来说,惯

性传感器可用于实现手部位置的跟踪，但会存在空间位置求解时加速度积分导致的误差累计问题，而惯性传感器在实现姿态跟踪方面具备很高的准确度，因此可将惯性传感器用于手部跟踪的研究中。

基于多传感器融合的手部跟踪，往往通过多个传感器获得姿态拟合出最优姿态解，以弥补各自的缺点。现有通过基于 IMU 的数据手套估计手部姿态的方法仍然存在着误差漂移问题和由于碰撞等原因导致的姿态明显不符合手部结构问题，为了提高手部姿态估计的性能，有研究者通过多传感器融合的方式提高姿态估计的精度。Jiang 等人【25】将数据手套与深度相机配合使用，将基于可穿戴的手部姿态估计方法与基于模型的手部姿态估计处理方法相结合，提高了手部姿态估计的稳健性和准确性。Park 等人【26】将佩戴于手部的陀螺仪与佩戴于头部的深度摄像头结合，利用陀螺仪良好的动态性能，弥补了手在高速运动时深度图像失真导致手部姿态估计准确性大大降低的问题。Liu 等人【27】通过建立描述手指间运动动力学的手掌模型，提高了手指跟踪精度。在此基础上，该研究人员结合深度相机 Kinect 对手掌位置进行跟踪，进一步减小了使用传感器进行位置估计的误差。将惯性传感器与其它传感器结合，虽然提高了手部姿态估计的稳定性和准确性，但也增加了用户的负担，提高了功耗和成本。基于多种传感器的手部跟踪。研究人员【28】，通过融合传感器的姿态信息和磁感应强度信息，解算得手部的位姿信息，探索了磁感应手势跟踪的方法。也有【29】等人探究脑电信号与惯性传感器融合的方式进行手势跟踪，利用脑电信号与惯性传感器互补，以得到更优的姿态解。

为了进一步降低姿态求解的误差，同时人体手部约束属于身体属性之一，有许多研究者将生理约束融入到手部姿态估计中，通过针对手部不同的生理学知识进行处理，设计相关手部模型【30】【31】减少姿态误差。Wen 等人【32】提出了融合偏差减少传感器误差的思想，通过对传感器测量误差的建模处理，提高了手部姿态解算的性能，然后将手部约束融合进手部姿态估计模型中，得到符合手部约束的手部姿态信息。Liu 等人【29】通过建立描述手指间运动动力学的手掌模型，结合视觉识别和定位方法，补偿由传感器精确导致的手掌位置估计误差，提高了手指跟踪精度。Li 等人【33】提出了将生理学手部知识融合进三维手模型之中，将图像中的肌肉、骨骼和皮肤与模型通过迭代形状配准和参数学习进行对应匹配，以保证得到的三维手模型的运动范围与姿态符合物理手部结构。手部生理学约束【34】【35】包含关节的运动角度存在的约束关系，手部各关节之间不是相互独立的，并且各关节拥有的自由度不同，某个关节的动作可能会引起其他关节做出相应反应，关节之间存在着约束和依赖关系，同样单个关节的活动范围也是受限的，这些关系导致手部各组成部分密不可分，共同形成了手这一整体，本文基于关节约束信息优化手部姿态估计模型，得到符合物理手部结构的手部姿态。

识别用户的交互意图能使交互系统提供更流畅的交互体验。【39】有意识的人类活动会导致大脑的神经细胞产生有节奏的电信号。放置在头皮区域的电极可用于检测反映

人类意识活动的脑电信号，从而控制假肢的运动。动作意图识别主要分为两个步骤，特征提取和分类识别。提取的特征主要包括时域【43】、频域【44】和时频特征【45】。在识别不同动作的过程中，单个特征的分类效果可能会有所不同。影响动作识别的另一个因素是机器学习算法的选择。目前，常用的识别动作意图的算法包括支持向量机(SVM)、决策树和神经网络。尤里等人【40】开发了一种基于脑电图的非侵入性脑机接口系统。该系统通过脑电数据检测用户的心理意图，平均准确率达到 80%。Zhong 等【46】利用时频信息融合方法获得表面肌电信号的时频特征，并引入深度信念网络来识别上肢康复动作。Ning 等人【47】提取时域和频域的表面肌电信号，并采用优化的支持向量机来识别上肢运动。Hou 等人【48】利用惯性传感器记录被测物体的运动轨迹，并利用长短期记忆来识别身体活动和相关动作单元。Sun 等人【49】利用惯性传感器获取假肢膝关节的姿态数据，并利用 BP 神经网络实现假肢膝关节的自适应速度控制。

根据现有研究，本文结合目前的手部跟踪应用及需求，对手部跟踪和交互技术展开研究，旨在实现更精确的动作捕捉，并根据本文的交互环境探索自然的手部交互方式。

1.3 主要研究内容

本文的研究内容主要分为三个部分：

1) 本研究通过开发手部约束特性，提出一种全手行为跟踪优化算法，降低了全手行为跟踪的误差。首先探究了手部姿态的常见构型和表示方法，建立以四元数为基础的算法模型；然后开发手部约束特性，通过将融合约束对全手行为跟踪算法进行优化，降低手与传感器错位所产生的误差，得到符合生理结构的手部姿态，提高手部姿态估计的准确度。

2) 本研究提出一种用户参与设计的空中交互手势集合，降低用户的学习成本，并给出了基于决策树的实现方法；最后研究人体震颤特征与用户行为交互意图的联系，提取人体运动信号中的震颤特征，根据特征信息进行交互意图区分，帮助计算机更好的理解用户交互行为，实现更流畅的手势交互，并设计用户实验并利用震颤特征识别用户行为中的交互意图，结果表明这种方法具有可行性。

3) 设计并实现了手势交互原型系统 iGestureGlove。在该交互原型系统中设计了一种无线传输数据手套，为基于姿态的空中单手手势交互提供硬件支持；结合全手行为跟踪优化算法，实现了自然的手势交互。

1.4 论文结构

论文的章节结构安排如下：

第一章是绪论部分，从本文的选题背景和研究意义展开叙述，随后针对手部跟踪的国内外研究现状做出简要描述，最后提出本文的主要研究内容，并对论文的总体结构进行介绍。

第二章介绍手部跟踪及交互技术的相关研究，从全手行为追踪技术和手势交互技术

两方面展开。全手行为追踪技术中介绍了基于电磁波、计算机视觉、惯性传感器的全手姿态估计技术；手势交互技术中介绍了构造手势集的研究，分析了基于原始数据的手势识别和基于姿态的手势识别。

第三章主要研究融合手部约束的全手行为跟踪优化方法。首先建立了基于四元数的手部模型，通过进行手势交互误差分析，将人手约束信息作为先验知识作为姿态优化求解算法融合进姿态求解模型中，利用通过将基于人手的关节约束建模为基于四元数的数值约束，得到符合真实人手物理约束条件的手部姿态，减少运动过程中的佩戴误差问题，相较于优化前的方法提高了关节姿态估计的准确度，最后通过实验验证了该优化方法提高了全手行为跟踪的准确率。

第四章研究了一种空中手势交互指令与手势集合，并探讨了一种基于震颤的意图识别方法。首先采用以用户为中心的设计方法设计基于姿态的手势集，通过向用户解释手部姿态的含义，并采用大声思考的方法记录用户的想法，建立一种以用户为中心的手势集合；然后根据该手势集合探索了一种坐标映射方法和手势识别方法，以实现交互过程中光标的长短距离映射以及交互指令映射；最后提出了一种用户交互意图识别方法，通过提取震颤特征对用户行为进行分析，设计实验并验证了该方法的可行性。

第五章面向空中手势交互领域，设计并实现了一个空中手势交互原型系统 iGestureGlove，介绍了系统硬件和软件两部分，最后对原型系统的实现进行进一步的介绍和展示，并介绍了 iGestureGlove 原型系统的应用场景。

第六章主要对全文的研究内容和系统实现做出总结，并展望了今后的研究方向。

2 手部追踪与交互方法相关研究

本章介绍了当前手部跟踪算法的常见实现方式，包括基于磁感应的手部跟踪、基于计算机视觉的手部跟踪、基于惯性传感器的手部跟踪，并分析了三种方法各自的优缺点；此外介绍了交互手势集的构建方法，并分析了基于原始数据层面的手势识别和基于手部姿态的手势识别方法。

2.1 手部追踪相关研究

2.1.1 基于磁感应的手部追踪

基于磁感应的手势追踪是利用佩戴在手指及手腕处的磁传感器捕捉手部运动，获取手势形状、方向以及动作变化信息的一种方法。

首先，用户需要在手部关键部位佩戴几个微型的磁传感模块，这些模块中集成了三轴磁感应传感器、信号放大及滤波电路、微控制器等。模块之间通过数码通讯解决方位标定与时间同步问题。控制端则产生稳定、精准的参考磁场。当手指运动时，会干扰磁场，并由传感器探测磁通密度的微小变化。然后，根据磁传感器输出的电压值变化，可以测算出各模块在参考坐标系下的实时空间位置和方向，这些数据反映了手骨骼的姿态信息。最后，读取不同时间段的姿态数据可以构建动态手势的骨骼运动学模型。

使用磁感应进行手部跟踪，已经得到了广泛的研究，Phil Meier 等【51】提出了通过多传感器共同检测同一目标点的方法，预设传感器已知的姿态偏差和位置偏差，为消除传感器姿态影响提供了一种有效的方法，然而未解决磁场空间姿态的影响，在实验中将其假设为固定值。此外，文献【52-53】研究了通过三轴正交发射线圈和三轴正交接收线圈的磁感应跟踪技术。文献【54】通过仿真方法比较了磁偶极子模型和互感模型在磁感应跟踪中的定位精度。【50】MagIK 系统是一种基于手部运动学模型的磁传感器手势跟踪系统。该模型将手部分割成多个刚体段，每个刚体段均安装有磁传感器以捕捉运动。然后基于优化算法校准传感器参数并提供高精度的手部定位，手指误差小于 10mm。

基于磁感应的手势捕捉方法具备无污染、高灵活性等优势，然而，亦存在一定局限。在其优势方面，磁传感不受光照或遮挡的影响，因此能够稳定地采集骨骼运动数据，支持自由空间交互。此外，其每轴分辨率高达毫米级，且磁传感器结构简单，可实现微小化，方便高密度佩戴，从而能够重建细致的手势。相较于视觉识别，磁感应手势追踪对计算复杂度的要求也较为低。

然而，尽管该方法具备较高的技术成熟度，仍存在几大缺陷。首要的问题是，传感器的佩戴可能会影响手部正常运动，降低用户体验的自然度。其次，网状分布的磁场本身容易受到外部磁源的干扰而发生漂移。此外，标定与信息融合的过程相对较为复杂。最后，传感器的功耗以及采集器的带宽也是限制因素。总体而言，磁感应手势捕捉方法

能够为精细化应用提供高质量的动作数据，但需要解决易用性问题。在进一步发展中，应着重解决传感器佩戴对用户体验的负面影响、外部干扰问题以及提升系统标定与信息融合的效率。

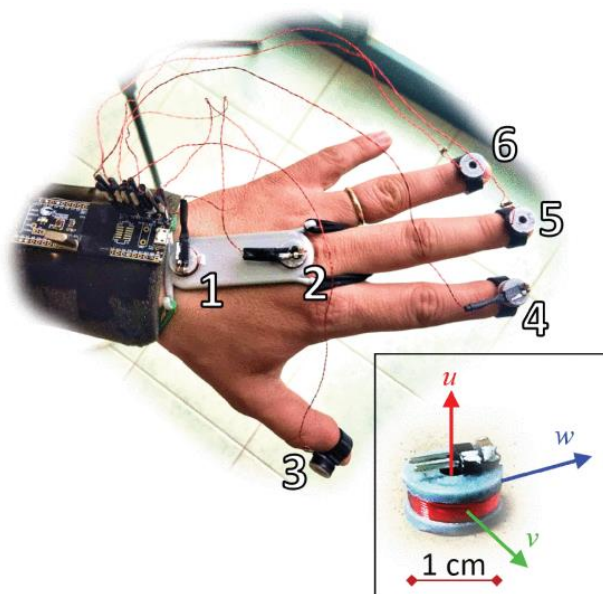


图 2-1 磁感应跟踪系统佩戴布局^[50]

2.1.2 基于计算机视觉的手部追踪

基于计算机视觉的手势追踪技术利用图像处理和机器学习算法捕获和分析手部运动，实现手势追踪。

首先，系统需要采集手部图像或视频作为算法分析的输入源。高质量的图像采集对整个系统识别性能至关重要。通常使用红外摄像头来获得更清晰完整的手部视角。其次，基于视频进行手部检测与分割，提取手部的区域光谱信息。这需要运用肤色检测、背景差分等图像处理技术区分前景手部和背景元素。正确的手部定位是手势特征提取的先决条件。然后，在分割后的手部图像上进行轮廓描边、关键点提取以获取手势形状和姿态的数据。例如，检测手指尖点来判断指向；通过掌心和指骨的轮廓信息识别手掌的方向和形态。这些都是重要的手势特征。基于上一步得到的特征数据，建立对应的机器学习模型以实现手势形状和动作的自动分类与识别。通常需要事先从样本图像中训练模型，让其学习不同手势的特征判别方法。最后，系统会连续跟踪和解析图像帧序列，以输出手势的时间序列信息和解释手势所表示的意思，转换为计算机指令或相应交互。整个过程实现无接触式的人机交流。

基于计算机视觉的手势追踪技术利用摄像头视角信息实现对手部运动的动态捕捉和建模，可广泛应用于无接触式人机交互。这种视觉化手势分析方法兼具独特的优势与不足。基于计算机视觉手势交互最大的特点在于其非接触式的人机交流形式，用户无需佩戴任何电子设备或标记物即可自然控制计算机系统，大大提高了使用的便捷性与友好度。此外，相比数据手套等方法，计算机视觉系统的硬件成本较低，普通摄像头即可实

现高准确度的手势追踪，其算法和模型训练也日趋成熟，可以实现细致精确的手势解析。然而，这类方法也存在一定局限性。首先，复杂的图像处理与机器学习推理对计算性能要求较高，交互过程仍然有较高的延迟。同时外界环境变化如光照条件以及遮挡很容易对系统稳定性产生干扰，个体差异如手型大小也难以通过单一模型兼容，取景范围的限制与占用也为用户空间交互方式带来不便。总体上，计算机视觉技术为手势交互带来革命性进步，但仍有很大优化空间。

2.1.3 基于惯性传感器的手部追踪

基于惯性传感器的手势追踪是一种通过佩戴式设备采集手部运动数据，实现动作捕捉和识别的方法。手势追踪的第一步是选择并佩戴惯性传感器(IMU)。IMU 一般包括陀螺仪、磁力计和加速度计，分别测量手部运动的角速度和线加速度。为捕捉整个手部骨骼的全详细运动，需要在手指、手腕等关键点佩戴多个 IMU，然后通过数据融合算法构建运动学链，获得手骨骼动力学模型。一般结合卡尔曼滤波等优化算法提高时间序列解析效果。最后，系统会输出手势的语义解释，转换为相应的人机交互控制指令。

总的来说，基于 IMU 的手势交互实现了无污染的详细动作捕捉，但需要佩戴多个传感器；算法复杂性也是优化方向。未来可望与计算机视觉或生物力学模型相结合，获得更丰富和精确的手部行为分析。

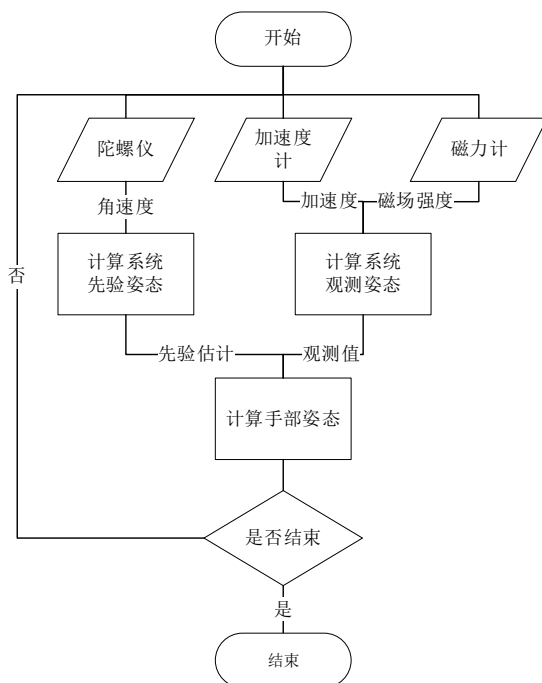


图 2-2 交互流程

基于惯性传感器的手势追踪技术使用佩戴在手或手指关节的 IMU 模块采集肢体运动数据，实现动作捕捉与识别。在优点上，IMU 能提供高保真的运动采样信号，频率可达几百赫兹，精确记录快速复杂手势。而且测量自带各关节的方位信息，便于构建多自由度的骨骼运动学链，计算肢体旋转和位移的参数。基于传感器的数据也不受环境光

影响,对外界条件适应性强。算法上,可获得良好的实时手势分类与回归训练模型。这些都有利于细致、动态的手势理解。

然而,存在几个主要缺陷。首先,模块化部署多个 IMU 造价昂贵,对消费类应用形成障碍。其次,需要用户配戴和配置设备,交互形式不够自然。再者,由于积分漂移误差的存在,长时间使用会逐渐降低测量精度和稳定性。最后,能耗和通信也是需考量的问题。总的来说,IMU 手势交互具备高性能动作追踪的潜力,但还需解决节能与成本控制等问题。

2.2 手势交互相关研究

2.2.1 基于原始数据的手势识别

基于原始数据的手势识别是指直接利用来自传感器的原始信号信息进行手势分类和识别,而不提取结构化的手势描述符特征。具体来说,采集系统会输出包含手部多维信息的动态序列数据,像从 RGB 或深度摄像头获取的图像帧;或者佩戴的电极阵列、超声波雷达接收到的电信号记录。这些数据保留了完整的手势运动细节。然后直接输入神经网络等机器学习模型中,模型会自行学习区分不同手势模式所需的关键特征。通过基于原始数据的方式完成训练,可以获得一个泛化能力好的手势分类或回归系统。

在基于原始数据进行手势识别的方法中,长短时记忆网络(LSTM)表现出对手势时序信息进行建模和识别的良好性能。然而,由于 RGB 图像通常具有较高的分辨率,将其直接作为 LSTM 网络的输入并不可行。通常情况下,研究者采用卷积神经网络(CNN)对 RGB 图像分段进行特征提取,然后将生成的特征序列作为输入传递给 LSTM 网络进行进一步处理。这样的策略在实际应用中旨在有效地捕捉手势的空间和时序特征,提高识别性能。通过这种协同工作的方式,CNN 负责提取静态图像特征,而 LSTM 则专注于处理时序信息,使得整个系统能够更准确地识别和理解手势动作。这种联合使用的方法不仅克服了 RGB 图像输入的挑战,还为手势识别任务提供了更强大的建模能力。Molchanov 等人【51】提出一个端到端的手势识别架构:首先,利用 C3D【52】来提取每个视频片段的特征;然后,将这些特征输入到 RNN 中提取时序特征;最后,将每个 RNN 输出的特征经过 softmax 层变换后输入到 CTC 层中,从而获取每个手势片段的起始位置、结束位置和手势类型。Camgoz 等人【53】也提出了一个类似的架构,区别是,他们利用 CNN 来提取每一帧图片的信息,并且利用的是双流 LSTM 架构。Cui 等人【54】也使用了类似的架构,并且使用了分阶段优化方法以获取更加准确的分类与检测结果。Cao 等人【55】提出了一个第一人称视角下的手势识别数据库,并且提出了一个 3D 神经网络架构。该架构首先将视频分为一些视频片段,然后将每个视频片段使用该网络提取特征,最后将特征输入到一个机器学习模块中,该模块的目的是为了将不同帧中的手部变换到同一个视角下,从而缓解第一人称视角下镜头视角变换的影响,最后利用长短时间序列模型提取每个视频片段之间的时序信息用来分类。

采用基于原始数据的手势识别方式具有显著的优势，主要体现在避免了繁琐的人工特征提取和处理步骤，从而减少了数据加工处理阶段的信息损失，同时保留了手部运动的完整细节，有利于进行差异特征学习。此外，该方法还能够有效实现对多源数据（如视觉、磁信号等）的并行建模，为综合信息的融合提供了便利。然而，需注意到该方法在实践中存在一些限制。

首先，为从原始数据中挖掘特征，该方法对大量样本的需求较高，可能对数据收集和处理带来一定的挑战。其次，该方法对计算机的计算性能要求较高，可能需要强大的计算资源支持，这可能对一些计算能力有限的环境构成一定的局限。同时，由于其较少的特征提取步骤，难以进行对特征的控制变量调整，这使得研究人员在使用该方法时可能面临较为复杂的黑盒问题，难以深入理解系统内部的运作机制。因此，在应用基于原始数据的手势识别方式时，研究人员需要在充分利用其优势的同时，认识到其局限性，并寻求合适的应对策略。

2.2.2 基于手部姿态的手势识别

基于姿态的手势识别是指通过分析手部骨骼框架的位置和运动情况来判断手势，这类方法模拟了人眼分辨手势的视觉注意机制。首先，需要构建手部的骨骼模型，典型的是使用图像或深度信息预测 2D 或 3D 的关键点坐标，比如手腕、指尖和指节的位置。这描述了一个姿态框架。然后在时间轴上追踪框架的变换，获取表达手部配置与运动的高层抽象信息。相较直接应用像素或信号，这种姿态表达的数据量减小了数量级，利于压缩代表互异手势的差异。因此，通常会提取一系列姿态数据特征，包括关键点距离、相对角度以及轨迹形变参数等。这些特征即形成了区分不同手形或动作的依据。最后，得到的结构化姿态描述子会输入分类器或回归器中，对其内在数据模式进行判别或映射，输出对应的手势标签或参数。当前各类深度网络框架都可完成这种从姿态特征到语义空间的映射，实现高效稳健的基于姿态的手势识别与理解。

基于手部姿态的手势识别方法是利用手部关键点的信息来进行手势识别。相比于基于 RGB 图像与视频的手势识别方法，手部姿态不受背景信息的影响，能够更好地关注手部的位置与运动信息，是一种具有较大发展潜力的方法。基于手部姿态的手势识别首先利用手部姿态检测方法获取手部的姿态信息；然后，利用传统特征提取方法或深度学习特征提取方法提取手部姿态的特征；最后，将提取的特征输入到分类器中进行手势分类。Smedt 等人【56】使用 3 个向量来表示手部的运动方向信息、旋转信息和手部的形状信息，使用时间金字塔方法来聚合不同时间尺度上的手部信息，并编码这些特征，最后输入到支持向量机进行训练和分类。Zhao 等人【57】提出了一种基于骨架的动态手势识别方法，该方法提取了 4 种手部形状特征和一种手部方向特征，用时间金字塔表示手部形状和手方向特征，得到最终的特征向量，并将其输入线性支持向量机分类器进行识别。Boulahia 等人【58】从指尖、掌心和手腕的三维坐标中提取出 HIF3D 特征表示手部

形状的高阶信息，并加上时间金字塔来捕获时间信息，最后使用支持向量机对得到的特征进行分类。

基于姿态的手势描述，如关键点坐标序列，数据规模缩小了数量级，有利于训练样本效率。特征表达语义清晰，便于理解与诊断。模块化设计也使工程方法简单可控。但前置的姿态估计与特征选择的人工介入，也可能带来误差累积与信息损失。

