软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software, [doi: 10.13328/j.cnki.jos.000000]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

DistantHands: 一种为大屏设计的远距离范围空中手势交互方法*

张永浩[1,2], 万炎广[1,2], 刘 舫[6], 程 坚[1,2], 刘心阁[3,4], 张 维[1], 马翠霞[1,2], 卞玉龙[5], 周 晓[7], 邓小明[1,2], 刘永进[3,4], 王宏安[1,2]

[1](人机交互北京市重点实验室(中国科学院 软件研究所), 北京 100190)

[2](中国科学院大学, 北京 100049)

[3](清华大学 计算机科学与技术系, 北京 100084)

[4](普适计算教育部重点实验室(清华大学), 北京 100084)

[5](山东大学 机电与信息工程学院, 山东 威海 264209)

[6](媒体融合与传播国家重点实验室 中国传媒大学, 北京 100024)

[7](中国科学院 空天信息创新研究院, 北京 100190)

通讯作者: 邓小明, E-mail: xiaoming@iscas.ac.cn; 刘永进, E-mail: liuyongjin@tsinghua.edu.cn

摘 要: 大屏幕是一种非常有效的信息传播平台,基于大屏幕的多媒体信息交互大大丰富了用户的信息浏览体验.近年来,空中手势识别技术为大屏幕的便捷使用提供了新的交互方式.与已有的工作不同,本文着重探讨距大屏幕较远交互距离范围内的交互式信息浏览任务.为此,我们专门设计了一个空中手势交互系统: DistantHands.为了选择最佳的空中手势集,我们设计了一个新的基于用户偏好的手势诱导阶段,在该阶段用户将实际体验手势以评估其合理性.为了解决手势识别系统中长期存在的无意识手势问题,我们提出了一套新的在线手势识别流程,其中无意识手势预防模块可以过滤掉无意识手势,从而带来更好的用户使用体验.最后, 我们比较了 DistantHands 和触摸交互在大屏幕交互中的用户体验,发现 DistantHands 在舒适性,连贯性,满意度,专注度等指标上都优于触摸界面,这表示我们的新的手势诱导阶段和手势识别流程能够增强用户体验.

关键词: 隔空手势;大屏交互;

DistantHands: Mid-Air Hand Gestures for Interacting with Large Displays at a Distance

Yonghao Zhang[1,2], Yanguang Wan[1,2], Fang Liu[5], Jian Cheng[1,2], Xinge Liu[3], Wei Zhang[1], Cuixia Ma[1,2], Yulong Bian[4], Xiao Zhou[6], Xiaoming Deng[1,2], Yong-Jin Liu[3], Hongan Wang[1,2]

[1](Beijing Key Laboratory of Human-Computer Interactions, Institute of Software, Chinese Academy of Sciences, Beijing, 100190, China)

[2](University of Chinese Academy of Sciences, Beijing, 100049, China)

[3](Tsinghua University, Beijing, 100084, China)

[4](Shandong University, Weihai, 264209, China)

[5](Communication University of China, Beijing, 100024, China)

[6](Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, 100190, China)

Abstract: Large displays provide new information windows in public spaces, and enable rich user experiences for information browsing. Natural interactions with mid-air hand gesture are highly desired to provide easy navigation in large displays. In contrast to the short-distance interaction with large displays in previous works, in this paper, we focus on interactive information browsing task on large displays at a distance. To achieve this, we specifically design a mid-air gesture interaction system. In order to select an ideal mid-air hand gesture set, we design a new gesture elicitation stage based on participants' preferences to select suitable gestures, i.e. user evaluation of gestures based on

* 基金项目: 北京市自然科学基金-海淀原始创新联合基金(No.L232028);国家自然科学基金(No. 62473356);北京市医院管理中心临床医学发展专项经费资助(No.ZLRK202330)

their experience. To deal with the longstanding issue of unintentional mid-air gestures during hand gesture recognition, we propose a new online hand gesture recognition pipeline, which can filter out unintentional gestures through an unintentional gesture prevention module and lead to a better user experience. Furthermore, we compare the interaction performance of our gesture system with touch interface on large displays. In contrast to previous studies showing that mid-air hand gesture is less satisfied than touch, we find that, benefiting from our effective gesture elicitation stage and our new gesture recognition pipeline, our new system outperforms touch to enhance user experience when interacting with large displays.

Key words: Mid-air Hand Gesture, Large Displays Interaction

1 引言

大屏幕在公共空间(如展厅、教室和办公室)中扮演展示多媒体信息的重要角色,公众可以浏览丰富多彩的信息[1]并与大屏幕进行交互.然而如何降低用户交互疲劳已成为一个亟待解决的问题.近年来,空中手势交互作为一种新兴技术被广泛应用于大屏幕交互领域.用户可以通过手部动作实现与大屏幕的直接交互,手势交互逐渐在大屏交互人机协同领域中发挥关键作用[2].与传统输入方式（如鼠标和键盘）相比,空中手势为大屏幕交互提供了一种便捷且高效的替代方案,特别适用于操作那些针对手势特殊设计的用户界面元素[3].与大屏幕常用的触摸交互方式相比,手势交互能支持屏幕布置在更适合观看的位置,允许多人同时观看内容.更重要的是因为用户无需直接触摸显示屏即可完成操作任务,空中手势交互还有助于维护公共卫生.