2.3 本章小结

本章主要介绍了手部跟踪以及交互技术的相关研究。首先介绍了分别基于磁感应、计算机视觉、惯性传感器的手部追踪方法，分析其各自优点与缺点；其次介绍了基于原始数据的手势识别技术和基于手部姿态的手势识别技术，分析其优缺点。依据上述研究，为探究基于可穿戴设备的全手行为追踪技术和更自然的手势交互技术提供支持。

3 基于姿态的全手行为跟踪

本章根据惯性传感器的姿态求解方法和手部结构约束特点,构造了全手行为跟踪优化方法,降低了手部与佩戴装置错位的求解误差。

本章在佩戴可穿戴设备进行手势交互时,人手大小的差异、个体活动习惯的多样性、运动幅度的不一致性均对姿态估计结果产生影响。为了降低执行手势过程中出现的误差,本章提出了一种手部姿态优化方法,将手部关约束作为先验条件进行手部姿态的优化,建立了全手跟踪优化模型,减少可穿戴设备的佩戴误差问题,相较于优化前的方法提高了关节姿态估计的准确度。在第二章中描述了目前较为常见的手部追踪的方法,有基于视觉的方法、基于惯性传感器的方法等,本研究选择了由惯性传感器组成的可穿戴设备进行全手姿态追踪。在比较不同手部模型和姿态表示方法的基础上,充分权衡了其优劣势,最终选择四元数作为姿态描述的方法,建立手部姿态模型,旨在提高对手部运动和姿态变化的准确捕捉,并为全手行为追踪的研究和应用提供更为可靠的技术支持。

综上所述,本章系统性地阐述了手部姿态的构型和表示方法,通过深入对比不同姿态表示方式,建立基于四元数的姿态表述模型,能够支持多种典型手部构型;此外,本文剖析了可穿戴设备的误差问题,通过将手部关约束纳入全手行为跟踪方法中,提出了一种全手姿态优化方法,该方法能使结果符合手部物理结构,降低了可穿戴设备的错位误差,通过实验证实了该方法的有效性。

3.1 姿态表示与关节特性

手部姿态可以用手掌的方向与各个骨节的方向表示,同时手部的关节也可以用相邻关节的旋转过程表示,这个过程可以用不同的方法来描述,例如欧拉角、旋转矩阵、旋转向量、四元数等。从运动学的角度来看,刚体的运动是刚体在空间中的一个变换过程,将手骨节看作空间中的刚体,同样关节姿态也可以表示成空间中相邻两个刚体的旋转过程。同时手是人类与周围环境交互的重要部分,由复杂的肌肉和骨骼组成,关节受限于骨骼结构而具有运动特性。本小节描述并分析了手部姿态与姿态描述方法,之后分析手部结构,描述手部关节的运动特性与自由度,为将其作为先验知识进行优化提供基础。

3.1.1 姿态描述方法

手部作为一个复杂的关节链系统,其姿态不仅取决于手掌作为刚体的姿态,每个手指节段的相对运动也会影响手部整体的形状。手指的弯曲程度由每个关节的弯曲角决定。因此,要完整描述一个手部姿态,需要手掌姿态参数以及每个手

指关节的弯曲参数。

由于手部关节非常多,包括各个指掌关节(MP)和指间关节(IP),除拇指外的其他手指关节又包括指间近端关节(PIP)和指间远端关节(DIP)两个关节,而且在不同的研究中,手部运动模型也不一致。有研究者提出了一种比较精确的具有 31 个自由度手部运动模型^[1],其中拇指 5 个自由度:掌骨和腕骨相连的关节 3 个、指掌关节(MCP)1 个、手指关节(IP)1 个;余下 4 根手指每根 5 个自由度;掌指关节(MCP)3 个、指间近端关节(PIP)1 个、指间末端关节(DIP)1 个;手的位置和方向 6 个自由度。也有研究人员将手部划为 24 个自由度的运动模型^[1],其自由度包括,拇指 5 个自由度;掌骨和腕骨相连的关节 2 个、指掌关节(MP)2 个、手指关节(IP)1 个;其他 4 根手指每根 4 个自由度;手的方向 3 个自由度。这些复杂的关节结构与运动学链协同工作,使手部可完成精细而复杂的动作。通过组合手掌的旋转角度以及每个手指关节的弯曲角度,可以精确描述复杂的手部姿态,这为分析复杂的手势提供了基础。研究者温^[1]等人根据手指在日常生活中的功能权重占比,分析食指拇指和中指在 Swanson 损伤评估模型中占整手的 72%,提出了一种三指简化模型,并通过关节间约束与手指间约束估计全手姿态。

手部姿态可以通过手相对于参考坐标系的旋转角度以及手指关节的旋转程度来描述。各个骨节和关节共同组成了手这一整体,在多传感器运动捕捉系统中,多个惯性传感器被佩戴在人手不同部位,其中每个惯性传感器可以看成是一个刚体。为了重新构建人体运动,需要合理描述每个刚体的运动状态。这些运动状态通常可以使用欧拉角、旋转矩阵和四元数来表示空间中刚体的姿态。考虑到计算效率及避免万向锁问题,多采用四元数法来描述刚体的旋转运动。四元数是一个四维向量,可以避免欧拉角表示法中的奇异问题,也可以直接转换为方向余弦矩阵进行运算。同时,四元数表示法还具有与旋转运动运算相关的很好的数学特性。

欧拉角作为刚体姿态表示的一种典型方法,其基本思想是将一个刚体的姿态描述为绕坐标系的三个轴依次进行的三个基本旋转,每个旋转由一个角度确定。从数学角度看,三维空间中的旋转属于 $SO(3)$ 群,是一个三参数的李群。而欧拉角给出了一种将其中的旋转用三个角度参数化的方法。从运动学角度可将一个刚体的姿态看成三个基本旋转的组合,每个基本旋转由一个对应旋转矩阵表示。这样欧拉角提供了一个将刚体姿态与三个基本旋转关联的简洁方法。

相比旋转矩阵和四元数,欧拉角更加直观和简洁,符合人的经验认知,易于理解。但是也存在一定缺陷,例如万向锁问题、多个解的存在等。另外,欧拉角表示并不唯一,不同顺序旋转轴所表示的姿态也存在差异。

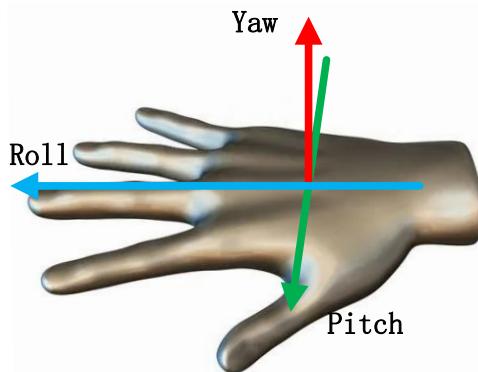


图 3-1 欧拉角旋转示意图

如图 3-1 所示，欧拉角表达直观简洁，仅需要三个角参数就可以确定刚体在三维空间的方向，这符合人们根据日常经验对姿态变化的感知习惯。不同欧拉角定义方式相对独立，每一个角度变化都可以理解为一个基本的坐标轴旋转，这与许多实际机械运动形态类似。例如飞机的驾驶操作即对应欧拉角中的滚转、俯仰和偏航三个变化。欧拉角运算简单，在手工计算和工程实现中都比较方便。基于这些优点，欧拉角在许多工程领域得到大范围运用。欧拉角的主要缺点是在某些特殊姿态(万向锁态)下会产生奇异性，导致表示不唯一和数值问题；与四元数相比，欧拉角表示在进行姿态的时间积分运算时精度较低，数值误差会逐步累积；欧拉角定义方式不唯一，需要按使用场合合理选择坐标轴顺序。

此外，旋转矩阵在刚体运动学中是描述三维空间刚体姿态的重要工具。从数学的角度来看，三维欧氏空间中的旋转变换属于特殊正交群 $SO(3)$ ，其群元素为 3×3 的正交矩阵。具体而言，设空间中的某刚体从初始坐标系 A 经过一个旋转运动 R 到达目标坐标系 B ，则这个旋转运动可用一个 3×3 特殊正交矩阵 R 来表示，其列向量为坐标系 B 的三个基向量在 B 坐标系下的坐标。

旋转矩阵 R 的特殊性质为一个正交矩阵且行列式为 1。这意味着它保持了欧氏空间中向量的长度不变，仅改变向量的方向。即对于刚体上的任一向量 v ，执行坐标变换 $v_b = R \times v_a$ ，向量 v_b 的长度和 v_a 的一致，不会改变。正是由于这一特殊的几何特征，旋转矩阵在描述刚体的运动学时极为方便，可以直接表达刚体姿态的变化、向量在不同坐标系之间的转换等。旋转矩阵完整而直观地表达了三维旋转运动，是刚体运动学分析中一种重要的工具。旋转矩阵的表示中存在着表达冗余的问题。为了避免万向锁，通常会运用其他手段作为对旋转矩阵的补充，从而构建出一个完整的刚体姿态表达方式。

为了避免姿态表示中的万向锁等问题，研究人员开始考虑使用其他参数化方法来表示刚体的姿态。19 世纪，爱尔兰数学家威廉哈密顿提出了四元数的概念。四元数作为复数的推广，由一个实部和三个虚部组成，三个虚部之间正交。与欧

拉角相比，四元数对刚体的姿态建模不存在奇异性问题。另外，单位四元数具有良好的积分性质，可以方便地对刚体的角速度进行积分以得到姿态的变化。这使得四元数十分适合描述刚体在时间段内的运动过程。

四元数最初被应用到计算机图形学中，用于实现三维场景中的旋转变换与插值。随着研究的深入，人们发现四元数在机器人学、导航、控制等领域也有广泛的应用价值。例如在无人机的姿态控制中，四元数可以避免欧拉角的奇异性问题，实现平稳的姿态跟踪。与欧拉角表示法相比，四元数也可以简化相关的运动学方程。总体而言，四元数为刚体姿态建模提供了一种有效的数学工具，它与欧拉角、旋转矩阵等传统方法可以互相转换，也可以根据应用场景需要进行替代。

描述三维空间中两个坐标系的相对姿态时，姿态四元数可以用公式 3-1 表示。

$$q = \begin{bmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{bmatrix} = \begin{bmatrix} \cos(\frac{\theta}{2}) \\ e_x \sin(\frac{\theta}{2}) \\ e_y \sin(\frac{\theta}{2}) \\ e_z \sin(\frac{\theta}{2}) \end{bmatrix} \quad (3-1)$$

在式 3-1 中， θ 为两个坐标系间的相对旋转角， e_x 、 e_y 、 e_z 为单位旋转轴方向向量在两个坐标系中的各个分量。该四元数提供了一个紧凑的刚体运动参数化表示，避免了欧拉角表示中的问题。四元数之间的旋转则用四元数乘法 \otimes 表示，见公式 3-2。

$${}^A_B q = {}^A_G q \otimes {}^G_B q \quad (3-2)$$

对于坐标系 B 旋转到坐标系 A，可以通过借助同一参考系 G 完成，先将 B 转到参考系 G 下，再通过 G 转动到参考系 A，四元数乘法很好的完成了这一过程。对于三维空间中的刚体运动，仅仅表示刚体的瞬时姿态是不够的，还需要描述姿态随时间的变化规律。这需要建立刚体姿态的微分方程模型。与刚体的瞬时姿态参数化表示密切相关的，是刚体的瞬时角速度。刚体的姿态变化可以看作是刚体固连坐标系相对于空间坐标系的转动，而这个转动速度就是刚体的角速度。因此，刚体的姿态变化方程与角速度之间存在着直观的对应关系。

采用四元数表达刚体姿态时，根据刚体角速度构造四元数微分方程的一般形式如公式 3-3。

$$\dot{q} = \frac{1}{2} q \otimes \omega = \begin{bmatrix} -\frac{1}{2}(\omega_x q_1 + \omega_y q_2 + \omega_z q_3) \\ \frac{1}{2}(\omega_x q_0 + \omega_y q_3 - \omega_z q_2) \\ \frac{1}{2}(\omega_y q_0 + \omega_z q_1 - \omega_x q_3) \\ \frac{1}{2}(\omega_z q_0 + \omega_x q_2 - \omega_y q_1) \end{bmatrix} \quad (3-3)$$

在公式 3-3 中, \dot{q} 为单位四元数, 表示刚体瞬时姿态; ω 为刚体坐标系下的角速度三维向量, $\omega = [\omega_x, \omega_y, \omega_z]$, \otimes 表示四元数乘法。上述四元数微分方程直观地反映了刚体运动学之间的内在联系。通过求解这一微分方程组, 即可对刚体的运动轨迹进行解析, 也为基于四元数的姿态估计和控制提供了理论基础。

四元数方法能够获得线性的姿态运动学方程, 直观地表达三维空间旋转, 进行简单的运算, 并方便进行轨迹插值, 这使得四元数方法计算量小且数值稳定, 非常适合用于复杂的手部姿态估计和控制场景。

3.1.2 手部关节特性

对于空间手部姿态估计问题, 关键是建立合适的手部运动学模型, 手部结构包含许多小骨和关节, 具有不同的自由度。但是从运动协调的角度来看, 手部运动可化简为一组基本运动模式, 这使得可以建立一个相对简化的手部模型, 同时保留手部运动的主要特征。具体来说, 人手包含一个球形手腕的结构, 实现手腕三轴运动, 然后是五个手指, 每个手指由多个关节串联而成, 手部关节与指骨示意图如图 3-2 所示。

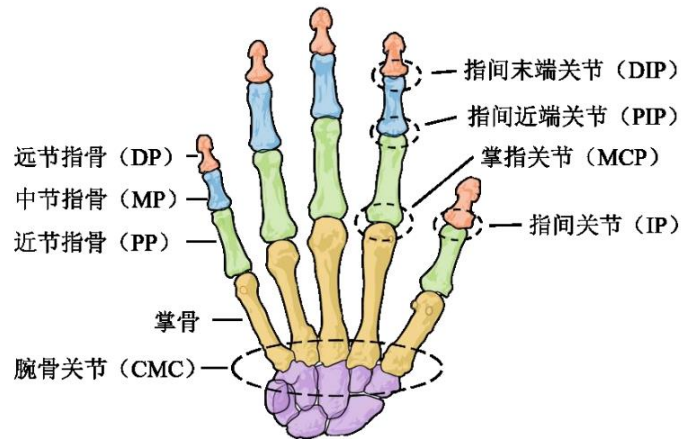


图 3-2 手部关节与指骨示意图^[1]

尽管每个关节都具有一定自由度, 但在许多抓握模式中, 多个关节会协同工作, 形成协同运动模式, 因此可以将每个手指简化为 2 到 3 节刚性物体链, 每一节相当于一个刚性物体, 这种模型化表示降低了模型复杂度。基于以上思路, 研究者们提出了许多手部运动学模型^[1]。这些模型在保留手部主要运动特征的同时, 具有少量参数和简单的运动学方程, 研究表明, 这类模型与复杂手部模型相比, 在大范围手部姿态估计算法中表现相近。因此, 设计合适的简化手部模型, 对空间手部姿态跟踪与预测问题是至关重要的。这种模型化思路也可扩展到其他复杂生物运动结构的分析中, 温豪等人提出了一种简化三指构型的手部模型^[1], 它保留了手在日常生活中 72% 的功能, 具备日常活动以及空中手势交互任务的能力。

Cobos 等人研究了人手各关节的静态运动范围约束,为建立符合生物学特性的手部姿态估计模型奠定了解剖学基础。具体来说,他们通过实验测量了受试者手部在无外力作用下的最大自主运动范围。这种约束反映了人手的骨骼关节、肌肉、肌腱等组织结构的生物力学特性。

表 3-1 手部关节活动范围【引用】

手指	关节	弯曲角度	外展角度
拇指	MCP	[0°~75°]	[0°~5°]
	IP	[-5°~75°]	0°
食指	MCP	[-30°~90°]	[0°~60°]
	PIP	[0°~110°]	0°
	DIP	[-5°~90°]	0°
中指	MCP	[-30°~90°]	[0°~45°]
	PIP	[0°~110°]	0°
	DIP	[-5°~90°]	0°
无名指	MCP	[-30°~90°]	[0°~45°]
	PIP	[0°~120°]	0°
	DIP	[-5°~90°]	0°
小指	MCP	[-30°~90°]	[0°~50°]
	PIP	[0°~135°]	0°
	DIP	[-5°~90°]	0°

通过表 3-1 可以看出,手部包含多个关节,其中可以划分为两类:(1) 指间关节(包括近端指间关节 IP、远端指间关节 PIP 以及指端关节 DIP); (2) 掌指关节(MCP)。指间关节为单自由度关节,仅允许围绕一个轴进行旋转运动,运动模式较为简单。而掌指关节为多自由度关节,能够实现较为复杂的运动。具体而言,掌指关节能够绕手掌横向轴、纵向轴以及与这两轴垂直的轴进行运动。此外,掌指关节之间受到覆盖手部的肌肉的制约,运动不能视为完全独立的。肌肉将多个掌指关节共同约束,产生协同运动。相比指间关节,掌指关节的运动模式更为复杂。

手部约束与人体生物力学特性相关,每个人会有一定的个体差异。但总体上主动约束反映了人类生物结构对手部运动的协调规律。在设计手部姿态估计算法时,需要考虑不同类型的约束,建立能够反映真实手部运动协同规律的模型。加入约束知识,可以提高姿态估计的准确性,也使得预测系统能够生成更符合生物学特性的手部运动。

对于通用的手部姿态跟踪问题,使用这些静态约束的效果可能并不显著。因为在手部日常运动时,通常不会出现大于最大静态范围的运动情况。但在某些特定应用中,这些基本约束为姿态估计系统提供了重要的先验知识。例如,在中风康复患者的机器人辅助手部训练中,由于肌力受损,患者手部只能在有限的静态

范围内活动。此时，加入静态运动范围约束，可以使姿态估计系统生成符合生物学特性的手部运动。这有助于确保中风患者的康复训练不会超出安全范围，避免引起二次伤害。同时理解由人体解剖结构导致的静态运动范围限制，对构建仿生手模型具有重要价值。虽然这些约束不一定通用地改进姿态估计的精度，但它们提供了有用的先验知识，以更准确地模拟生物运动。

综合利用手部约束信息构建手部运动学模型，恰当利用先验的生物学知识，可以使得手部姿态算法在特定应用场景中获得更好的效果，这种生物约束的引入也为人机交互、运动控制等相关领域提供了宝贵的启发。

3.2 全手行为跟踪方法

本节将描述一种手部行为跟踪方法，基于该方法将手掌和手指骨节姿态作为系统状态，构建线性卡尔曼滤波模型，使用角速度更新系统状态，以磁力计与加速度计拟合姿态为观测值，实现对四元数表示的手部姿态的最优估计，进行手部行为追踪。

3.2.1 IMU 姿态求解方法

本小节介绍了一种融合磁力计、陀螺仪、加速度计的 IMU 姿态求解方法，为全手行为跟踪方法提供一种理论支持。单一类型的传感器在进行姿态估计时，由于自身误差的累积常会出现明显的漂移和抖动。这主要源自两方面，一是传感器本身的制造工艺误差。例如 MEMS 惯性传感器中的零偏误差和初始离散误差。这类误差在单个传感器中会不断累积，导致输出迅速偏离真值；二是传感器运行环境带来的干扰。例如温度变化会影响陀螺仪的偏置，磁场干扰则会影响磁强计的准确性。这同样会导致单传感器的输出存在漂移。

因此，研究者针对手部姿态问题广泛采用了传感器融合策略。一种典型方法是在滤波框架下对数据进行优化融合。这种模式综合两种传感器各自的优势，可获得更准确、稳定的手部姿态输出。设计高性能的多传感器融合方案，是提升手部姿态估计准确性的重要途径。传感器的测量值都是基于传感器的姿态给出的，即可以通过数学模型推导得到初始状态到传感器姿态的姿态差异，从大地坐标系到传感器坐标系的用四元数表示的旋转过程。本方法可以通过将分为重力场与地球磁场两个系统分别进行姿态推算，并且为了方便标记传感器坐标系与大地坐标系，称为 L 和 G。

在传感器静止的状态下，加速度计只存在重力加速度，本文中用 ${}^L\mathbf{a}$ ，而大地坐标系的重力加速度是一个常量，且用向量 ${}^G\mathbf{g}$ 表示，满足只在竖直方向上有数值，而其他分量没有加速度存在，为了计算的规范性，将加速度计的测量值与重力加速度都进行标准化处理。同样，磁力计可以测量出传感器所在的磁场数据，本文

用 Lm 表示传感器坐标系下的单位向量, 用 Gh 表示真实的地球磁场中的单位向量, 即 Z 轴竖直向上, 并且 X 轴指向正北方向。

单位四元数可用于操作三维向量的旋转, 本文用四元数 L_Gq 来表示大地坐标系 (G) 中向量与传感器坐标系 (L) 的单位向量的旋转。所以 L_Gq 应满足将传感器坐标系下的单位向量 La 转化为大地坐标系下的单位向量 Gg , 将传感器坐标系下的单位向量 Lm 转化为大地坐标系下的单位向量 Gh 。故写成代数方程的形式如下式 3-4。

$$\begin{cases} R^T({}^L_Gq){}^La = {}^Gg \\ R^T({}^L_Gq){}^Lm = {}^Gh \end{cases} \quad (3-4)$$

通过式 3-4 可以确定 L_Gq 的代数解, 即获得 La 与 Lm 的函数。通过式 1 去获得向量轴介于地球坐标系 G 与传感器坐标系 L 的旋转四元数 ${}^L_Gq_{acc}$, 通过可以得到其余两轴介于地球坐标系 G 与传感器坐标系 L 的旋转四元数 ${}^L_Gq_{mag}$, 为了方便计算, 引入了一个中间过程坐标系 I 以连接两式, 故中间过程坐标系下的向量表示通过 ${}^L_Iq_{acc}$ 矫正后无偏差的 Z 轴, 然后再用 ${}^I_Gq_{mag}$ 对中间过程坐标系的向量的余下两轴进行校准, 故 L_Gq 满足式 3-5。

$${}^L_Gq = {}^L_Iq_{acc} \otimes {}^I_Gq_{mag} \quad (3-5)$$

如此便建立了对于四元数的求解理论部分, 通过重力场进行竖直方向的矫正, 求解出 ${}^L_Iq_{acc}$ 的代数解, 解算出中间坐标系下的向量, 然后利用磁场信息校准其余两轴的方向, 求解出 ${}^I_Gq_{mag}$ 的代数解。接下来讲述具体的计算过程。重力场测定竖直方向并求得 ${}^L_Iq_{acc}$ 代数解 (俯仰角和翻滚角)。因为磁力计较容易收到周围环境的干扰, 加速度计容易受到运动的干扰, 因此结合磁力计与加速度计进行姿态求解的方法有较高的噪声, 得到一个有较大的噪声的传感器姿态。

基于陀螺仪的姿态求解则是通过三轴陀螺仪采集角速度矢量, 由查阅文献知, 刚体的运动服从角速度矢量微分方程式 3-6。

$$q_{t+1} = \phi(\omega, \Delta t) q_t \quad (3-6)$$

在式 3-6 中, q_{t+1} 与 q_t 分别表示在 $t+1$ 时刻和 t 时刻的物体姿态。 $\phi(\omega, \Delta t)$ 则表示对角速度 ω 和采样周期 Δt 进行积分后的状态转移矩阵。通过陀螺仪采集到的连续的数据, 即可推导出连续的姿态四元数。但是由于陀螺仪本身有测量误差, 并且测量误差随着时间的增长, 积分后的误差累计越来越大。在短时间内是稳定连续的, 但是长时间的积分会导致整体误差比较大。

因此有许多研究者进行姿态的计算。互补滤波算法利用陀螺仪、加速度计以及磁力计所具有的互补性能进行多传感器数据融合。该算法通过建立反馈控

制器，将姿态解算过程中的误差反馈进行补偿，以校正陀螺仪积分漂移引起的低频误差。此外，该算法还保留了陀螺仪的高带宽性能，能够对高动态运动进行准确捕捉。通过这种方式，互补滤波算法兼顾了低频和高频信号，满足了精确姿态解算的动态和静态需求。卡尔曼滤波器基于贝叶斯递归估计框架^[1]，是一个连续时间线性系统的最优递归数据处理算法。它利用目标系统的状态方程建立动力学模型，并构建过程噪声和观测噪声的协方差矩阵。通过预测和更新递归步骤，卡尔曼滤波器对系统状态进行估计，并通过反馈调节获得最优解。相较于互补滤波，卡尔曼滤波器充分利用所有传感器信息，通过统计优化获得最佳状态估计。此外，其数学框架的严格性也使滤波过程更加稳定可靠。总体来说，这些多传感器融合算法各有优势，应根据实际需求进行选用与设计。

3.2.2 全手行为跟踪方法

本小节研究了一种全手行为跟踪方法，基于传感器姿态求解方法，根据手部姿态的定义和特性构建卡尔曼滤波系统，其包含系统状态方程与观测方程两部分，观测系统是根据加速度计和磁力计的数据对手部各个骨节姿态进行解算，而系统状态方程则是根据通过陀螺仪采集到的连续的数据，进行二次积分推导骨节姿态。在本方法中，所提到的骨节姿态均基于贴合在骨节上的惯性传感器，得到的全手姿态亦为附着的传感器的姿态，本方法假定传感器姿态即为附着的骨节姿态。

在本文的手部姿态估计方法中，将 8 个手部骨节姿态信息作为系统状态，对三指关节进行约束空间求解，每个骨节姿态以四元数形式表示，系统状态记作 \mathbf{x} ，如公式 3-7 所示。

$$\mathbf{x} = [H \quad T_{FM} \quad T_{PP} \quad T_{DP} \quad F_{PP} \quad F_{MP} \quad M_{PP} \quad M_{MP}] \quad (3-7)$$

在公式 3-7 中， H 表示手的姿态，贴合食指与中指的掌骨； T 、 F 、 M 分别表示手指为拇指、食指和中指； FM 、 PP 、 MP 、 DP 则分别表示拇指的掌骨、近端指骨、中端指骨、远端指骨。

通过相邻指骨姿态的数值运算，可以得到相邻指骨间关节的姿态，对关节姿态进行约束空间下求解，获取系统状态的先验估计与获取观测系统的值。该系统包含及系统方程和观测方程两部分。系统方程使用陀螺仪进行姿态的变化求解，而观测系统则没有累计误差、利用加速度计与磁力计求解得到的四元数姿态。故给出系统的先验估计值和观测系统得到的观测值与系统状态的关系表达式 3-8。

$$\mathbf{x}_t^- = A\mathbf{x}_{t-1} + \boldsymbol{\omega}_t \quad (3-8)$$

在式 3-8 中, x_{t-1} 表示 t-1 时刻的系统的当前姿态, x_t^- 表示将 t 时刻的系统状态作为先验知识, 通过角速度的积分得到系统状态得先验估计。A 表示通过角速度计算得到的状态转移矩阵, 将 t-1 时刻的状态转移到 t 时刻。 ω_t 则由传感器的陀螺仪噪声组成。

观测方程则不随时间增加而累计误差, 根据加速度计和磁力计的数据对传感器姿态进行解算。本文基于加速度计和磁力计的姿态解算通过两次旋转完成, 第一次旋转是通过加速度计的输出解算得到, 相应的四元数描述了旋转过程中的翻滚角和俯仰角信息; 第二次旋转是通过磁力计的输出解算得到, 相应的四元数描述仅包含了传感器的偏航角, 第二次旋转基于第一次旋转之后的坐标系。在得到观测系统的旋转姿态之后需要与当前系统状态进行转换, 观测系统与当前系统状态的状态转化表达式如式 3-9 所示。

$$z_{k+1} = Hx_{k+1} + v_{k+1} \quad (3-9)$$

在式 3-9 中, H 表示系统状态姿态坐标系与观测系统姿态坐标系之间的转换矩阵, 在本文中为单位阵, z_{k+1} 表示 k+1 时刻下观测系统所得的姿态值, 与系统状态组成相同。观测系统通过加速度计计算传感器坐标系的向量和大地坐标系下的向量的差异, 然后通过磁力计测得的传感器坐标系下的向量, 使得其与大地坐标系下向量方向一致, 然后求解方程组计算两种坐标系下偏航角的差异, 并且结合两个四元数, 求解得到了传感器坐标系与大地坐标系下的旋转过程, 亦可以表示传感器坐标系与大地坐标系的姿态差异, 也即为所贴合指骨的姿态。

在得到系统观测姿态值与系统先验姿态值后, 分别计算其对应的误差协方差矩阵^[1], 系统先验误差协方差增益可以将陀螺仪误差转换得到, 观测系统的误差协方差矩阵则根据磁力计与加速度计的误差通过泰勒的一阶展开式给出, 根据其对应的误差协方差矩阵, 计算并更新卡尔曼增益 K, 进而将观测系统计算所得手部姿态与系统先验手部姿态融合求解, 得到最优手部姿态值。

$$x_{k+1} = x_{k+1}^- + K(z_{k+1} - x_{k+1}^-) \quad (3-10)$$

在式 3-10 中, x_{k+1} 表示当前系统 k+1 时刻下求解所得最优手部姿态值, x_{k+1}^- 表示在 k+1 时刻下手部姿态的先验估计值, K 表示通过误差协方差矩阵获得的卡尔曼增益, z_{k+1} 表示 k+1 时刻下观测系统所得的姿态值。在利用卡尔曼增益完成系统状态值的更新之后, 进一步对系统状态估计的误差协方差进行更新, 达到动态校正系统状态值的目的。

3.3 全手跟踪优化方法

本小节首先分析了全手行为跟踪的误差, 结合人体手部特性提出了一种融合人手生理约束的全手跟踪优化方法。手部姿态估计经常需要考虑生理学上的

手部关约束，大量解剖学实验表明，由于骨骼、肌肉和软组织的生理结构限制，手部运动存在客观的生物力学约束。而将这些约束关系纳入全手行为跟踪的方法中^[1]，则可以得到更加符合人体构造的手部姿态。本研究提出的全手跟踪优化方法基于这一思路，使用多组惯性传感器来捕捉手部运动，并结合 3.1 节阐述的生理学约束关系进行数据融合，以提高姿态估计的准确性。通过实验发现，由于相邻的惯性传感器无法完美贴合骨节并佩戴位置容易受影响，在计算关节姿态时存在误差。为了降低这一误差，本方法将生理学约束引入传感器数据融合的过程，使最终估计结果更贴近真实手部运动。通过这种生理学约束的融合方法，本研究所描述的手部虚拟姿态可以有效反映真实手部的运动状态，全手行为跟踪优化算法结构如图 3-3 所示。

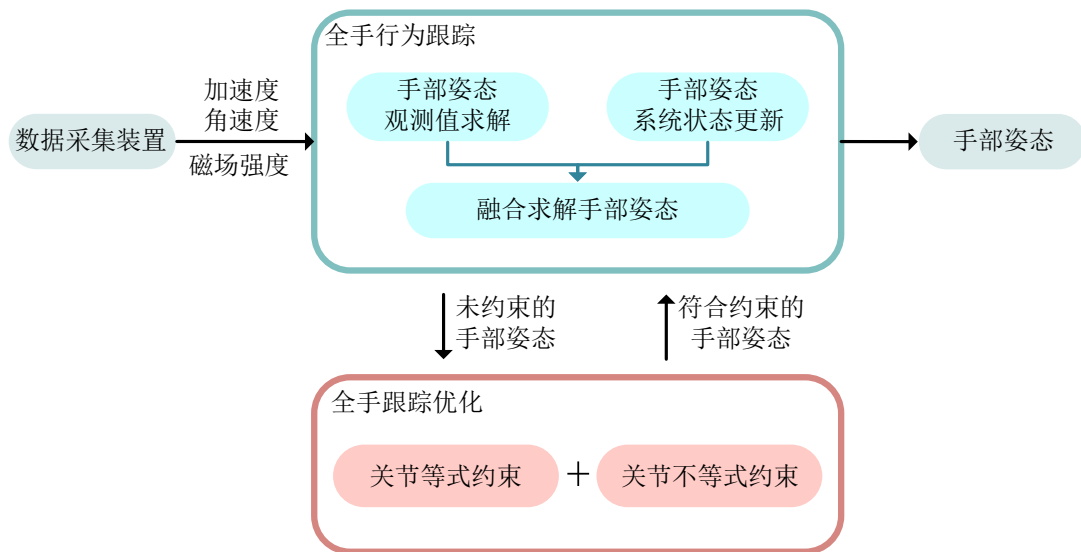


图 3-3 全手行为跟踪优化算法结构图

因此本节将介绍交互过程中全手追踪的误差，并引入关节约束特性对 3.2 小节中的行为跟踪方法进行优化，从而降低运动过程中的误差问题。

3.3.1 全手行为追踪的误差分析

全手行为追踪中的误差可以分为系统误差、环境误差、运动过程误差。姿态估计系统的误差，传感器和姿态估计系统本身具有一部分误差，由估计方法的误差和传感器误差组成，有研究者^[1]通过加入零均值高斯白噪声和融合偏差的方式降低这一部分误差。环境误差通常是由空间中磁场不一致导致的，磁力计对环境磁场非常敏感，任何磁场变化都会对姿态估计产生影响，干扰磁力计的读数。研究人员会通过事先测量环境磁场分布，建立磁场映射，实时校正磁力计测量。

运动过程误差则是发生在手部运动时产生的，它对于不同的佩戴方式、手套材质以及传感器分布有着不同的差异。数据手套不像人体皮肤和肌肉富有弹力和

韧性，往往在进行手部动作时受限于材质的延展性，骨节在带动传感器运动时，传感器受到手套的约束，与真实手部姿态产生误差，在执行大范围的移动时，误差更为明显。数据手套利用惯性传感器与骨节的绑定记录手指运动，为手部姿态估计提供关键的数据，而惯性传感器(IMU)通常固定在手指骨节表面，当手指进行活动时，手套与皮肤之间的相对滑动，皮肤与骨骼之间相对滑动，会使 IMU 的位置相对于骨骼发生变化。这种相对错位现象在手指进行大幅度弯曲运动时尤为明显，佩戴错位示意图如图 3-4 所示。

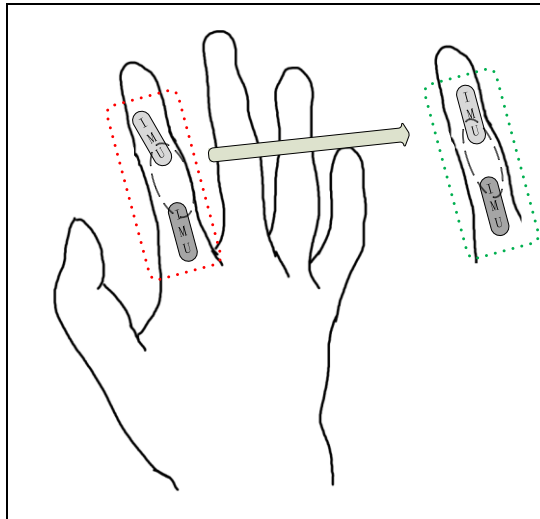


图 3-4 传感器佩戴差异示意图

在图 3-4 中，左侧为佩戴错位图，右侧为正确固定时示意图。同时，手指之间的直接接触会对姿态估计产生一定干扰，具体来说，当用户同时弯曲多个相邻手指时，这些手指的末端可能发生相互接触。因手套表面绑定有测量各手指关节弯曲角度的弯曲传感器，手指之间的直接接触会对相应的弯曲传感器产生额外的机械力，导致关节角度测量结果出现错误。例如，当食指和中指同时并拢弯曲时，两指的接触会相互施加额外力矩，使惯性传感器测量的中指和食指的弯曲角度比实际弯曲角度要小。这种手指直接接触引起的弯曲角度测量误差会导致最终手指姿态跟踪的误差。此外，手指之间的密集接触还可能导致手套表面的惯性传感器发生错位，这也会对后续的姿态解析产生不利影响。

在运动测量中，惯性传感器的佩戴位置误差也是影响测量精度的关键因素之一。这主要是由于人体手部关节运动的复杂性所导致的。具体而言，执行手部运动时，指尖经常随着手指侧转和扭转而运动，使佩戴在手指上的惯性传感器位置发生漂移。这种运动过程中导致的传感器与骨节错位问题，会影响手部姿态的估计，带来较大的误差。在上述过程中，如果佩戴方式没有良好的纠正，则这类误差在用户执行手势过程中一直存在。针对这类问题，本研究将关节的特性引入到手部追踪系统中，降低运动过程误差。

3.3.2 构建手部非线性约束模型

为了更好地将手部约束融合到系统模型，结合手部约束特点以及融合方法需要，本小节将 3.2 节中的手部关节角度约束转化为基于四元数姿态的等式与不等式约束，并融合进手部跟踪模型中。本文将手部关节分为单自由度关节与多自由度关节，单自由度关节包括拇指 IP 关节与其余手指的 DIP、PIP 关节，该类关节只能进行弯曲的动作，只有一个自由度。多自由度关节包括各个手指的 MCP 关节，该类型关节不仅可以进行弯曲动作，还可以进行外展/内收的伸展动作。本节针对单自由度关节和多自由度关节分别建模，将关节执行动作角度的范围约束，转化为针对本文系统状态约束，进而进行数值求解。

本文将手部约束作为先验的知识融合到卡尔曼滤波系统的状态方程中，求解得到符合手部约束先验姿态值。卡尔曼滤波系统有系统状态方程与观测方程两部分，观测系统下的手部姿态是根据加速度计和磁力计的数据对传感器姿态进行解算，而系统状态方程则是根据通过陀螺仪采集到的连续的数据，进行二次积分推导姿态四元数，测量误差随着时间的增长，积分后的误差累计越来越大，在本文中，结合手部约束可以减少这种误差，融合关约束的系统状态方程可以描述为如式 3-11 所示。

$$\mathbf{x}_{k+1}^- = \mathbf{A}\tilde{\mathbf{x}}_k + \boldsymbol{\omega} \quad (3-11)$$

在式 3-11 中， \mathbf{x}_{k+1}^- 表示在 $k+1$ 时刻下手部姿态的先验估计值， \mathbf{A} 则表示由角速度积分得到的状态转移矩阵， $\boldsymbol{\omega}$ 表示符合高斯分布的系统过程噪声， $\tilde{\mathbf{x}}_k$ 则表示在 k 时刻下满足手部约束的手部姿态值，也是上一时刻的符合生理学约束的系统状态。

由于本文的系统状态为手部骨节姿态，因此需要通过相邻手部骨节姿态来获取该关节的姿态，假设手部上的两个相邻指骨为 L1 和 L2，两指骨所形成的关节则需要通过两骨节运算表示。

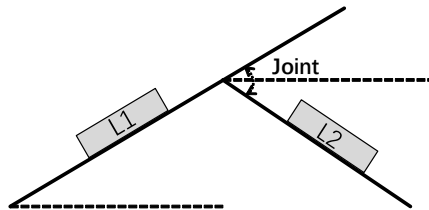


图 3-5 通过相邻骨节求解关节示意图

在图 3-5 中，*Joint* 关节可以由骨节 L1 基于骨节 L2 的旋转过程来表示，即为将 L1 转换为 L2 的旋转过程为该关节的姿态，两者的姿态都是通过大地坐标系所表示，例如 L1 的姿态则是从大地坐标系旋转到传感器坐标系 L1 的旋转过程，所以将大地坐标系作为桥接坐标系，即 L1 的逆旋转表示为传感器坐标系 L1 旋

转到大地坐标系的旋转过程，在此状态下再进行大地坐标系到传感器坐标系 L2 的旋转，即得到从传感器坐标系 L1 旋转到传感器坐标系 L2 的旋转过程，也即为骨节 L1 基于骨节 L2 之间所形成的关节，如公式 3-12 所示。

$$q_{Joint} = X_{L2} \otimes (X_{L1})^{-1} = \begin{bmatrix} h_1(X_{L1}, X_{L2}) \\ h_2(X_{L1}, X_{L2}) \\ h_3(X_{L1}, X_{L2}) \\ h_4(X_{L1}, X_{L2}) \end{bmatrix} \quad (3-12)$$

在公式 3-12 中， h_i 代表姿态四元数系数的函数， i 对应该关节姿态四元数的系数编号， X_{L1} 表示 L1 骨节所对应的系统状态。上述过程得到系统估计所得的关节姿态，要对关节姿态进行约束，则需要得到每个关节姿态的约束范围，也即将针对关节的角度约束，转化为针对关节的姿态约束。

本文将关节约束分为了单自由度关节约束与多自由度关节约束，单自由度关节约束主要特征就是只存在一个方向上的转动，在单自由度关节约束建模中，假定该关节的弯曲角度为 θ ，对应姿态绕 x 轴进行旋转，则该关节的姿态如公式 3-13 所示。

$$q_{Joint} = \begin{bmatrix} \cos\left(\frac{\theta}{2}\right) \\ \sin\left(\frac{\theta}{2}\right) \\ 0 \\ 0 \end{bmatrix} \quad (3-13)$$

通过公式 3-13 可以得到弯曲角度为 θ 的四元数姿态，假定该角度的取值范围为 $[\theta_0, \theta_1]$ ，则通过简单求导即可得到各个值的最大值与最小值范围，结合公式 6 可以即可得到姿态的范围约束结果，如公式 3-14 所示。

$$q_{joint} = \begin{bmatrix} \min\left(\cos\left(\frac{\theta}{2}\right)\right) \leq h_1(X_{L1}, X_{L2}) \leq \max\left(\cos\left(\frac{\theta}{2}\right)\right) \\ \min\left(\sin\left(\frac{\theta}{2}\right)\right) \leq h_2(X_{L1}, X_{L2}) \leq \max\left(\sin\left(\frac{\theta}{2}\right)\right) \\ h_3(X_{L1}, X_{L2}) = 0 \\ h_4(X_{L1}, X_{L2}) = 0 \end{bmatrix} \quad (3-14)$$

在公式 3-14 中， $h_1(X_{L1}, X_{L2})$ 和 $h_2(X_{L1}, X_{L2})$ 为具有最大值与最小值的不等式约束， $h_3(X_{L1}, X_{L2})$ 和 $h_4(X_{L1}, X_{L2})$ 则为固定值的等式约束，因此可以将这个拆分成等式约束和不等式约束两部分。结合 3.2 小结内容，通过将表 3-1 中的各个角度约束范围代入公式 8 和公式 10 中，得到针对关节四元数的值约束，如表 3-2 所示。

表 3-2 手部关节约束转化为数值约束的一般形式

手指	关节	关节姿态约束表达式
拇指	MCP	$0 \leq q_1, 0 \leq q_2 < 0.6088,$ $-0.0133 < q_3 < 0.0133, -0.0218 \leq q_4 < 0.0218$

	IP	$0.7933 \leq q_1 \leq 1, -0.0436 < q_2 < 0.6088, q_3 = q_4 = 0$
食指		$0 \leq q_1, -0.2588 < q_2 < 0.7071,$
	MCP	$-0.183 < q_3 < 0.183, -0.2588 \leq q_4 < 0.2588$
	PIP	$0 \leq q_1, 0 \leq q_2 \leq 0.8192, q_3 = q_4 = 0$
	DIP	$0 \leq q_1, -0.0436 < q_2 < 0.7071, q_3 = q_4 = 0$
中指		$0 \leq q_1, -0.2588 < q_2 < 0.7071,$
	MCP	$-0.138 < q_3 < 0.138, -0.1951 \leq q_4 < 0.1951$
	PIP	$0 \leq q_1, 0 \leq q_2 \leq 0.8192, q_3 = q_4 = 0$
	DIP	$0 \leq q_1, -0.0436 < q_2 < 0.7071, q_3 = q_4 = 0$
无名指		$0 \leq q_1, -0.2588 < q_2 < 0.7071,$
	MCP	$-0.138 < q_3 < 0.138, -0.1951 \leq q_4 < 0.1951$
	PIP	$0 \leq q_1, 0 \leq q_2 < 0.866, q_3 = q_4 = 0$
	DIP	$0 \leq q_1, -0.0436 < q_2 < 0.7071, q_3 = q_4 = 0$
小指		$0 \leq q_1, -0.2588 < q_2 < 0.7071,$
	MCP	$-0.1531 < q_3 < 0.1531, -0.2164 \leq q_4 < 0.2164$
	PIP	$0 \leq q_1, 0 \leq q_2 < 0.9239, q_3 = q_4 = 0$
	DIP	$0 \leq q_1, -0.0436 < q_2 < 0.7071, q_3 = q_4 = 0$

在表 3-2 中给出了手部关节约束转化为关节姿态四元数的数值约束结果，其中 q_i 对应公式 3-14 中的 $h_i(X_{L1}, X_{L2})$ ，也即为系统状态的非线性约束，在得到如表中所示的等式约束集合和不等式约束集合之后，将其代入式 3-11 中，之后利用 Active-set 方法进行约束求解，即可得到符合生理约束的手部姿态。

本文所用卡尔曼模型系统状态为各惯性传感器的姿态，生理学上的关节角约束无法直接作用在系统状态值上求解，因此将相邻骨节姿态差异视作手指两骨节所形成关节的姿态，将手部的关节角约束转化为基于四元数的数值等式与不等式约束集合后，得到关于基于骨节姿态值的等式与不等式函数集合，利用投影法将针对系统状态的约束集合融入卡尔曼滤波系统中，将关节姿态的约束问题转化为系统状态的约束集合求解问题，在约束求解中，本文利用 Active-set 方法约束求解，最终得到符合约束的系统状态。

3.3.3 非线性约束下的姿态求解

本文将关节约束融入到卡尔曼滤波系统的状态方程中，针对状态方程中约束问题求解，需要定义最小化约束的目标函数，目标函数用于衡量无约束状态下手部骨节姿态与符合约束的手部关节范围的接近程度，目标函数的设计可以使用欧氏距离作为目标函数的度量方式。对于每个手部骨节计算无约束状态下的姿态与约束范围之间的欧氏距离，将所有关节的距离进行加权求和，得到一个总体距离值作为目标函数的输出，用于度量无约束状态与约束面上的点之间的差异，目的是最小化这个差异，以找到与无约束状态最接近且符合手部关节范围约束的姿态，用 \hat{x} 表示。对于 \hat{x} 的求解运算，具体可以描述为如式 3-15 不等

式约束优化问题。

$$\begin{cases} \tilde{\mathbf{x}} = \operatorname{argmin}_{\tilde{\mathbf{x}}} (\tilde{\mathbf{x}} - \mathbf{x})^T (\tilde{\mathbf{x}} - \mathbf{x}) \\ h(\tilde{\mathbf{x}}) = 0 \\ g(\tilde{\mathbf{x}}) < 0 \end{cases} \quad (3-15)$$

在式 3-15 中, $\tilde{\mathbf{x}}$ 表示符合约束方程组的手部姿态最优值, 也即满足约束关系的系统状态, \mathbf{x} 则表示无约束状态下的系统手部姿态, 求得符合约束并与当前系统状态欧氏距离最短的手部姿态, $h(\tilde{\mathbf{x}})$ 表示关节姿态四元数满足的等式约束集合, $g(\tilde{\mathbf{x}})$ 表示关节姿态需要满足的不等式约束集合。针对求解目标函数的最小化问题, 有许多不同的最优化算法, 包括梯度下降、共轭梯度法、牛顿法等。选择合适的算法取决于问题的特点和约束条件的复杂性, 对于简单的约束条件, 如角度限制或线性约束, 可以使用基于梯度的方法进行求解, 但对于更复杂的约束条件, 如非线性约束则可以使用其他的优化算法, 如牛顿法、Active-set 等算法进行求解。对于含有非线性约束的最优化问题, 除了线性规划算法外, 还可以采用其他数学优化方法求解。具体而言, 牛顿法是一种典型的非线性规划算法, 它利用二阶泰勒展开逼近目标函数和约束函数, 将原问题转化为一系列二次规划子问题。通过迭代求解子问题, 牛顿法能够快速收敛到条件的最优解。Active-set 算法则是按约束的工作状态将其划分为工作集与非工作集, 以确定迭代搜索的方向。总之, 针对具有复杂非线性约束的问题, 采用专门非线性规划算法能够获得更高的计算效率和更精确的最优解。

本文综合考虑采用 Active-set 方法进行约束求解, 它是一种常用的数值优化方法, 该方法通过迭代地将约束条件中的一部分视为等式约束, 并通过求解相应的等式约束子问题来逼近原始非线性约束问题的最优解。根据上述不等式约束优化问题可知, 在求解系统状态最优解时, 需要将系统状态的约束按照等式约束和不等式约束进行分类预处理, 具体分类步骤以及结果已经在上小节中给出, 约束求解流程图如图 3-6 所示。

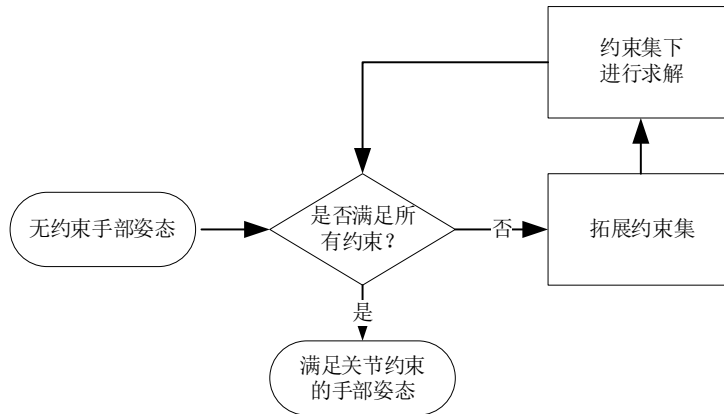


图 3-6 进行手部姿态约束求解的流程简图

在图 3-6 中, 将当前无约束的手部姿态 \mathbf{x} 作为初始状态, 初始状态下, 判断当前无约束手部姿态是否符合所有的手部约束, 如果满足所有的约束信息, 则不需要进行约束求解, 若存在手部姿态不满足的约束条件, 则在每个迭代步骤中, 将系统状态不符合的约束条件添加到约束集合中, 并将其转化为等式约束, 就可以得到一个等式约束子问题, 其中目标函数与原始问题相同, 约束条件仅包含等式约束。通过求解等式约束子问题, 可以得到一组候选解, 表示在无约束情况下的手部关节姿态。在获得等式约束子问题的解之后, 需要检查候选解是否满足所有的关节姿态约束。如果候选解违反了某些约束条件, 将其添加到约束集中。重复执行上述步骤, 直到候选解满足所有约束条件且约束集不再变化或者达到最大迭代步数为止, 这表明已经找到了满足手部关节姿态约束条件的最优解 \mathbf{x}^* 。

3.3.4 实验分析

本实验为手部姿态误差计算与对比实验, 目的在于探究优化方法估计手部姿态的准确度, 探究与未优化方法相比是否降低了可穿戴设备错位误差。鉴于使用角度尺以及组合测量装置作为测量手指关节角度的测量装置时需要研究人员记录角度值, 研究人员的操作会导致姿态的变化以及读数的不稳定, 而光学设备在环境光线充足条件下具有较高的准确度, 因此本实验使用商业手部姿态捕捉装置 Leap Motion 进行手部姿态捕捉, 并将其作为手部的参考姿态。

1) 参与者与实验环境

本次实验招募了 12 名参与人员其中男性 9 名, 女性 3 名, 参与者年龄分布在 20 岁至 25 岁之间($M = 22.08, SD = 2.11$), 惯用手均为右手。

实验所用计算机的显示器为 27 寸 2560*1440 分辨率戴尔品牌, 交互设备为包含惯性传感器和信号处理单元的数据手套, 测量装置为 Leap Motion 运动捕捉装置。

2) 实验任务与实验设计

在本实验中, 参与者需要根据提示做出不同的手势进行数据采集和对比。实验将伸展手势(Open)、握拳手势(Close)和 OK 手势(Ok)作为本次实验的三种手势, 手势示意图如图 5 所示。本次实验要求参与者右手佩戴数据手套依次执行伸展手势、握拳手势和 OK 手势, 每种手势执行 10 轮, 每次在抓握姿态稳定后采集 10 秒钟并保存数据。由于 Leap Motion 装置通过视觉的方式进行姿态捕捉, 因此需要本实验环境保持光源稳定, 执行动作时用户运动范围一致, 掌心保持面向镜头, 并且在执行手势时小臂与腕关节保持稳定。

选择这三种手势是由于伸展手势和握拳手势往往是人手关节日常运动的最大与最小范围，使用这两种手势能够体现出姿态估计方法在边缘范围下的准确性，而 OK 手势则是日常生活中常用的具有语义的手势，其表现为食指与拇指自然弯曲，其余三指自然伸展，选取该手势能够体现出用户执行有语义手势时的准确程度。

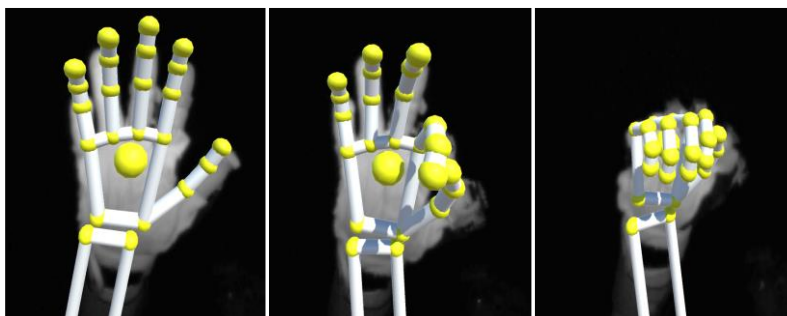


图 3-7 伸展手势(Open)、OK 手势(Ok)和握拳手势(Close)示意图

实验变量：（1）姿态估计方法：优化的手部跟踪方法、未优化的手部跟踪方法。（2）手势：伸展手势、握拳手势和 OK 手势。

实验类型：within-subject

在通过实验得到用户执行任务过程中的数据后，对数据进行预处理操作，例如 Leap Motion 在进行手势捕捉时出现了误判左右手、数据采集过程中由于佩戴时的剧烈运动导致传感器桥接处断连等情况，本文根据实验时的标注以及对结果的分析，采用插值法对误判左右手的数据进行替换。将 Leap Motion 的手部关节姿态值作为手部姿态参考值，利用传感器数据分别通过优化后的方法和未优化的方法进行姿态估计，将两种方法所得到的姿态值的坐标系进行转化到与 Leap Motion 所得姿态的同一个坐标系下，再进行逆旋转运算，得到姿态的误差。

4) 实验过程

实验开始前要求参与者坐在电脑显示器前，在实验人员的帮助下佩戴好数据手套并熟悉了解任务流程，帮助用户调整手臂合适的高度，参与者自然执行伸展、握拳和 ok 手势，调整坐姿注视显示屏幕，此过程直到参与者能够表示熟练手势动作后结束；参与者的初始状态为手掌自然张开，当参与者听到执行手势时，执行相应的行为，系统开始采集参与者佩戴的数据手套数据以及 Leap Motion 的姿态数据，当数据采集 10 秒钟时研究人员告知用户本轮结束，每一轮间隔 30 秒，此过程一直重复直到实验完毕。

5) 实验结果

通过实验，本研究获得了用户三种手势下的传感器数据和 Leap Motion 的姿态数据，首先本研究将传感器数据通过未优化手部跟踪算法和优化后手部跟

踪算法得到两种估计姿态，之后将 Leap Motion 姿态和两种估计姿态的坐标系统一化，并求两种方法的误差。

本文展示食指 PIP 在三种手势下优化后方法与未优化方法的误差对比结果，如图 3-8 所示。在图 6 中由上而下分别为误差姿态四元数的值， q_1 越接近 1 表示误差越小，其余四元数值越接近 0 误差越小。



图 3-8 食指 PIP 在三种手势下的姿态误差

通过图 3-8 发现，融合约束方法的对于食指 PIP 关节姿态求解优于未融合约束方法姿态，为此本文针对各个参数进行了单因素方差检验(ANOVA)，判断两种方法的姿态误差是否具有显著差异。进行参数检验时样本需要满足正态分布，针对两种方法下各个关节姿态误差进行正态分布检验，本文使用对齐秩变换(ART)^{【***】}对不符合正态分布的数据进行转换处理。结果显示食指 PIP 的参数 q_3 ($F_{1,33} = 396.4$, $p < 0.001$, $\eta^2 = 0.46$) 和 q_4 ($F_{1,33} = 396.3$, $p < 0.001$, $\eta^2 = 0.46$) 存在显著差异，即两种方法对误差的影响显著，在食指 PIP 上偏航角的估计中融合约束方法误差 ($M = 0.25$, $SD = 0.32$) 低于无约束方法误差 ($M = 0.26$, $SD = 0.37$)，对于翻滚角的估计中融合约束方法误差 ($M = 0.25$, $SD = 0.32$) 低于无约束方法误差 ($M = 0.26$, $SD = 0.37$)。而姿态参数 q_1 的误差均值在两种方法之间并未表现出显著差异 ($F_{1,33} = 1.724$, $p = 0.1936$)，而姿态参数 q_2 的误差均值在两种方法之间并未表现出显著差异 ($F_{1,33} = 0.065$, $p = 0.7989$)。

该结果表明通过引入手部约束对手指弯曲方向的变化影响并没有表现出显著差异，但降低了佩戴错位问题导致的其他旋转轴的误差。

为了进一步分析误差离散程度与对比，绘制姿态误差箱线图 3-9。在图 3-9 中的优化方法与未优化方法对比图可以发现，优化方法对姿态系数 q_3 、 q_4 的误差均值小于未优化方法的结果，优化后的方法的分布更加集中，相较于未优化方法的误差更接近 0 值，根据关节特性可知，PIP 只具有弯曲方向的自由度，即其姿态四元数参数 q_3 、 q_4 的值为零值，然而运动引起的错位误差，则会给估计所得姿态引入其他方向分量的误差，使得传感器与骨节之间发生偏转，导致姿态出现误差。在全手姿态跟踪中，引入生理学约束降低了佩戴错位误差导致的姿态估计误差，优化方法相较于未优化的全手行为跟踪方法，能够获得更真实的手部姿态。

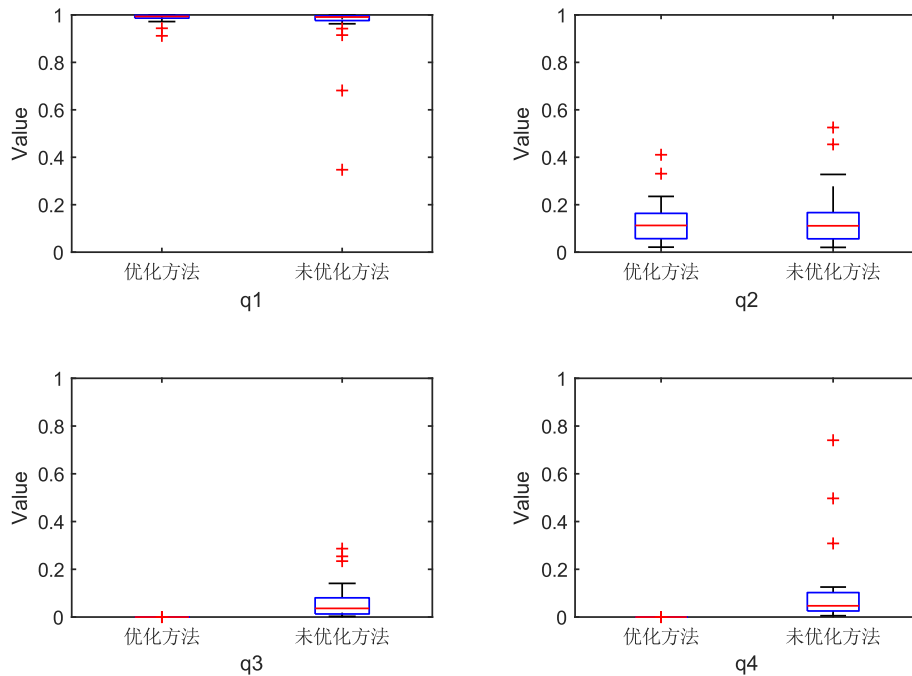


图 3-9 食指 PIP 姿态误差均值箱线图

本文设计的手部姿态测量实验，目的在于探究优化后方法的准确度是否优于未优化方法。优化后方法在求解 PIP 姿态尤其在参数 q_3 、 q_4 的求解上表现出更好的准确性，误差显著低于未优化方法。手指 PIP 只有一个自由度，即只可以进行弯曲动作，不能进行旋转等其他动作，姿态四元数参数 q_3 、 q_4 值为 0，能够真实的反应出手部的姿态信息。综上所述，融合约束方法在手部姿态求解中优于未优化方法，能较好地降低由于运动错位导致的估计误差问题，在实际应用中，融合约束方法可以在手势交互领域作为一种更优的选择。

3.4 本章小结

本章提出了一种全手行为跟踪优化方法,其通过佩戴数据手套进行手部姿态估计,通过实验表明,结合约束的全手跟踪优化方法相较于未优化方法,在 PIP 姿态四元数两个分量的误差具有显著差异。与现有方法相比,不需要使用额外的模态信息,令手部姿态符合真实手部结构,降低了可穿戴设备中佩戴错位的问题,这为接下来的基于姿态的手势交互技术提供了稳定的手部姿态。

4 基于姿态的手势交互技术

本章旨在探索并设计以用户为中心的自然手势交互方式，包括用户参与的交互手势集设计、构建映射模型和基于手部震颤特征的交互意图识别三个方面。现有研究的手势集不能满足本节基于姿态的平面手势交互需求且研究者提出的手势集合可能存在偏见问题，因此本章首先基于以用户为中心的设计方法提出手势集，向用户解释手部姿态的含义，采用大声思考的方法记录用户的想法，最终建立基于姿态的交互手势集合。

针对用户提出的交互手势集合，本章探索了一种坐标映射方法和手势识别方法。根据用户参与式的手势设计实验中，用户提出当使用手掌的偏航转动控制光标时转动角度范围小，无法进行长距离移动的问题，提出长短距映射模型实现光标的移动控制，发掘手势集合手指关节功能，构建决策树进行手势分类。

在手势交互中，难免会出现系统与用户交互意图不匹配的问题，例如用户交互过程中活动手腕，一般来说交互系统会一直将用户行为转化为交互指令而出现的未知错误，这也被称为空中手势交互的 *Midas Touch* 问题，同时理解用户的交互意图也能提高交互的自然性和流畅性。用户行为是多方面作用的结果，即包含震颤信息，也由用户主观意图驱动。通过研究发现，健康人群的震颤信号是一种线性随机过程，而用户的交互意图则是一种用户自发的主观行为，具有非线性特性，两者都是通过神经元传达到肌肉发生作用，因此探究震颤信号中偏离线性的非线性信息，为识别用户的交互意图提供了一种解决思路。本章依据震颤信号的特征进行分析，设计用户实验采集用户行为的数据，提取震颤信号的特征以识别用户是否具有交互意图和探究交互意图强弱的影响。

综上所述，本章首先分析平面交互的基本功能，再通过用户实验设计基于姿态的手势集合；然后针对用户提出的手势集合提出了一种坐标映射方法和手势识别方法；最后分析健康人群的肌肉震颤与交互意图的联系，设计实验评估不同行为下的震颤特征，并通过多种分类器进行分类，结果表明利用肌肉震颤判别交互意图具有可行性。

4.1 基于姿态的平面交互手势设计

本节主要内容为建立以用户为中心的平面交互系统的手势集合。在手势交互领域，平面交互是指用户与二维平面上的信息进行交互，典型的二维平面包括智能手机、大屏交互界面，用户通过手势与平面空间的物体进行交互。该研究针对平面手势交互进行设计，第一部分分析了平面交互中的典型功能，确定了平移、选择、抓取、旋转和缩放等交互过程的基本功能。第二部分通过用户参与式设计方法，收集用户自然手势和偏好，设计出一套以用户为中心的基于姿态的手势集合。第三部分探讨了该

手势集合的实现方法，包括手势坐标映射与转换为设备指令两步，使各种日常手势都可以直观地用来控制平面交互过程，实现了一套符合用户习惯和偏好且支持多种典型操作功能的平面手势交互集。

4.1.1 平面交互基本功能

本小节探讨在平面手势交互中常见的功能与逻辑，将平移、选择、抓取、旋转和缩放作为本研究平面交互的基本功能。

在平面交互中，由于只有二维空间的限制，手势设计的自由度相对受限，因此选择恰当的基本功能尤为关键。本文从手势交互的相关研究中^[1]，总结了平面交互的核心功能，移动控制、定位与确认对象、变形控制三个部分。结合王刚^[1]等人的研究，本研究将平移、选择、抓取、旋转与缩放作为平面交互的核心功能，其能够有效覆盖平面交互的主要场景与需求。其中平移和选择满足定位与确认的基本需求，抓取、旋转和缩放构成了对象变形控制的主要形式。它们覆盖面广，容易记忆与识别，同时遵循用户认知习惯，是平面单手空中交互设计的最佳选择，基本功能的功能描述与场景见表 4-1。

这五类基本功能可以涵盖用户在平面界面操作时的主要需求，也为基于手势的平面交互建立了控制基础。后续工作将在此框架内对特定的手势体系进行优化和评估，以丰富平面空间场景下基于手势的交互方式。

表 4-1 五类基本功能描述与场景

交互功能		功能描述	场景
平移	上	向上移动	在平面空间中，控制光标/对象进行上下左右的移动。
	下	向下移动	
	左	向左移动	
	右	向右移动	
点选		选择对象	选择平面空间中的对象，后续执行抓取、旋转、缩放手势。
抓取		抓起对象	抓取平面空间中的对象，移动对象位置到目标区域。
旋转	逆时针	使对象逆时针旋转	对平面空间中的对象进行旋转，旋转到目标状态。
	顺时针	使对象逆时针旋转	
缩放	放大	使对象放大	调整对象的大小到目标状态。
	缩小	使对象缩小	

在表 4-1 中，针对平移功能，用户需要频繁进行界面内的位置变换操作，如平移地图或图片，因此设计平移手势尤为重要。选择功能可以让用户精确地选择特定目标，是大多数交互的基础。抓取功能模拟的手部抓握动作，可实现对象的抓取和移

动。旋转功能通过手掌的旋转来控制对象的方向变换。最后，缩放功能允许通过手势控制界面或对象的显示比例。

尽管先前研究提出了各种基于手掌和手指姿态的手势集^[1]，经过对比考虑，本研究认为仅依靠手部姿态难以有效实现平面交互所需的这些功能，尤其平移功能无法得到很好的映射，例如用户执行平移向左的功能时，只需要将手从左边移动到右边，在这个移动的过程中，用户的手部姿态不会发生变化。

此外，完全依赖用户记忆预设姿态的交互方式，存在较高的认知和使用难度。为充分发挥手势交互的直观性与易用性优势，本研究将采用以用户为中心的构建手势集实验。

4.1.2 用户设计手势实验

本小节设计了一项用户参与式的设计手势集合实验，实验目的为设计一组以用户为中心且基于姿态的空中手势集合。面向用户自主设计手势的启发式实验，最大的优点在于能够挖掘人类手部运动的自然映射习惯，获得高度符合用户直觉与认知的手势体系。统计研究显示，此类手势的可理解性和可记忆性明显优于研究者自顶向下预设的手势集。此外，面向用户的手势设计还能够显著降低交互记忆与认知负荷，减少用户的训练时间，提升体验流畅度。本研究通过让普通用户根据操作需求自由设计手势，获得了更加可用与人性化的平面交互手势体系，为后续的交互系统改进和优化奠定了基础。

1) 实验参与者

本次实验招募了 7 名参与者参与实验，其中 5 名男性和 2 名女性，年龄在 23 岁~25 岁 ($M=23$, $SD=1.5$)，其中 6 人惯用手为右手。

2) 实验任务和设计

本实验试验任务为用户观看功能演示动画并进行手势设计。本实验采用了启发式的研究方法，即通过功能动画的示意和解释，激发用户自主设计手势的思维。向用户展示平面空间中五类功能的动画，包括平移、选择、抓取、旋转和缩放，并解释其含义、应用场景和设计要点，作为手势设计的参考指引。要求用户采用大声思考法^[1]，即在设计手势时将想法大声讲出自己的想法，研究人员通过录制和分析用户的言语数据，挖掘手势设计的认知逻辑，建立用户思维模型。



图 4-1 用户正在设计手势

为了让用户更加直观的感受功能，研究人员绘制了每个功能的动画，见表 4-2，并向参与者描述每个功能的作用和效果。研究人员告知用户需要完全基于姿态进行交互的前提下，令用户针对平面交互的五种功能进行手势设计，其中某些功能对应多个子功能，平移分为上移、下移、左移、右移，旋转分为顺时针旋转、逆时针旋转，缩放分为放大与缩小，实验全程记录用户的设计手势与思考内容。在表 4-2 中，示意图箭头左侧表示操作前的对象信息，箭头右侧表示操作完成后的结果。

表 4-2 平移、点选、抓取、旋转、缩放

功能	功能效果示意图
平移	
点选	
抓取	
旋转	
缩放	

在本研究中，引入手势集合的一致性系数^[1]对手势集合进行分析，手势的一致性系数定义如公式 4-1。

$$AR(r) = \frac{|P|}{|P|-1} \sum_{P_i \in P} \left(\frac{|P_i|}{|P|} \right)^2 - \frac{1}{|P|-1} \quad (4-1)$$

在公式 4-1 中，P 是参与者的数量，P_i 是选择手势 i 的人数。AR(r)>0.5 表示具有非常高的一致性，0.5>AR(r)>0.3 具有较高的一致性，0.3>AR(r)>0.1 具有一致性。一致率越高，代表更多的参与者提出了相同或高度相似的手势提议，参与者对于执行某个功能的手势理解存在更高的一致性，该手势具有更好的可猜测性，从用户体验角度来看更自然、更容易被接受，获得了更高程度的用户共识。

3)实验过程

实验开始前，向用户说明实验目的，要求用户填写知情同意书与基本信息调查问卷。之后向用户解释基于姿态的手势交互方法，明确告知用户通过手部姿态进行功能控制与手势设计，手部姿态即手部姿态包括手掌的俯仰动作、翻滚动作、收展动作，所有手指的关节角度以及所有关节的旋转过程，不能通过手的空间位置进行手势设计。研究人员向用户解释需要设计的五类功能，要求用户在大屏幕前使用单手进行空中手势设计，并告知用户在设计过程中，将自己的思考过程大声说出来，对任务的理解、手势选择的顾虑、怎样思考的、为什么这么设计等。每当用户设计完手势后，需要对自己设计的手势填写两个 7 点 Likert 问题：你是否认为该手势容易执行？1：非常困难，7：非常容易；你是否认为该手势能够与该交互功能直观的联系到一起？1：非常困难，7：非常容易。当用户提出一组完整的手势集合之后，对用户进行访谈，探究的问题主要在于用户提出的手势集合中，给出最容易、最难和最满意的手势，并要求用户给出原因。最后再重新复述一边实验目的，实验完毕。

4)实验结果：

实验总共收集了 90 个手势，根据收集到的手势分析特征，对具有相同特点的手势进行合并处理，将其合并为同一类手势，例如选择功能中，有用户提出五指张开的情况下食指弯曲表示点击，和用户提出其余手指并拢的情况下食指弯曲表示点击，这两种动作的核心都是食指的来完成点击功能，即将用户想要表达的手势核心动作如果一致，则可以归为同一类手势。

通过收集到的手势集合，根据公式 4-1，分别计算各个功能的一致性分数，手势集的平均一致率为 0.35，体现出手势集合具有较高的一致性，结果如图 4-2 所示。

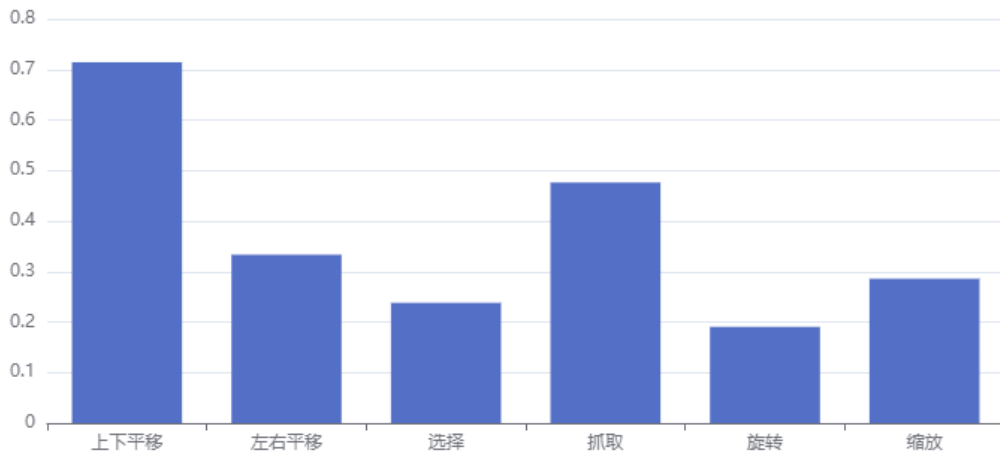


图 4-2 用户参与设计手势的一致率分数

在图 4-2 中，用户在光标上下平移的功能上，具有较高的一致率 $AR(r) = 0.71$ ，只有一名用户使用手掌俯仰和四指弯曲表示移动，其余用户均选择手掌俯仰控制光标上下移动。左右平移和抓取功能的一致率分数均超过了 0.3，这两类功能具有较高的一致性，选择、旋转和缩放功能的一致率分数也超过了 0.1，这两个功能的手势集合具有一致性。

用户在进行光标上下平移的手势设计时，均提出了使用手掌俯仰的转动进行控制，有一位用户表示“我想根据指尖的方向进行判断”，将指尖位置和光标位置联系起来。唯一不同手势的用户则是提出了触发式的方案，将手俯仰旋转和四指弯曲组合在一起，当检测到该动作时，光标以匀速移动，直到用户将手伸直放平。用户在光标左右平移功能上，出现了不同的结果，有四名用户选择了手掌偏航转动动作控制光标左右平移，有一名用户说“我的手偏航转动的时候，活动范围有点小，不像手俯仰转动那样大范围”，于是放弃了使用手掌偏航作为左右平移的光标控制，选择了使用手掌翻滚转动作为光标左右控制。还有一名用户则是使用拇指指向手心作为区分，使用手掌的俯仰进行控制。也有用户沿用了自己的设计风格，采用了触发式的方案，使用手掌翻滚和四指弯曲控制左右移动。

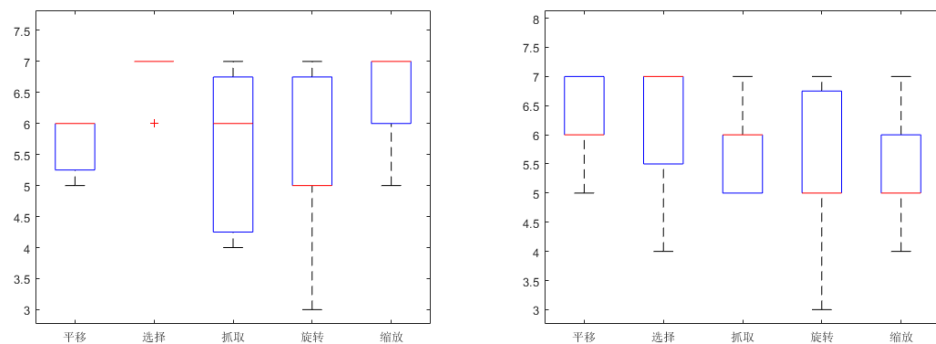
在选择这一手势上，用户几乎都没有过多思考就给出了手势，有三名用户使用食指点击进行选择，有一名用户说“这个手势就像点鼠标一样，点一下就选择了”，有两名用户使用食指和拇指捏住的手势，有一名用户说“选择可能会很频繁的触发，因此我选择食指与拇指捏住这个手势，因为这个很容易执行”，有两名用户选择握拳作为选择手势。

在抓取功能上，有五名用户选择了五指并拢，有一名用户一开始也选择了五指并拢，但是发现执行该手势时再执行移动手势受限，于是换成食指伸出，食指的方向控制抓起后的物体移动。同样有一名用户沿用了自己的设计风格，两指捏合的方式控制抓取，其余手指弯曲控制触发式的移动指令。

旋转手势中，有三名用户通过五指张开触发旋转指令，模拟抓住这个物体进行旋转控制，有两名用户则是将自己的手看作旋转的物体，执行翻滚动作控制对象的旋转，有两名用户采用触发式的指令进行控制，伸出食指和中指，分别通过偏航和俯仰进行旋转。

缩放手势中，有四名用户通过调整手指指尖的距离进行放大与缩小，有一名用户则是使用多个手指的，有一名用户采用手指和旋转来进行操作，拇指表示触发缩放，使用手掌翻滚来进行放大缩小的控制。

在用户提出手势之后，让用户立刻回答两个 7 点 Likert 问题，统计这五种功能的得分，绘制出盒图如图 4-3 所示。



a) 用户针对提出手势易用性评分

b) 用户针对提出手势直观性评分




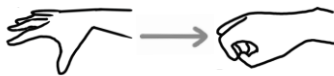


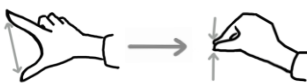
图 4-3 用户手势评分箱线图

在图 4-3 中，在手势易用性评价中，选择手势有六名用户打了 7 分，一名用户打了 6 分，都认为自己针对选择提出的手势容易执行。针对平移手势则用户提出，在进行左右移动时，手的旋转角度受限，不容易控制光标移动到屏幕的边界。抓取手势在易用性上得分较低，分析用户的想法得出，用户认为执行抓取手势之后，还要控制抓取的物体移动，所以认为该功能较难执行。在手势直观性评价中，用户 7 对自己提出的旋转手势打了 3 分，该用户提出的手势为伸出食指和中指，手掌俯仰控制旋转角度。

针对于不同的用户而言，各自喜欢的风格可能也会有很大区别。有的用户喜欢模拟现实的操作，用户 5 提出像点击鼠标一样使用食指点击执行选择功能，将大拇指和食指指尖的距离看作物体的对角线执行放大和缩小的功能。有的用户并不喜欢将功能细分到具体的手指，用户 3 说“我不喜欢细分手指的功能，我觉得手势交互最好的就是我大范围的操作，手抓一下能实现大部分功能就好”。有的用户不擅长使用连续的控制，喜欢触发式的控制。用户 7 所提出的所有手势，都是用手做出一个指令，系统识别后启动对应的功能。用户还考虑到了执行频率的问题，对于执行频率高的功能，选择更简单的手势。

本研究中用户自定义手势集的开发遵循最大群体优先的原则。具体而言，对于每个操作功能，通过统计分析用户提交手势提案的相似度分布，选择相似度最高的手势组作为该功能的映射。这可以尽可能避免一个手势对应不同操作导致的命令冲突。当不同操作之间手势集存在冲突时，赋予用户群体更大的子集自定义的手势以优先权，即多数人优先。这种方法能够在平衡操作一致性与用户自主性之间取得最优解，手势集合见表 4-3。

表 4-3 最终选择的手势集

交互功能		动作描述	手势
平移	控制光标进行上/下移动	以拇指掌指关节与小指掌指关节为轴进行转动。	
	控制光标进行左/右移动	以手掌所在平面法向量为轴进行转动。	
点选		食指做出点击动作、余下手指自然伸展。	
抓取		五指自然握拳。	
旋转	控制目标逆时针/顺时针转动	五指弯曲，以手掌所在平面法向量为轴进行转动。	
缩放	对目标进行放大操作	食指与拇指捏合、余下三指自然握拳，食指与拇指自然张开至呈 L 形。	
	对目标进行缩小操作	食指与拇指张开呈 L 形、余下三指自然握拳，食指与拇指自然捏合。	

4.2 平面空间下手势交互实现方法

本节主要内容为探究以用户为中心，基于姿态的手势集合实现方法，主要分为坐标映射与手势指令实现两部分。全手行为跟踪方法是平面空间手势交互的核心部分，由于人手与交互空间维度的不一致性，在平面空间交互中还需要交互映射技术，将人

体交互空间的手势行为转换为大屏空间坐标系的指令，其中关键是对手势空间和大屏空间的转换，本节分为两部分分别介绍系统所用的空间映射方法及手势集分类方法。

4.2.1 手部姿态与平面空间坐标映射

本小节主要介绍了基于手部姿态的平面交互映射模型。映射模型的目标是将用户手势角度变化映射为平面光标的位置控制。具体而言，以用户手部向量为基准坐标，测量用户单位时间内手掌俯仰和偏航运动的角度变化。然后将角度变化量映射为平面空间中光标竖直和水平方向的位移。这种映射方式参考了用户习惯，即使用手掌转动控制光标上下移动。为使控制更加自然连续，将中指指骨末端相对手腕的变化距离映射为光标变化距离。

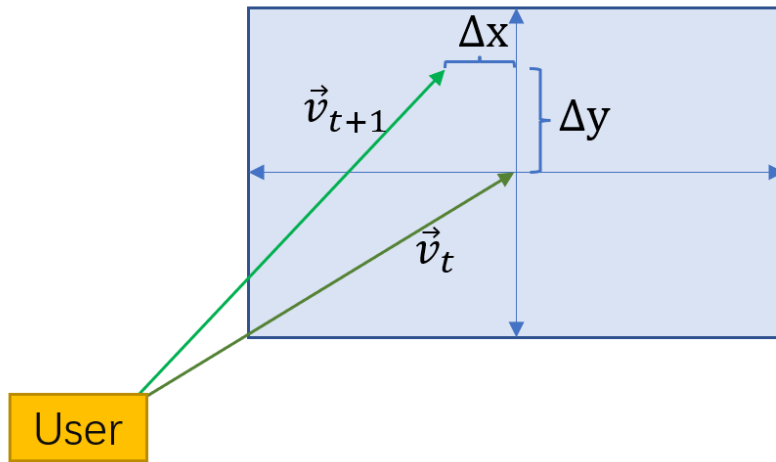


图 4-4 坐标系转换模型

在图 4-4 中， v_t 表示用户 t 时刻下的初始向量， $v_t = [0, D, 0]$ ，而 v_{t+1} 则表示利用手部姿态变化量旋转后的向量，当用户在单位时间内执行旋转动作，该过程用四元数 ${}_{H_t}^L q$ ，执行偏航动作四元数 ${}^{H_{t+1}}_{L_t} q$ 表示，该转动过程的姿态四元数满足公式 4-2，用户手所在向量 \vec{v}_t 的转动结果如公式 4-3。

$${}_{H_t}^{H_{t+1}} q = {}_{L_t}^{H_{t+1}} q \otimes {}_{H_t}^L q \quad (4-2)$$

$$v_{t+1} = R({}_{H_t}^{H_{t+1}} q) v_t \quad (4-3)$$

将公式 4-3 向量的 v_{t+1} 的 x 分量和 z 分量的值映射到平面空间中竖直方向的移动和水平方向的移动。这种映射方式受到 4.1 节中用户的启发，用户选择手掌俯仰转动作为控制光标上下移动的手势时，提出根据指尖的位置进行控制。鉴于不同的手势需要手指运动实现，因此本研究将其抽象为将中指掌骨视为移动的向量，掌骨末端相较于手腕的空间移动，映射为光标的空间移动。

上述过程将手部姿态的变化角度映射为了 2D 空间中的移动距离。原始数据到设备控制指令的映射是实现手势交互的关键步骤。常见的基于原始数据的映射方法包括线

性映射、非线性映射、基于模型的映射等。线性映射模型简单直接，但会引入失真。考虑到人手腕关节转动范围和运动时间特点，设计合理的阈值，实现光标位移的非线性平滑控制。这种模型使手势控制更加符合人机交互的自然特性，提升了交互体验。线性映射通过线性函数建立输入和输出之间的对应关系，该方法简单直接，但容易引入极端值或失真。非线性映射如指数、对数等复杂映射函数，可以提供更平滑自然的控制效果，但需要精心设计。基于模型的映射是指运用机器学习模型来学习输入和输出之间的复杂映射关系，可以自动适应不同用户的偏好，实现个性化交互，但需要收集大量训练数据。

基于 4.1 小结中用户提出的手掌运动范围受限的担忧，本文改进了这一映射方案。有研究者^[1]提出了一种 Go-Go 交互技术，使用交互式增长了用户的非线性映射。映射函数的设计在平衡映射的准确性、连续性、容错性等方面需要综合考虑。同时，评估和迭代优化也是确保映射质量的重要环节。

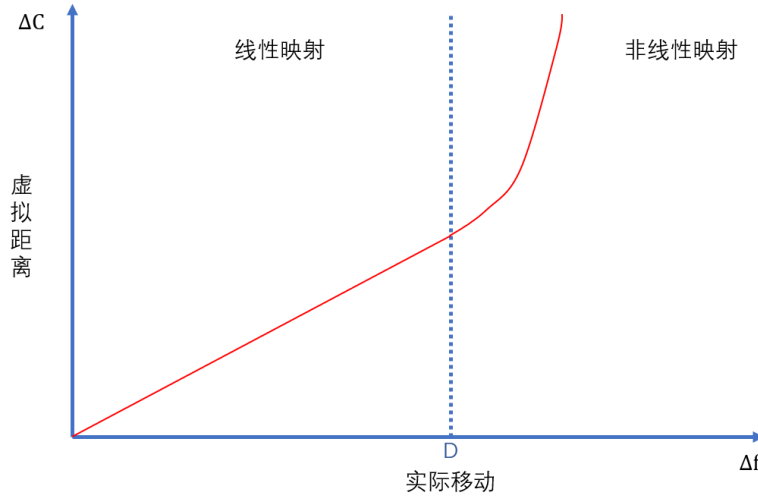


图 4-5 基于分段函数的非线性映射图

在图 4-5 中，横轴表示手部姿态旋转的角度，纵轴表示光标在平面中的位移距离。这表明在映射时是一段分段函数，有研究者^[1]指出人手腕关节的俯仰角度是 110 度的范围，偏航却只有 30-50 度的范围，这为本研究设计分段函数的临界点与函数提供了参考。综合考虑，本研究建立映射分段函数，如公式 4-4 所示。

$$\Delta C = \begin{cases} a \times \Delta f, & \Delta f < D \\ a \times \Delta f + (\Delta f - D)^2, & \text{other} \end{cases} \quad (4-4)$$

在公式 4-4 中， ΔC 表示光标移动位置变化量， a 表示映射系数，对应于手掌转动幅度与光标移动位置映射系数， Δf 表示掌骨相对于掌心移动距离变化量，通过公式 4-3 给出， D 表示在单位时间内手指移动的距离临界值。根据该公式进行坐标映射，实现了基于手部姿态的平面交互映射模型。

4.2.2 基于手部姿态的交互指令实现

本小结主要介绍了 4.1 小节中用户提出手势集的识别方法，分析在每种手势下手指的功能和状态，建立决策树模型进行手势分类。分析 4.1 小结用户提出的手势集合，发现可以将其转化为指令型的静态手势和动态的控制，针对基于佩戴设备采集的手部和手指关节数据，通过决策树将这类信息作为判断属性，递归地构建出一个树形的手势分类器。

在用户提出的手势集合中，平移手势仅需要手掌的偏航与俯仰操作，不需要手指参与操作；点选只需要食指弯曲，其余手指不需要弯曲；抓取动作则是执行握拳手势，所有手指都弯曲到最大角度；旋转的手势更像模拟抓住了物体进行旋转，通过手掌偏航控制旋转方向；缩放则是中指、无名指和小指弯曲到最大角度，通过食指和拇指指尖的距离来控制缩放程度，拇指和食指之间的距离可以利用关节角度通过前向动力学进行估算。为了更直观的表达五种手势中各个手指的参与情况，本节总结了不同功能下手指参与的关节数目，见表 4-4。

表 4-4 手指在不同手势中弯曲关节数

功能\手指	拇指	食指	中指	无名指	小指
平移	0	0	0	0	0
点选	0	1	0	0	0
抓取	2	3	3	3	3
旋转	1	2	2	2	2
缩放	0	0	3	3	3

通过表 4-4 可以发现，仅通过判断食指的弯曲关节数，就可以区分点选、抓取、旋转手势，但这样更容易误识别，综合多个手指考虑以提高手势识别模型的准确率。

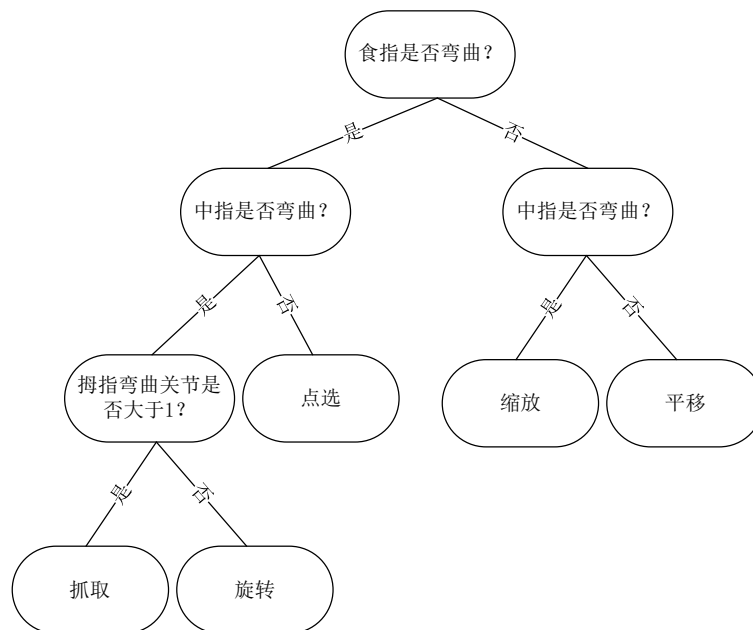


图 4-6 手势识别的决策树结构

本节通过构造决策树进行手势分类。决策树由于其模拟人类决策过程的特性，被广泛应用于基于佩戴设备的静态手势识别任务中。在该任务中，使用包含多种动作传感器的智能数据手套采集手部和手指的运动数据，并人工标注对应的手势类别。然后利用决策树算法以手部运动特征为判断依据，学习拆分规则，建立决策树模型。该模型能够根据输入的手部运动数据，从树根遍历到叶子，得到手势的分类结果。决策树模型表现出较强的解释性，通过树的结构和决策规则，能直观地理解计算机为何将一组运动数据识别为某个类别。同时，决策树也支持增量学习，当数据集增加时只需更新部分树分支，这些特性使其在基于佩戴设备的静态手势识别等领域中被广泛采用。根据手势集合特征构造的决策树结构图如图 4-6 所示。

在图 4-6 中，通过判断食指是否弯曲，则将手势集合划分为平移、缩放一类，点选、抓取、旋转一类。再通过判断中指是否弯曲，可以区分出平移和缩放、点选手势，由于抓取和旋转都用到了五根手指，因此需要判断关节数目来进行区分，在本决策树中选择了拇指进行判断，因为旋转手势对拇指只有一个关节弯曲，更容易进行分辨。

4.3 基于姿态的交互意图识别方法

本小结主要探究通过震颤信号识别用户行为是否具有交互意图及交互意图强弱，为空中手势交互的 Midas Touch 问题提出一种解决方法。Midas Touch 问题是人机交互领域常见的问题之一，指当用户进行自然动作，改变手掌或手腕的位置和角度时，这些无明确交互目的而出现的姿态变化，也可能被交互设备所采集的传感数据误判为用户有意进行的操控手势指令。例如用户手掌的无目标晃动，就可能会被错误识别为某种具体的指向或选择动作，并导致执行未预期的操作。

本节主要探究人体震颤信号与交互意图之间的关系，从神经学角度分析了静息态震颤、动作态震颤及含交互意图的动作态震颤的特征，健康人群的震颤是一种线性随机过程，而交互意图则并不符合简单线性随机模型，具有一定的非线性性质，因此探究震颤信号的线性随机过程特征和偏离程度，可以分析用户行为是否蕴含交互意图。本节基于震颤振幅、频率等时间/频域特征以及符合线性随机过程的特点，构建了 14 个特征。在不同难度交互任务中采集参与者的各类震颤信号，利用这些特征可以区分含交互意图的动作性震颤和不含交互意图的震颤信息。然后本节探究了交互意图强度强弱与区分含交互意图的动作性震颤和不含交互意图的震颤的联系，结果并未发现存在相关关系。本节旨在通过健康人群的震颤信号特征，分析用户的交互意图，以实现更自然的手势交互。

4.3.1 手部行为震颤与交互意图

本小节主要描述了行为震颤的特征和交互意图对行为的影响，通过震颤信号来区

分用户行为是否具有交互意图及不同意图强弱对结果的影响，以通过这种方式更好的理解用户交互意图，为空中手势交互中 Midas Touch 问题提出一种解决思路。

神经学的相关研究将人体的震颤分为静息性震颤、动作性震颤，在实验分析中，令参与者执行具有交互意图的运动，提取过程中含有交互意图的动作性震颤(IAT, Intent Action Tremor)，而令参与者不具备交互意图情况下采集静息性震颤(RT, Rest Tremor)和维持姿态的动作性震颤(PAT, Postural Action Tremor)。通过震颤的特性挖掘震颤信号的特征，通过震颤信号识别运动中是否蕴含交互意图。

震颤是指身体部位近似有节律的不自主运动，主要见于面部、下颌、上颚、眼睛、躯干和四肢。四肢震颤的典型表现为不规则、持续的运动，其频率和振幅可能发生波动。正常或增强的生理性震颤常呈现杂乱无序的宽频谱特征；而病理性震颤则更为规则、缓慢，噪声小，频谱中含有尖锐峰值。频谱分析可提供震颤信号在频域的组成信息：宽频谱表示多种频率成分贡献显著，集中峰值表示某单一主频率成分占主导地位。频谱特征可用于分析和判断震颤的类型、程度及状态。

自 1993 年以来，有两种主要的震颤分类在使用。第一种是基于观察到震颤时身体部位的活动状态，第二种是基于潜在疾病或条件的病因。本研究面向健康人群，因此采用另一种分类方式。而根据 Bain (1993)和 Deuschl 等人(1998)的报告，第一种对震颤的分类为当相关肌肉没有被激活，身体部位得到充分支撑以抵抗重力时发生休息性震颤；当相关肌肉被激活时发生的动作性震颤，包括姿势震颤、动能震颤、等长震颤、意图震颤和任务特定震颤。本论文基于该分类方式探究震颤特征与交互意图之前的联系。

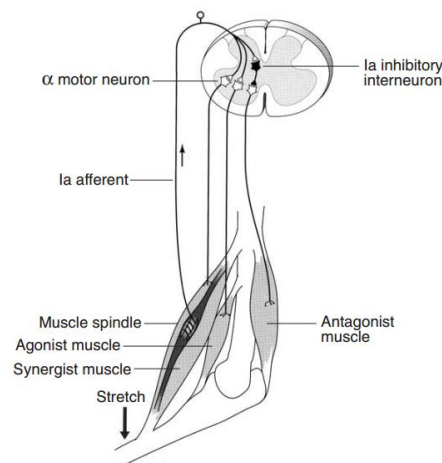


图 4-7 神经控制震颤信号示意图^[1]

研究者 Elble 等人^[1](1986)指出正常震颤包含两个不同的振荡成分：第一个成分是肢体拉伸反射所形成的欠阻尼振荡，频率由肢体的质量和刚度决定；这一成分的产生源自肢体本身的被动力学特性，当外力或内力干扰肢体时引起的振荡，外加质量通常会降低频率，外加刚性会提高频率，该成分与肌电爆发无关，但当震颤加剧时，运动神

经元的放电会通过反射性调节。第二个成分与肌电爆发直接相关。研究者 Deuschl 等人^[1]表明静息性震颤频率通常在 4hz-6hz，而姿势性震颤频率通常在 4hz-12hz 之间。研究者 Elble 等人^[1]指出正常的生理震颤通常集中在 8-12 赫兹的范围，但不同人也可能存在较大差异。同时，研究者^[1]指出震颤是一种线性随机过程，具有时间反转对称性等特点。

而交互意图的形成是一个复杂的认知和决策过程，难以简单地用线性模型描述。具体来说交互意图的产生涉及复杂的大脑认知活动，会受到情感、经验、环境、目标等多种内外部因素的影响，这是一个多维的非线性过程。在交互前、中、后，用户的意图也会不断发生变化。这同样是个动态调整的过程，难以用静态的线性关系解释。不同用户由于其个体差异，在相似环境和目标下，会产生不同的交互策略，这种多样性也决定了非线性的特点。用户意图表达和意图理解中的语言、表情、手势等都是复杂的交互要素，互相影响也呈现出强非线性的特点。

总的来说，交互意图的形成和演化是一个多源、动态、个体化复杂的过程，很难用简单的线性模型进行表达，应该视为一个非线性的现象加以研究和处理。这为通过震颤信号的线性特征差异分析交互意图提供了理论基础。

4.3.2 交互中手部震颤提取方法

本小节使用巴特沃斯滤波器进行信号过滤，保留 4-12hz 的震颤信号，根据震颤信号的振幅、频率和符合线性随机过程的特性，描述了震颤信号具有偏度和峰度特征和具有时间反演对称性，将这些特征作为特征集的备选以进行实验设计与分析。

1) 震颤信号提取

根据时间序列可以对震颤进行测量，最简单和直接的是测量其振幅(平均振幅)。如果信号平均值为零且无整体漂移趋势(如震颤加速度记录)，那么振幅即为信号的均方根，也基本等于标准差。为去除某些频率分量，可以对信号进行滤波处理，这相当于部分或全部去除其傅立叶变换，然后进行反变换生成新的滤波信号，如果存在整体漂移趋势，高通滤波可用于消除低频成分，从而更准确测量振幅。判断去除多少低频部分需要谨慎，例如呼吸、心跳频率影响位移记录，而震颤频率一般认为至少有 4 赫兹，并且一般不高于 12hz。除振荡振幅外，从参考点偏离的总量也可反映病理学情况。位移信号可通过傅立叶分析分离漂移和震颤成分。滤波前后示意图如图 4-8 所示。

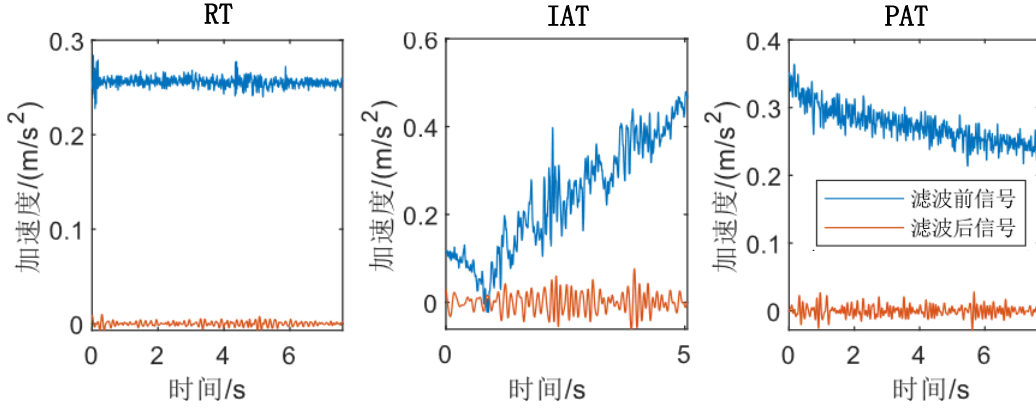


图 4-8 RT、IAT、PAT 的滤波前后信号

在图 4-8 中, 使用 10 阶巴特沃斯高通滤波器($f_1 > 4 \text{ Hz}$)进行带通滤波, 用于消除手部动作的低频信号, 再用 10 阶巴特沃斯低通滤波器($f_2 < 12 \text{ Hz}$)将高频噪声过滤掉, 保留静息性震颤信号(4hz-6hz)与动作性震颤信号(6hz-12hz)。通过这种方式能够去除运动的影响。综上, 时间序列分析可定量分析震颤的振幅、频率等参数, 振幅测量需要考虑信号预处理, 以提高测量准确性。

2) 线性随机过程特征

有研究者^[1]指出健康人的震颤信号是一种线性随机过程。而人的交互意图也是一种非线性的行为, 通过分析震颤信号的线性随机过程所具有的特征, 可以区分含有交互意图与不包含交互意图的行为震颤。

生理性震颤具有高斯随机性且基本过程为线性, 检测震颤是否含交互意图的简单方法是观察数据值分布是否偏离正态分布。常用方法是计算高阶矩, 这些高阶矩可以检测出信号分布与正态分布的偏差。例如第三阶矩(偏度)反映分布的对称性, 第四阶矩(峰度)反映分布尖峰情况。正弦波等规则对称振荡其偏度近似为零, 但峰度明显低于正态分布。

综上, 如果生理震颤为高斯随机过程, 可以通过计算高阶矩, 判断信号分布是否偏离正态分布, 以检测信号中是否存在额外的意向性成分, 这是一种简便有效的方法。公式 4-5 给出了偏度的计算过程。

$$m_3 = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^3 \quad (4-5)$$

公式 4-5 为偏度计算公式, 其中 n 表示数据总数, x_i 是第 i 个观察值, s 表示观察值的标准差, \bar{x} 表示观察值的均值。偏度是描述概率分布形状对称性的一个统计量。它反映分布曲线相对于中心的左右尾部延伸的情况。偏度为正表示右侧尾部更长, 负偏度则表示左侧尾部更长。偏度绝对值越大表示分布曲线越不对称。零偏度对应的是对称分布, 如正态分布。判断分布的偏斜方向和程度, 偏度是最有效的度量标准, 可判断实际分布与理论对称分布的差异。

峰度描述概率分布曲线的尖峭程度，反映尾部与峰值区域曲线形状特征。正态分布峰度为 3，称为正常峰度。当实际分布的峰值区域相对较尖峭时，峰度大于 3，称为尖峭峰度；当峰值区域较为平缓时，峰度小于 3，称为平坦峰度。

$$m_4 = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^4 - 3 \quad (4-6)$$

公式 4-6 为峰度计算公式，其中 n 表示数据总数， x_i 是第 i 个观察值， s 表示观察值的标准差， \bar{x} 表示观察值的均值。正态分布曲线代表了大多数实际分布均匀和对称地延伸到两个方向的尾部区域，与其峰值区域契合，这是一种典型和标准的分布形状。当实际分布曲线尾部相对于峰部突出时呈现较尖峭的非对称形状，如指数分布，其峰度大于 3；反之，当实际分布曲线整体较为平缓时呈现钟形，其峰度小于 3。通过与标准正态分布的比较，峰度直观反映分布整体形状的差异特征，是判断分布形状的重要指标之一，针对震颤信号的正态分布拟合如图 4-9 所示。

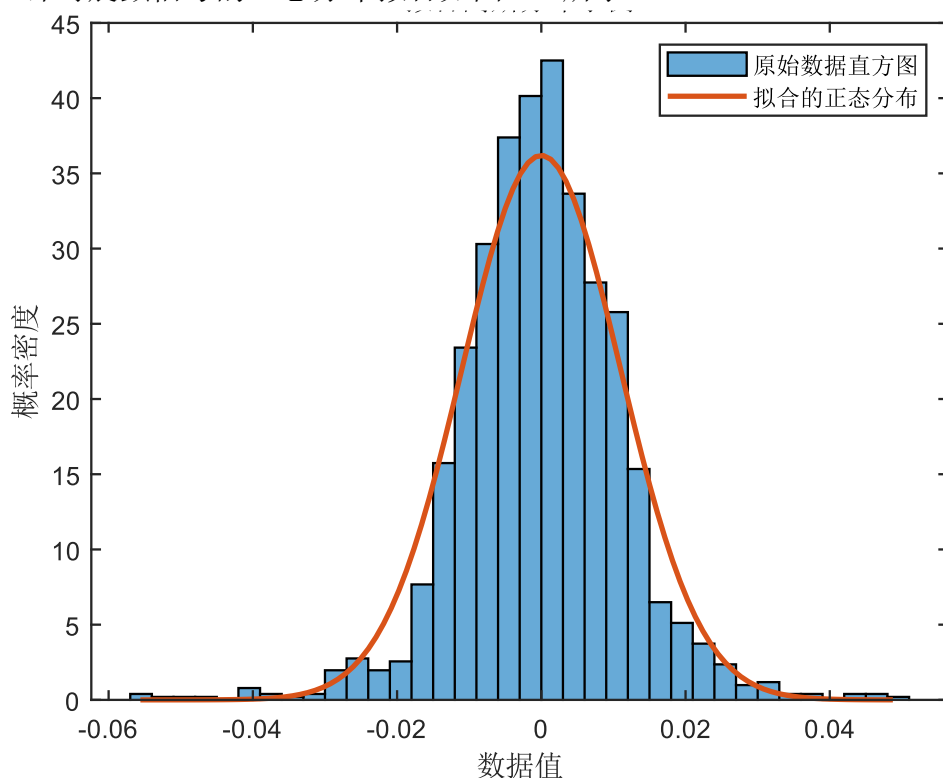


图 4-9 健康人群震颤信号分布直方图

有研究者指出【Edwards and Beuter 2000】，震颤信号具有时间反转对称性，即在时间反转下性质不变，而非线性过程的性质通常会改变，也可以通过这种方法来判断是否蕴含交互意图的非线性差异。时间不对称性的一般测量可以用来诊断正确的非线性类型。其中一种方法是从相似值分别预测向前和向后的值(具有相同滞后)，并比较两者差异的大小。对所有可能的起始值均进行这样的预测，将方差相加即可得到一个滞后相关的时

间不对称性度量。时间不对称性的另一种度量方法是计算点和滞后点之间差异的立方和，见公式 4-7。

$$R = \max \left\{ \frac{\sum_i (x_i - x_{i-j})^3}{(\sum_i x_i^2)^{3/2}} \right\} \quad (4-7)$$

在公式 4-7 中， j 为时间滞后参数， x 为统计量。线性高斯过程具有时间反转对称性，而非线性过程不具有该特点，通过计算时间序列在正反两个时间方向上的预测误差，可以判断时间不对称性，从而识别信号的非线性成分，实现对含有交互意图行为的非线性过程的有效检测。

如上所述，震颤信号的线性随机过程所具有的特征可以有效判断信号的线性与非线性特性。频谱分析的常用方法还包括功率谱密度分析和主频率分析。功率谱密度反映信号功率在频域上的分布情况。通过离散傅里叶变换可以计算得到静息态和运动态震颤信号的功率谱密度值，作为判断震颤强度的指标。主频率分析即识别信号中占主导地位的频率成分。通常震颤信号会包含一个主要的频率成分，这可以反映震颤的类型和状态。短时傅里叶变换的时频分析可以有效提取该主频率。因此，频谱分析中功率谱密度和主频率等参数的计算，可以从频域角度更全面地分析判断震颤的特征。

4.3.3 基于震颤的交互意图识别

本小节描述通过震颤特征进行交互意图的识别，主要分为识别震颤信号是否用户行为蕴含交互意图以及交互意图强弱。首先介绍用户震颤信号的分类，有无交互意图的震颤信号类型，用于通过震颤区分行为是否具有交互意图；然后描述构建震颤的特征集合、对特征进行标准化处理、相关性检验去除高相关特征、使用重复 5 折交叉验证进行结果评估、使用准确率、召回率、F1 分数作为指标进行结果分析；最后设计用户实验进行分析。

根据 4.3.2 小结内容，在本章中将用户有意图的行为所产生的震颤称为意图性动作震颤(IAT, Intent Action Tremor)，将用户没有交互意图的维持动作产生的震颤称为维持性动作震颤(PAT, Postural Action Tremor)，而用户肌肉没有被激活时无交互意图的震颤称为静息性震颤(RT, Rest Tremor)，三种震颤信号原始数据见图 4-10。

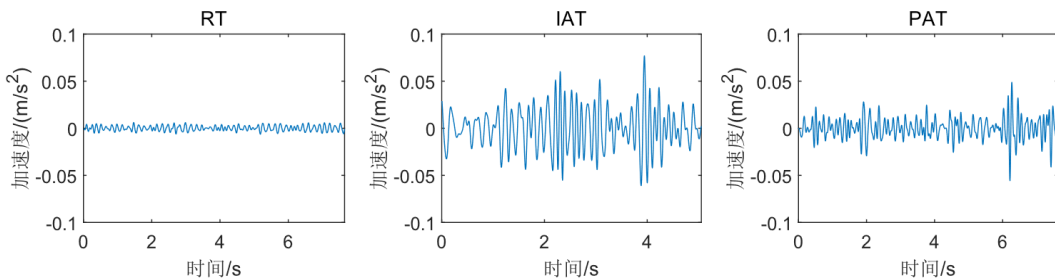


图 4-10 RT、IAT 和 PAT 震颤信号波形

在图 4-10 中, 最左侧为用户肌肉没有被激活时的震颤信号 RT, 震颤幅度较低; 中间为用户执行交互任务, 具有交互意图的运动性震颤 IAT; 右侧为用户维持姿态的运动性震颤 PAT, 这个过程用户不具有交互意图。

这三种震颤中只有 IAT 具有交互意图, 于是根据交互意图的强烈程度, 设计不同的交互任务将对 IAT 进行进一步区分, 判断用户所产生的 IAT 信号强烈程度, 在不同交互意图情况下, 用户震颤信号如图 4-11 所示。

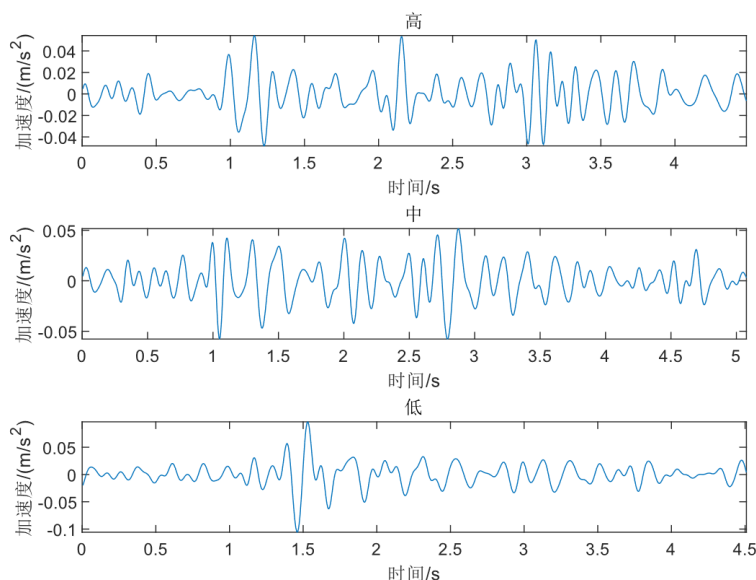


图 4-11 三种交互意图强度的 IAT 波形

1) 构建初始特征集合

根据 4.3.2 小节内容, 针对基于震颤信号的分析, 本节首先需要对信号进行特征提取, 获得一个初始的多维特征空间。具体而言, 可以提取时域特征如均值、方差、四分位值、峰度峰值等, 反映信号在时间轴上的统计特性; 此外还可以提取频域特征比如不同频带下的功率谱密度, 表示信号在频率分布上的信息; 偏度、峰度等高阶统计量也能反映信号的非线性动态; 主频率直接反映震颤的频率形态。所有这些构成了一个初始的 14 维特征向量空间, 它为后续构建新特征提供了基础, 14 维特征向量见表 4-5。

在表 4-5 中, 从震颤信号的分布中, 提取了 6 种时域特征、3 种频域特征和 3 种线性随机过程的特征。时域特征如均值、标准差等能反映信号幅值分布的集中趋势和离散程度; 峰值特征反映信号的极大极小值; 均方根反映信号能量大小。频域特征通过功率谱分析信号频率组成, 主频率反映信号主要成分; 随机过程特征如偏度和峰度描述信号概率分布的对称性和尖锐程度, 线性对称性反映高阶统计量之间的相关性。综上, 这组特征集合提供了丰富的统计学和信号处理角度的信息, 能够可靠地刻画复杂

信号的时域、频域和概率分布等关键属性。在此基础上，本小节可以进行更高层的特征选择、转换与建模，以实现特定的理解、诊断与预测任务。

表 4-5 构建特征集合

特征类别	特征名称
时域特征	均值(Mean)
	均方根(RMS)
	四分位点(Q ₁ ,Q ₂ ,Q ₃)
	标准差(SD)
	峰值(PV)
	峰-峰值(PPV)
频域特征	4-6Hz 功率谱密度(LPSD)
	6-12Hz 功率谱密度(HPSD)
	主频率 (FT)
线性随机过程特征	偏度(m3)
	峰度(m4)
	线性对称性(R)

2) 标准化处理

考虑到直接利用提取得到的 14 维原始时域和频域特征对复杂交互模式进行识别会带来一定的限制，本研究对特征进行适当预处理以消除特征值数量级上由量纲单位带来的差异。具体来说，可以对所有特征应用 Z-Score 标准化方法，使每个特征值分布映射到均值为 0 方差为 1 的统一高斯分布空间。Z-Score 标准化公式见公式 4-8。

$$Z = \frac{x - \mu}{\sigma} \quad (4-8)$$

在公式 4-8 中， z 是 Z-score，表示标准化后的值， x 是原始数据点的值， μ 是样本数据的均值， σ 是样本数据的标准差。通过这个公式，原始数据点 x 被转换为相对于均值 μ 的偏差，然后除以标准差 σ 进行标准化。这样处理后，Z-score 表示数据点在原始分布中的位置，负值表示在均值左侧，正值表示在均值右侧，而数值的大小表示与均值的偏离程度。

通过这种线性转换，不同特征之间的绝对差值会被相对关系比例替代。这样在新的映射空间内，所有特征都可比较，都会对后续的模式识别产生均衡影响。标准化操作还使信号的稳定统计特性被保留，有助于更稳健地反映交互动态。采用这类量纲归一化的预处理技术，是使多个异质特征协同支持非线性判别模型(如 SVM 与随机森林等)训练的常用和有效方法。其结果是不同特征可比，模型参数更稳定，从而提高交互意图分类的整体性能。

本节手动提取出的特征涵盖时域、频域以及线性随机过程，但直接用于后续机器学习方法分类前，有必要对特征空间进行适当筛选与缩减。一方面，当前特征集可能存在冗余信息，使用全部特征不仅难以提高模型性能，反而会引入过拟合风险；另一方面，高维特征在实际应用中也会大大增加计算量。基于对上述 14 维震颤信号初始特征的分析，本文使用斯皮尔曼等级相关系数进行特征选择，去除强相关特征，优化特征子集的规模与表达能力。

3) 相关性检验

在进行 Z-Score 标准化消除特征量纲后，用斯皮尔曼等级相关系数评估各特征对之间的线性依赖性，该系数通过综合两个变量值之间的单调趋势，判断它们的统计相关程度。具体来说，可以求取标准化后特征矩阵中任意两列之间的斯皮尔曼系数。根据系数的绝对值，选择互相关性较弱的特征构成最后使用的子集。这通过相关性过滤可以去除冗余特征，使得保留特征对目标交互意图的条件独立性更强，避免多重共线性问题。采用非冗余、独立性较好的低维特征子集，可以使得基于决策树、贝叶斯网络等机器学习模型实现对复杂交互行为的更稳定、准确判别与分类。该特征选择方法计算高效，且通过阈值设置可以平衡特征数目与预测性能。

通过先后采用 Z-Score 标准化和斯皮尔曼等级相关系数方法处理原始特征集合中的 14 个特征，获得了一个经过量纲归一化和相关性筛选的优化特征子集。该子集在消除了特征量级差异的基础上，进一步剔除了互相关性高的冗余特征。其中，标准化通过线性映射实现了不同特征在数量级分布上的一致化处理，有利于它们对后续模式识别任务的均衡作用。而斯皮尔曼系数过滤则依据变量间单调函数关系的紧密程度，移除了相关性高的重复特征。因此，保留的特征子集在表达复杂交互动态信息方面可互补；并在支持向量机、K 近邻法等机器学习模型判别交互意图时，发挥更为独立的条件概率作用。这样，构建于该优化特征上的分类器可以获得更高的判别稳定性、准确率与泛化能力。

4) 机器学习模型

在进行了针对原始特征集合选择与转换优化后，得到了一个无量纲低相关的特征子集。基于该优化特征，本文选择机器学习模型去识别表示不同意图的用户行为。具体来说，将研究和对比支持向量机(SVM)、K 近邻(KNN)、决策树与朴素贝叶斯分类器在使用该特征子集条件下的准确度和性能。

SVM 通过寻找高维空间的最优决策边界，可以处理复杂的交互特征；KNN 通过实例示例直观地判别交互类别；而决策树与朴素贝叶斯可以从特征匹配中学习可解释的交互识别规则。设计实验评估这些模型对复杂动态特征刻画、防止过拟合、泛化于复杂交互情况等能力。最终选择合适的机器学习算法，使用优化特征识别不同交互意图和交互意图强弱的行为，进而达到获取用户交互意图的目的。

5) 结果评估指标

本文采用了准确率、召回率、F1 分数指标对分类模型的结果进行评估。其中，准确率直接反映了分类器将测试样本正确划分到真实类别的比例；F1 得分考虑了模型对每个类别的精确率与召回率的调和平均，评价了分类的精确度与覆盖面；而召回率特别强调了分类器对某个类别内部样本的正确识别比例。这三种指标从不同侧面反映了多类分类的性能好坏：准确率关注整体判别效果，F1 得分关注模型对每个类别区分特性的掌握程度，召回率对某特定类别内部样本覆盖面的成效。将它们一并采用给出全面的性能描述，并通过与其他基准模型的比较确定实验中分类器的效果提升与独特优势。采用这种多指标联合评估的方法，可以从多角度判断分类模型的综合效果。

本实验采用重复 5 折交叉验证进行性能评估，令机器学习分类方法得到更稳定和可靠的精度指标。具体来说，首先将全样本集随机分成 5 个子集，每次选择 1 个子集作为测试集，余下 4 个子集合并形成训练集；在该训练集上训练模型，计算其在对应的测试集上的精确度指标。如此循环 5 次，每份子集分别作为一次测试集，最终获得 5 个交叉验证结果。对该 5 个结果求均值可以得到该次划分情况下的模型泛化能力评分。通过使用重复交叉验证，不同数据划分和模型训练情况下对精度影响的变化都有权衡，使得最终的泛化性能评估更加稳定准确，可以对分类器的效果进行更加可靠的衡量。

4.3.4 实验分析

本小节旨在通过震颤信号特征识别用户行为中是否蕴含交互意图及识别交互意图的强弱，为空中手势交互中 MidasTouch 等问题提出一种解决思路，进而建立更自然的空中手势交互系统。手势交互中 MidasTouch 问题可以被描述为手势交互过程中用户的任何手部动作都会被计算机解释为指令并执行，这会导致用户行为误触发交互功能。本实验的实验目的为探究通过震颤信号识别用户行为是否蕴含交互意图及蕴含交互意图的强弱，并以此针对空中手势交互中存在的 MidasTouch 问题提出一种解决方法。

1) 参与者与实验设备

实验征集 9 名用户，均为所在学校的大学生，包括 4 名女性和 3 名男性，年龄分布在 20 岁到 23 岁之间($M = 20.55, SD = 1.013$)，惯用手均为右手，均无吸烟习惯和手部脑部疾病历史。48 小时内有 1 位同学有饮酒行为，3 位同学有喝咖啡的行为。有两位同学近期有服药经历，分别食用维生素及感冒药治疗感冒、食用阿莫西林治疗喉咙痛。

实验设备为惠普投影仪，投影尺寸为 100 寸；采集装置为 mpu9250 的惯性传感器，采样率为 200hz。



a) 实验设备及佩戴

b) 实验环境

c) 参与者正在参与实验

图 4-12 实验环境与交互设备图

2) 实验任务与实验设计

在本实验中，参与者需要根据提示控制光标移动到指定区域，并按照提示进行数据采集。本次实验要求参与者右手佩戴数据采集装置执行交互任务，每名参与者需要参与 3 个目标区宽度的交互任务，每个任务有 3 个实验组块，参与者每个实验组块中需要执行 10 轮。目标区宽度的 3 个水平采用 3×3 拉丁方在不同的参与者间进行均衡。

实验变量：目标区宽度 (30pt、90pt、180pt)。

实验类型：within-subject

实验过程中的提示分为视觉提示和听觉提示。视觉提示为交互区域高亮，提示用户需要将光标移动/停留在高亮区域内；听力提示为交互过程中的叮咚声，叮咚声提示用户要开始移动光标。

本实验根据 Fitts 定律模型^[31]构建交互任务，根据公式 4-9 中的难度指数 (ID)，给定了固定距离 D 下三种不同难度系数。

$$ID = \log_2 \left(\frac{D}{W} + 1 \right) \quad (4-9)$$

公式 4-9 为设计实验时研究人员应该使用难度指数 (ID) 的 Shannon 公式，其中 D 表示两个交互目标中线之间的距离， W 表示目标区的宽度。在本实验中，实验难度指数分别为 5.67、4.14、3.22，认为难度指数越高的交互任务用户需要做出的努力越大并且交互意图越强烈。

在通过实验得到用户执行任务过程中的指尖六轴数据后，对数据进行预处理操作，删除一些异常数据值。根据 4.3.2 小节中描述，提取不同行为下的震颤特征，进行数据统计与分析。本实验确保所有参与者信息的隐私受到保护，遵守人体研究的伦理准则。

本次实验探究用户含有交互意图的行为与震颤特征之间的联系。本小节首先对数据进行了预处理，根据 4.3.1 小节消除高频噪声信号以及低频的行为信号，之后进行数据筛选，删掉传输过程中出现错误的的数据。之后根据正常人震颤的线性随机过程特性，分别提取 m_3 、 m_4 特征、 R 特征以及 LPSD、HPSD 特征进行分析，并提取构造的特征，使用支持向量机、 k 最近邻、决策树、朴素贝叶斯分别对三种难度下运动性震颤和三种类型的震颤进行分类，计算其准确程度。

3) 震颤信号收集

在实验中，使用惯性传感器收集参与者指尖加速度、陀螺仪六轴数据，在交互任务中收集用户的 RT、IAT、PAT 信息，如图 4-13，在区域 1 记录 7s 包含休息性震颤信号（RT）的指尖数据；在区域 2 记录包含蕴含交互意图动作性震颤信号（IAT）的指尖数据；在区域 3 记录包含无意图姿势性震颤信号（PAT）的指尖数据；实验中探究不同交互意图强度的 IAT，则是从任务中的区域 2 中收集。

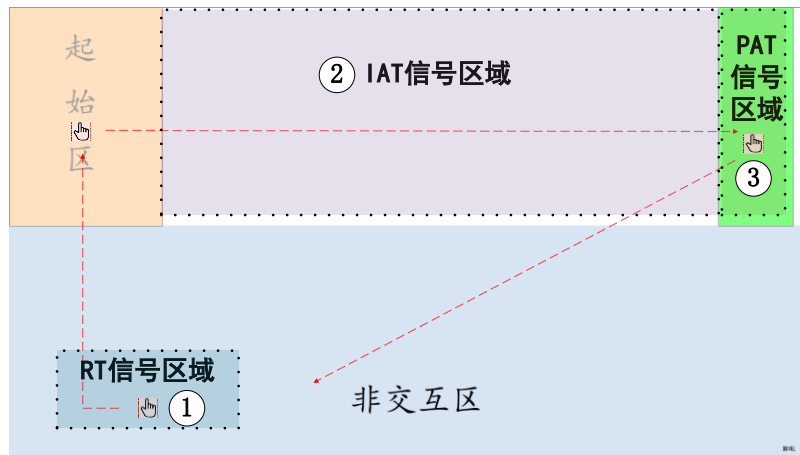


图 4-13 交互任务中收集用户信息区域

在本实验中，交互实验中采集用户的 RT、IAT、PAT 信息，用户状态分别为将手放在桌子上静止、在空中将手从一端旋转到另一端、在空中保持姿势静止。

4) 实验过程

参与者在实验人员的帮助下佩戴好采集设备并熟悉了解任务流程，帮助参与者能够表示熟练交互系统和任务流程后结束，明确告知用户所有的声音提示都提示开始运动，所有的高亮提示都提示停留高亮区域；正式实验阶段，用户任务目标为从起始区出发移动到目标区，一次交互任务分为了 5 个阶段：1) 用户将手静止在桌面上，保持 7s 静止；2) 系统发出声音提示，用户将手抬起并将光标移动到起始区；3) 移动到起始区后，系统发出声音提示，用户将光标移动到目标区；4) 用户保持手部静止 7s，系统高亮目标区提示；5) 采集完成后，系统发出声音提示，用户将手移动到非交互区。针对交互过程中的 5 个阶段，其交互过程中关键位置如图 4-14 所示。

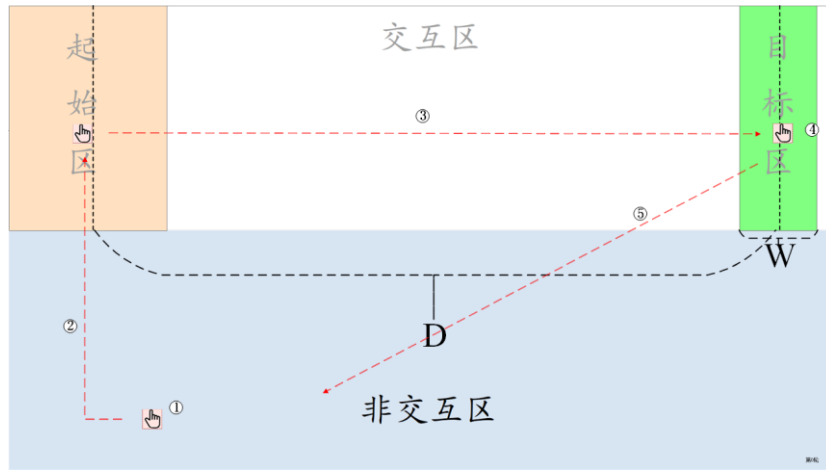
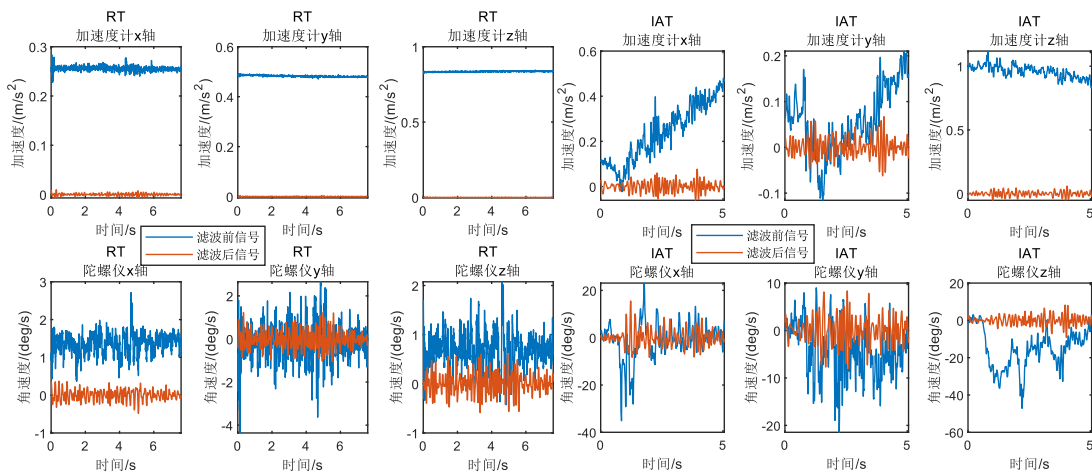


图 4-14 实验交互界面

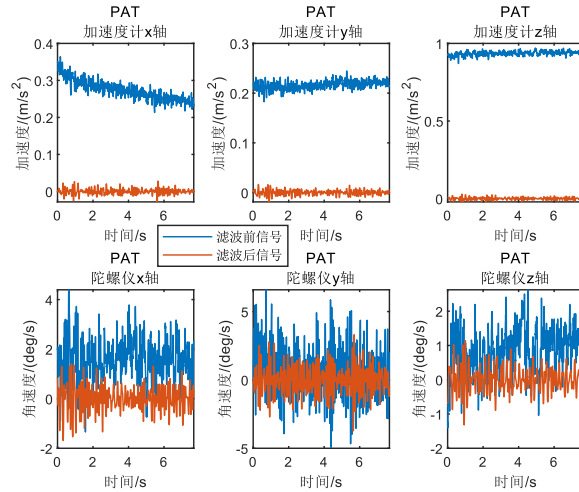
5) 实验结果

通过实验，本研究获得了用户指尖加速度和角速度六轴原始数据，通过数据预处理，去除低频运动影响以及高频噪声，得到 RT、IAT、PAT 震颤信号。其中 IAT 为具有交互意图的运动性震颤，而其他两种无交互意图。其中，IAT 包含三种难度下即强、中、低交互意图的运动性震颤信号，本研究首先利用震颤特征对 RT、IAT、PAT 进行区分，之后探究不同交互意图是否影响三种信号的识别，进而为空中手势交互中 Midas Touch 问题提出一种解决方法。

首先对数据进行数据预处理，保留运动过程中的震颤信号，去除其他噪声。数据使用 10 阶巴特沃斯高通滤波器($f_l > 4 \text{ Hz}$)进行带通滤波，用于消除手部动作的低频信号，再用 10 阶巴特沃斯低通滤波器($f_h < 12 \text{ Hz}$)将高频噪声过滤掉，保留 RT 信号 ($4\text{Hz} - 6\text{Hz}$)与动作性震颤信号($4\text{Hz}-12\text{Hz}$)。根据交互实验中标记与幅值检测，删除无效的数据，RT、IAT、PAT 滤波前后信号见图 4-15。



a) RT 滤波前后六轴信号 b) IAT 滤波前后六轴信号



c) PAT 滤波前后六轴信号

图 4-15 RT、IAT、PAT 滤波前后对比图

在对采集的手部和手指运动数据进行预处理后，得到 RT、IAT、PAT 的信号，其中 IAT 是用户含有交互意图的震颤信号，本实验接下来探究是否可以通过分析特征的差异，准确辨别出存在明确交互意图的 IAT 信号和无交互意图的 RT、IAT 信号。这样可为后续通过增强对复杂交互意图自身的识别，解决现有手势交互系统中存在的意图判断误操作问题提供可能性。

通过 4.3.1 小节所述，提取六轴信息的 m3、m4 特征，以实验中 block 为单位，以偏度为横轴，峰度为纵轴绘制分布图，见图 4-16。

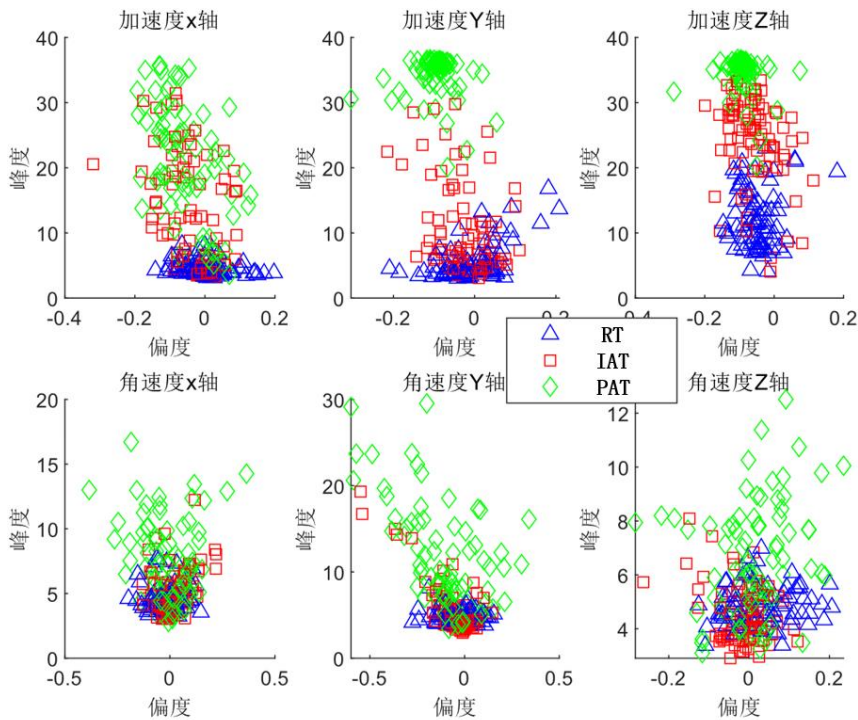


图 4-16 不同震颤类别的峰度偏度分布图

在图 4-16 中, 三种震颤信号的峰度在加速度 Y 轴与 Z 轴表现出差异显著, 通过 Wilcoxon 秩和检验中 RT 与 PAT 具有显著差异($Z = -9.12, p < 0.01$), RT 与 IAT 具有显著差异($Z = -10.9, p < 0.01$), 在六轴数据中, RT 的峰度均值高于 PAT 和 IAT, PAT 和 IAT 在加速度 Z 轴差异显著($Z = -10.2, p < 0.01$)。峰度表示为信号分布的尖峭度, 通过图 4-16 可以得出在加速度 Z 轴与 Y 轴上, RT 峰值最高, 而 PAT 的峰值次之, IAT 信号峰值最低。

综上, 仅通过峰度偏度分布, 可以通过加速度 Z 轴与 Y 轴数据区分 RT、IAT、PAT。接下来本实验分析震颤的时间反演对称特征 R 进行进一步探究。

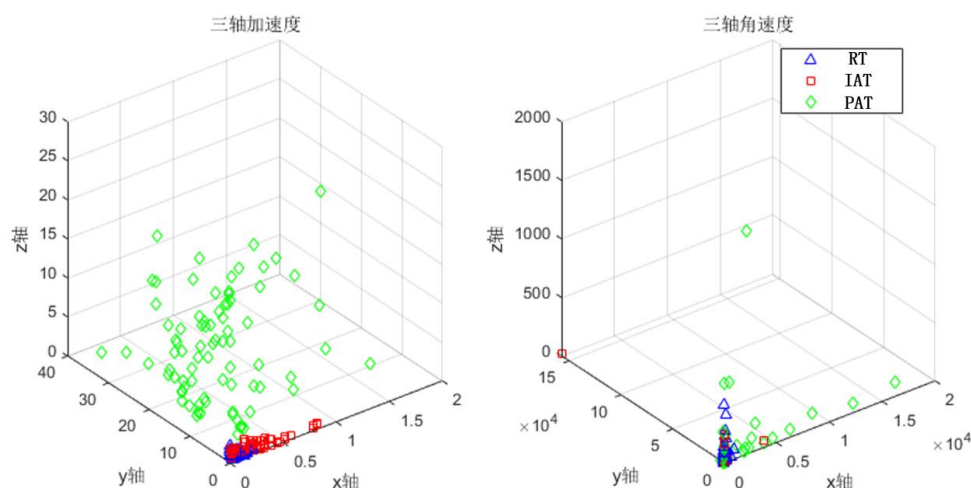


图 4-17 不同震颤类别的三轴分布图

在图 4-17 中, 在加速度三轴构成的立体空间中, RT、IAT、PAT 分布出现分层, 但同样无法通过该特征区分不同难度下的 IAT。

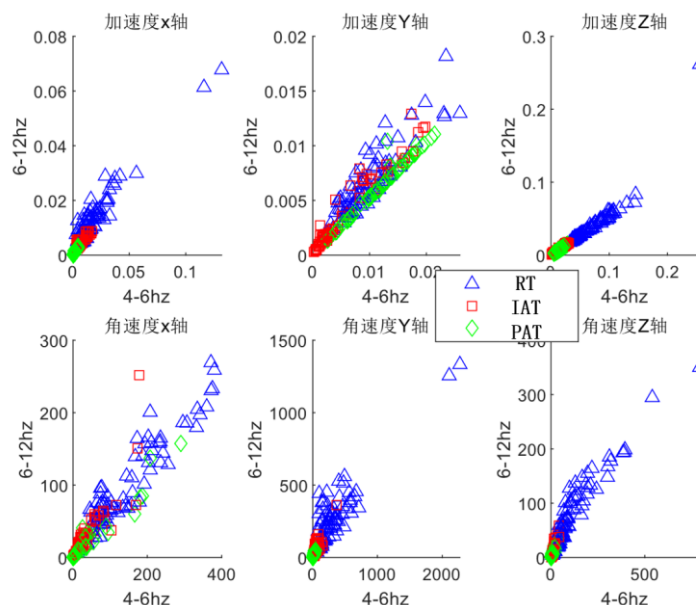


图 4-18 不同震颤类别的 LPSD-HPSD 分布图

为了进一步探究能否通过 LPSD 和 HPSD 进行区分, 本实验通过提取六轴信号的 LPSD 和 HPSD, 本实验以 LPSD 为横轴、以 HPSD 为纵轴, 绘制 L-HPSD 分布图, 如图 4-18。在图 4-18 中, IAT 的 LPSD-HPSD 均值大于 RT 震颤和 PAT, 这表明具有交互意图的震颤行为信号的 LPSD-HPSD 值均高于无交互意图的震颤行为。

通过上述分析可知, 仅通过简单的特征组合或单独特征, 能够区分出 RT、IAT、PAT, 本实验结合 4.3.3 节提出的特征集合, 分别从六轴中提取特征, 之后进行标准化处理, 消除不同特征之间的尺度差异; 之后针对特征集合求特征之间的斯皮尔曼等级相关系数, 剔除相关系数较高的变量, 特征之间相关系数结果如图 4-19。

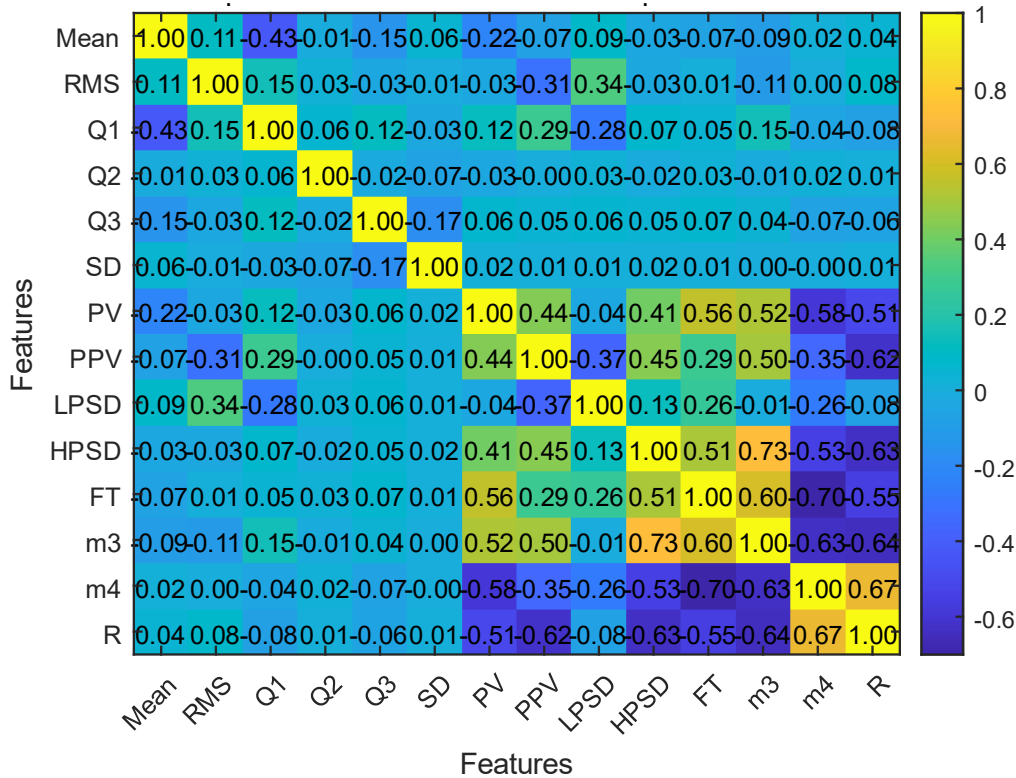


图 4-19 相关系数图

在图 4-19 中, FT 和 m4、HPSD 和 m3 相关系数超过了 0.7, 相关系数绝对值越接近 1, 说明变量间的相关性越强, 综合考虑, 在研究之中剔除 HPSD 和 R 两个特征变量, 之后使用 SVM、KNN、Decision Tree 和 Naive Bayes 进行分类。表 4-6 给出了 RT、IAT、PAT 分类的准确率、F1 分数和召回率。

表 4-6 区分 RT、IAT、PAT 结果

分类模型	Accuracy	F1 Score	Recall
SVM	0.305	0.289	0.315
KNN	0.837	0.829	0.849
Decision Tree	0.970	0.968	0.974
Naive Bayes	0.551	0.501	0.511

通过实验发现, 在通过震颤特征对 RT、IAT、PAT 分类时, 准确率最高为 97% (Decision Tree), 通过震颤特征对 RT、IAT、PAT 分类具有较高的准确率, 能够很好

的识别三种行为震颤，接下来探究不同交互意图强度与区分 RT、IAT、PAT 是否有影响。

在实验中，根据不同目标区宽度划分为了 3 个难度的交互任务，难度系数越高的任务用户越专注，认为用户具有更高的交互意图，统计不同难度任务的执行时间如图 4-20 所示。

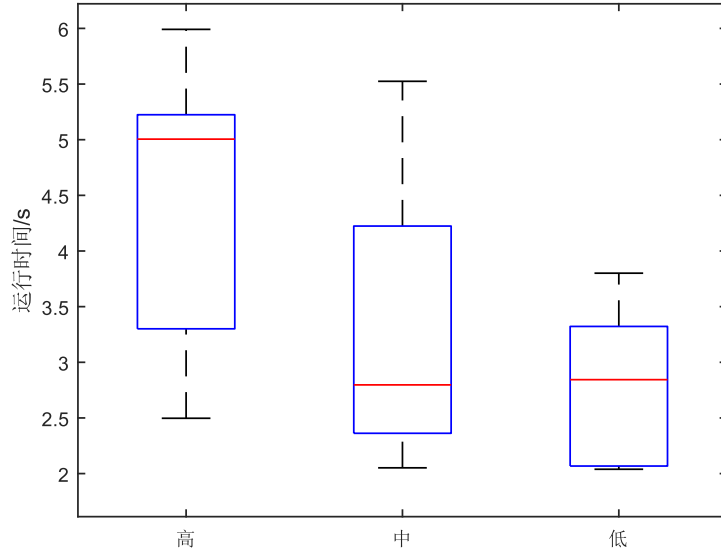


图 4-20 不同任务难度下平均用时箱线图

在图 4-20 中，图中分别为任务难度系数高、中和低下每名用户的平均执行时间，时间均值分别为 4.36s、3.29s、2.82s。使用 Lilliefors 方法进行正态性检验，用户每种难度下的执行时间均符合正态分布，使用 ANOVA 进行显著性检验。实验结果展示，不同任务难度对实验用时有显著影响($F(2,24) = 4.84, p < 0.05, \eta^2 = 0.28$)。组内测试的对比结果显示平均用时在难度高与难度中之间没有表现出显著性差异($F(1,16) = 3.40, p = 0.08$)，难度高与难度低之间则表现出显著差异($F(1,16) = 11.22, p < 0.01, \eta^2 = 0.41$)。

综上所述，任务难度越高则用户所需时间均值越长，这表明不同难度的交互任务对执行时间有影响，在本实验中，认为用户在困难的任务中具有更加强烈的交互意图。为了进一步探究不同交互意图强度对于通过震颤区分用户行为是否有不同的表现，本研究基于实验设计中的三种不同强度交互任务，利用决策树 (Decision Tree) 进行区分，计算其准确率、F1 分数及召回率，所得结果如图 4-21 所示。

在图 4-21 不同强度交互意图下通过震颤识别用户行为中，准确率、F1 分数、召回率均超过了 90%，其中低交互意图相较于其他交互意图分类结果表现最好，但交互意图强弱与分类准确率之间并未表现出正相关或负相关。

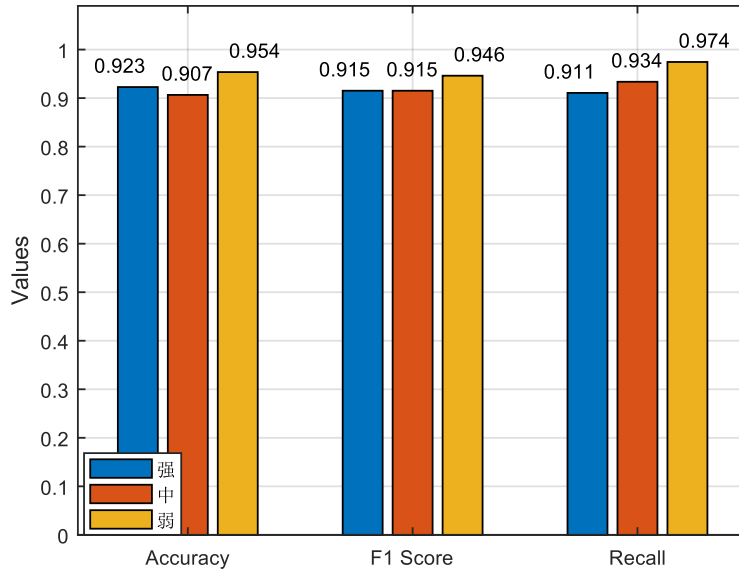


图 4-21 不同交互意图强度下区分 IAT、RT、PAT 结果

综上所述，通过震颤信号的特征，区分 IAT 与无交互意图震颤信号具有较高的准确率 97%，而对于不同交互意图强度对于区分 IAT 与无交互意图震颤信号中，发现低意图强度下分类效果最好。因此通过用户行为能够区分用户有交互意图和无交互意图的震颤，通过用户震颤识别交互意图，进而为解决 Midas Touch 问题提出了一种可行的解决方案。

4.4 本章小结

本章旨在得到用户的手部姿态后，利用手部姿态进行手势交互，并完善手势交互中的相关问题。首先，利用以用户为中心的方法设计实验，实验过程中采用大声思考的方式获取用户对手势的设计想法，最终建立了基于姿态的手势集合。这可以避免现有手势集合无法满足本研究的平面手势交互需求和减轻设计者主观偏见的问题。其次，针对用户设计的手势集合，研究了一种坐标映射方法和手势识别方法。坐标映射方法为用户反馈的手掌偏航转动控制光标时转动角度范围小的问题提出了一种解决方法；手势识别方法则通过决策树利用手指关节特征进行分类。最后，本文分析了健康人群肌肉震颤信号与交互意图之间的关系，设计实验采集不同行为下的震颤特征，并利用多种分类器进行分类，区分 RT、IAT、PAT 时，Decision Tree 分类器达到了 97% 的准确率，在不同交互意图强度下区分 IAT 时，准确率也均超过了 90%，这表明通过震颤信号识别交互意图具备可行性，进而通过识别用户行为是否具有交互意图时，系统选择是否进行功能映射，这为空中手势交互系统的 Midas Touch 问题提出了一种解决思路。

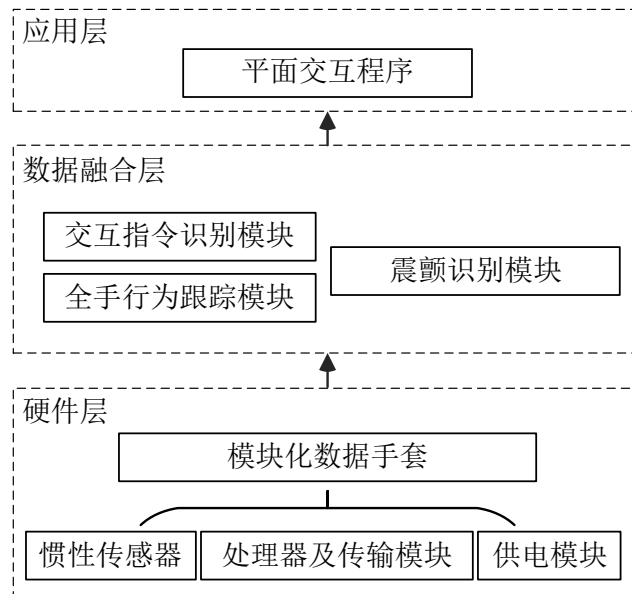
5 iGestureGlove 原型系统设计与实现

本章研究了一种自研数据手套及空中手势交互原型系统 iGestureGlove。在第四章中，本文研究了一种用户提出的手势集合及实现方法并探究了震颤信息与用户行为意图之间的联系，因此本章基于前文的研究，设计并实现了一个空中手势交互原型系统 iGestureGlove。

本章首先介绍了 iGestureGlove 原型系统的总体设计方案，包含系统硬件和软件两部分，在系统硬件部分中介绍了一种模块化的数据手套，在软件部分介绍了本原型系统的软件结构；其次对本原型系统的实现进行进一步的介绍和展示，并介绍了本原型系统的应用场景。

5.1 iGestureGlove 的总体设计方案

iGestureGlove 手势交互原型系统从逻辑上分为硬件层、数据融合层、应用层三部分，iGestureGlove 系统框架如图【】所示。



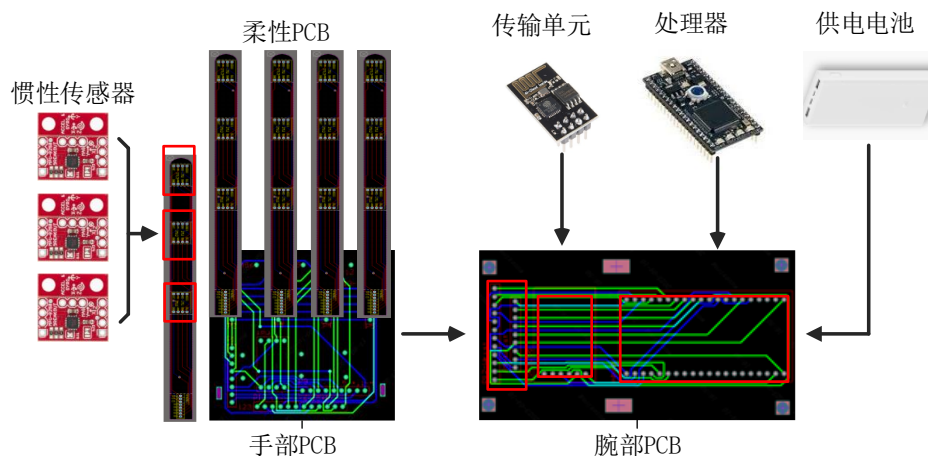
图【】 iGestureGlove 系统框架图

在图【】中，硬件层作为系统的输入载体，所用模块化数据手套由惯性传感器、处理器、传输元件等设备组成，捕捉用户手部的详细运动数据，这为上层的手势跟踪和识别提供了基础数据。数据融合层由交互指令识别模块和震颤识别用户意图模块两部分组成，其中交互指令识别模块基于硬件层传感器的数据，采用数据融合算法，实时跟踪全手在三维空间中的行为，并基于所得姿态进行交互指令识别；识别用户意图模块则根据

行为震颤特征识别用户行为是否包含交互意图。应用层的交互程序设计了多种平面交互任务，通过数据融合层提供的用户姿态及交互指令实现交互目的。

5.1.1 原型系统硬件设计

iGestureGlove 原型系统的硬件部分为一种自研的模块化数据手套，其由惯性传感器、无线传输模块、处理器、供电设备组成，原型硬件组成设计图如图【】所示。模块化指将数据手套从物理层与逻辑层划分为若干个可扩充、可复用、抽象化的小模块。在本数据手套中，将惯性传感器作为最小单元，集成到单条柔性 PCB (Printed Circuit Board) 中，实现手指模块的设计，惯性传感器数目可以根据需求以及手指关节数目差异进行自主选择；其次将手指所附着的柔性 PCB 集成到手部 PCB 中，实现整手模块的设计，柔性 PCB 数目可以根据系统构型及功能需求自主选择；最后设计腕部 PCB，集成无线传输单元与处理器并与整手模块连接。



图【】 原型系统硬件组成设计图

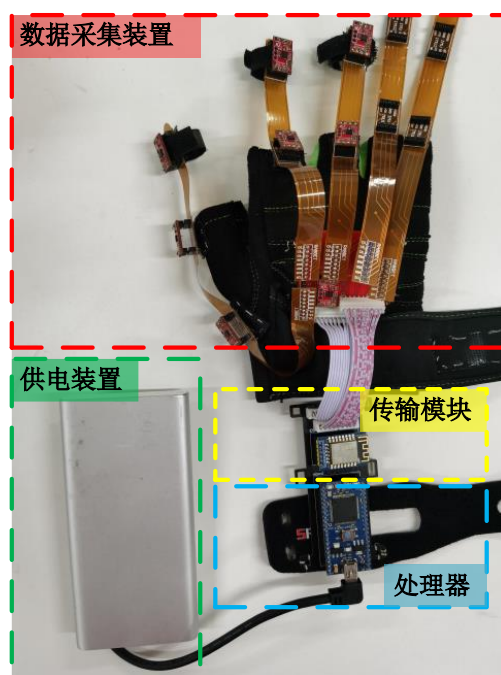
在图【】中，惯性传感器功能为采集附着骨节的运动信息；处理器则收集惯性传感器数据并整合为手部运动原始信息，进行数据预处理并输送到传输单元中；无线传输单元将手部原始信息通过 TCP/IP 网络串口发送到终端中。

iGestureGlove 原型系统的硬件部分所采用的器件清单如表【】所示。

表【】 数据手套器件清单

名称	型号	功能
惯性传感器	SparkFun MPU-9250	获取手部运动信息
PCB	柔性板、FR-4 双层板	连通元件并提供模块化接口
处理器	mbed NXP LPC1768	读取并预处理传感器数据
传输单元	ATK-ESP8266	将处理后数据发送至终端
供电设备	18650 锂离子电池	为数据手套供电

在表【】中，惯性传感器采用型号为 MPU-9250 的九轴传感器，这是一款融合了三轴加速度计、三轴陀螺仪和三轴磁力计的集成式惯性传感器模块，该传感器具有测量范围广、输出速率高、功耗低以及尺寸小等优点，通过附着骨节上提供手部运动信息。集成惯性传感器的柔性 PCB 和手部腕部 PCB 则是自主设计，分析手部构型以及惯性传感器接口，提供一组可复用的单指以及全手连接装置。处理器采用的是 mbed 微控制器，具有低功耗和高性能计算能力等优点，通过 SPI 通信协议与惯性传感器进行数据传输，对原始数据进行预处理操作。传输单元采用型号为 ATK-ESP8266 的 WIFI 通讯模块，具有性能稳定、接口丰富且成本较低的优点，通过 TCP 协议与终端实现数据通信，将手部运动信息从数据手套发送到计算机中。供电设备则采用由 18650 锂离子电池组成的便携充电设备中，该模块能够为数据采集装置提供稳定的 5V 电压，满足工作过程的能量需求。原型系统所用模块化数据手套具有低耦合、高复用、拓展性强、便携的特点，能满足不同交互环境的需求，实物图如图【】所示。



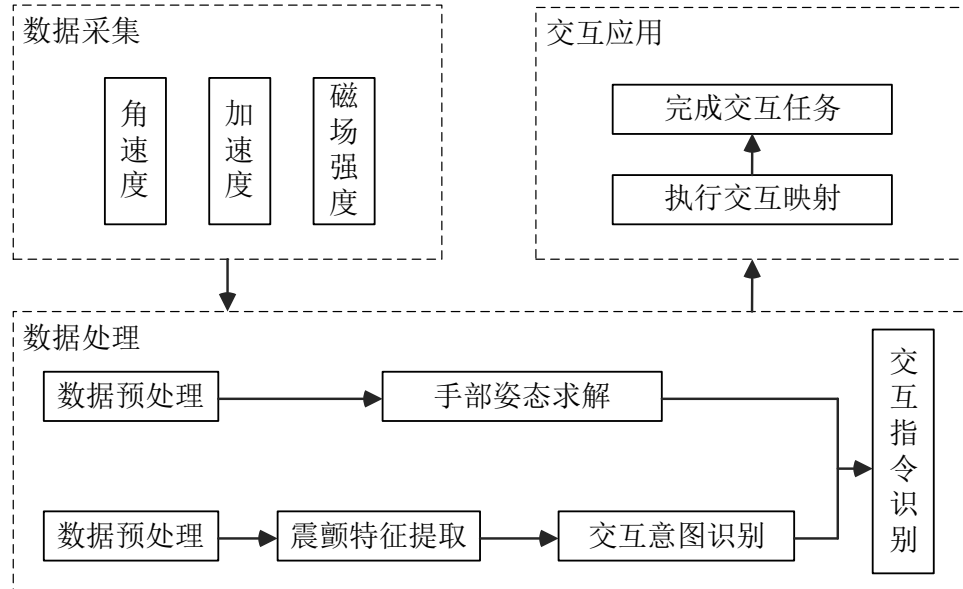
图【】 iGestureGlove 模块化数据手套

5.1.2 原型系统软件结构

iGestureGlove 主要包括数据采集模块、数据处理模块及交互应用模块，其具体结构如图【】所示。

iGestureGlove 原型系统结构中的数据采集层部署在微控制器中，所用开发语言为 C++ 语言，需要根据需求预定义柔性 PCB 及惯性传感器数目，使用 SPI 通信协议获取惯性传感器的角速度、加速度及磁场强度信息，并使用 TCP 通信协议通过网络串口传输到计算机中。数据处理层和交互应用层在计算机中实现，数据处理层中手部姿态求解算法

为第三章内容，震颤特征提取以及交互意图识别则使用第四章中模型识别，交互指令映射与交互手势识别则通过第四章的决策树模型实现，交互应用层所用开发语言为 C# 语言，开发平台为 Visual Studio，交互应用层通过虚拟串口获取数据处理层的数据，提供了展示手部姿态、执行交互任务的功能。



图【】 原型系统软件结构图

iGestureGlove 原型系统软件的开发环境如表【】所示。

表【】 软件开发环境清单

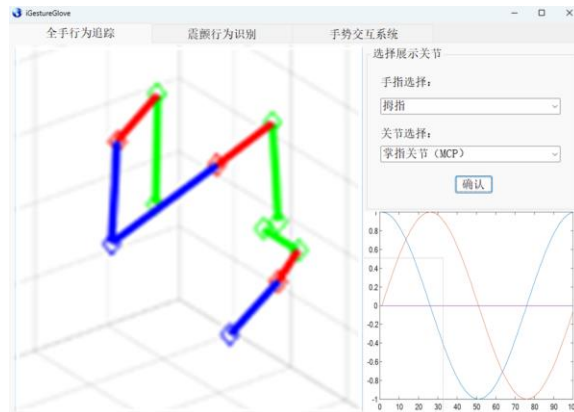
开发环境	版本
操作系统	Windows 11, 64 bit
开发语言	C++, C#
开发工具	Keil Studio, Visual Studio 2019
依赖开发包	.NET Framework 4.0

5.2 iGestureGlove 原型系统实现与应用

iGestureGlove 原型系统的软件界面实现如图【】至【】所示，主要由全手行为追踪、交互意图识别和手势交互系统三部分组成。用户可以通过该原型系统得出手部姿态的精确信息，通过虚拟手的视觉反馈观察手部状态；此外还可以得出当前行为的震颤信号，通过用户行为震颤特征判定用户行为是否包含交互意图；最后，用户可以执行交互模块中预设的交互任务，实现自然的手势交互。

在图【】中，该部分可将手部姿态以虚拟手的形式展示在界面中，通过佩戴数据手套执行手部动作，虚拟手同步展示在界面左边区域，同时还可以令用户自主选择关节查看关节姿态。iGestureGlove 原型系统通过提取震颤特征以实现用户意图检测，选用了第

四章中准确率最高的决策树分类模型进行识别，判断当前阶段用户行为是否为包含交互意图。



图【】 虚拟手及关节姿态界面

iGestureGlove 原型系统提供的交互模块则允许用户通过手势完成交互任务，系统实现的交互指令包含移动、选择、抓取、旋转和缩放五类交互功能，使用第四章中以用户为中心的手势集合，通过映射函数实现对光标的平移控制，通过决策树模型实现手势的分类，为手势交互提供基础支持。

基于上述研究，本系统在交互模块中设计了四种交互任务，分别是选择指定对象、移动对象到指定区域、旋转对象到指定角度和缩放对象到目标大小。交互任务界面如图【】所示。



图【】 平面交互系统界面

传统基于镜像的镜像疗法理论认为通过观察非患病肢体移动可以激活病侧脑区，促进功能康复。但是这种方法存在自身运动感觉反馈不足的缺点，患者难以准确判断肢体移动的细微变化。而 iGestureGlove 系统可以准确获取各手部关节的运动数据，并通过虚拟手显示为用户提供视觉反馈。相比传统镜像疗法，iGestureGlove 系统克服了反馈不足的问题，可以让患者更准确地感知自己的肢体运动。同时，虚拟手的显示也增强了视觉引导作用，有助于激活患病侧脑区，促进运动功能改善。因此，基于 iGestureGlove 系统的镜像疗法可以获得比传统镜像疗法更好的康复效果。



图【】 传统镜像反馈疗法¹与交互系统反馈

图【】展示了 iGestureGlove 原型系统在移动环境下的应用场景。用户戴着数据手套，可以在行走或移动时进行手势交互操作，而不受空间位置的限制。系统可以实时判断用户的手势是否包含交互意图，从而实现流畅的手势交互。这展示了 iGestureGlove 允许用户突破固定桌面环境的限制，实现真正的移动互动体验。相比被动的手势识别交互，iGestureGlove 实现了基于意图的主动交互模式，大大提高了流动环境下的交互效率与用户体验。

5.3 本章小结

本章围绕 iGestureGlove 原型系统设计与实现展开论述，首先探究了一种模块化数据手套的设计方案与实例；其次阐述了 iGestureGlove 原型系统的软件结构，分别包括部署在微处理器的数据采集层和部署在计算机端的数据处理层和交互层；最后展示了 iGestureGlove 原型系统的实现与应用。

6 总结与展望

6.1 总结

6.2 展望

参考文献

[1]