空中手势交互的优势显著,但同时也面临着几个关键问题.首先,目前尚未形成针对特定交互任务的标准功能手势集合[4],如果未能设计合理的手势集合,用户需要花费更多时间学习才能熟练使用交互系统,同时也更容易产生手臂疲劳;其次,系统识别到不完整或错误的手势会导致无意识手势的出现并触发意料之外的功能,然而现有的空中手势识别方法往往忽略了无意识手势的影响[5–9]; 第三,大多数现有的空中手势系统只能支持短距离的交互,用户在使用时可能会感到操作受限,并且交互控制器的摆放位置也可能限制用户对大屏幕信息的整体理解.例如,现有的 Leap Motion 手势控制器虽然能实现手部姿态跟踪,但其工作距离最多只有 80cm[10];最后,尽管现有研究[3,5,6]表明基于 Leap Motion 的近距离手势交互在用户体验上可能不如触摸交互, 但在允许较大距离范围的手势系统设置下,关于触摸和手势交互系统的对比研究仍然相对缺乏.

与现有工作不同,本文主要研究如何允许用户在大屏幕交互中,通过空中手势在较大且灵活的距离范围内进行操作,希望为大屏交互自然的交互系统[8]提供解决方案. 首先,我们进行了一系列系统化的实验,探索用户在大屏幕交互中的空中手势偏好,并选择最优的手势集合以确保这些手势符合大多数人的操作习惯.由于有必要探索用户在一个新的系统中最自然的手势集合[11],因此无法应用现有的研究[6,12,13]来直接定义手势集合,所以我们提出了一种新的三阶段的手势集合诱导方法来探索用户的偏好.在第一阶段,我们采用 Wobbrock 等人的方法通过问卷调查[14–17]收集所有可能的手势.在第二阶段,我们遵循 Dong 等人[18,19]的研究,让用户筛选出第一阶段产生的所有手势的子集.与现有的两阶段研究不同,本文额外设计了一个基于用户实际体验系统后的手势评估阶段,这个阶段能够有效的提高最终手势集在我们系统中的用户体验.

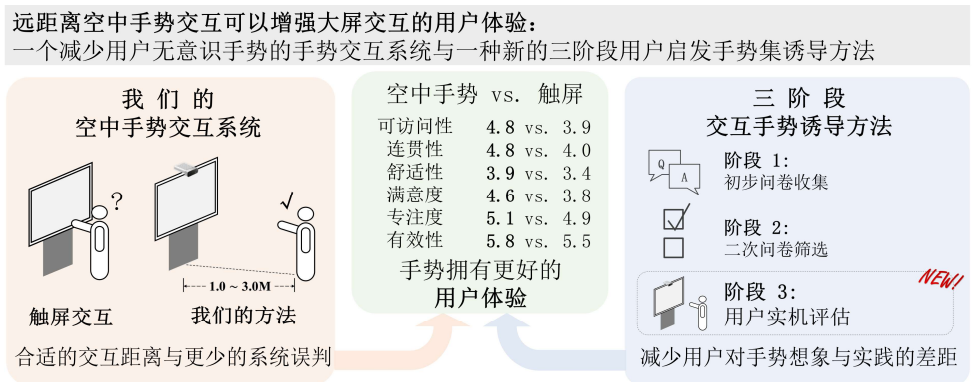


图 1: 本文总览.我们提出了一个全新的空中手势交互系统,能够允许用户在更远的距离操作大屏.

本工作设计了 DistantHands(见图 1), 一个支持在灵活距离范围内操作大屏幕的空中手势交互系统.该系统通过人手分割、手势投票和手势解释器三个模块实现在线手势识别.为了实现更广交互距离下的手势理解, DistantHands 采用了先进的基于视觉的手势采集硬件和神经网络模型,以获取精确的手势识别结果.

为了避免无意识手势带来的问题, 我们提出了两个创新设计.首先,在进行手势识别时,无意识手势预防模块对相邻时间窗口的手势进行投票;同时设计了负采样策略,收集大量的不同于功能手势的负样本(即单帧无意识手势)用于提高正样本的识别正确率.其次,受 Pavlovic 等人[9]的研究启发,他们认为使用滑动时间窗口有助于从时序上将功能手势无意识手势分开,本文提出了一种基于状态机的手势解释模块来减少无意识手势序列对手势识别的影响.与典型的动态手势识别方法[20–28] 不同,即需要预先分割好的手势片段作为输入, DistantHands 避免了计算密集的手势分割过程,同时实现了有效的空中手势识别.与其他基于状态机的方法[29–33]使用视频序列作为输入不同,我们利用精简的手部姿态特征和有限状态机来理解手势.为了缓解用户长时间使用可能产生的手部疲劳,我们特殊设计了用户界面组件,同时应用高效的导航和快捷手势,并设计了有效的指点策略,这些策略都大大减少了用户的大幅度肢体运动和高频次操作.

此外,我们将 DistantHands 与触屏交互在相同的硬件设定下进行用户实验对比,以研究哪种方式可以带来更好的用户体验.本文通过评估主观指标(如易用性、易学性、操作效率、舒适性、可访问性、连贯性、满意度)得出结论.

本文的主要贡献可总结为以下三点:

- 我们提出了一种新的空中手势系统 DistantHands,以实现在较大的交互距离范围内的大屏幕的交互.该系统使用深度学习模型跟踪手部和识别手势,并利用高效的导航、快捷手势和指点策略来减少手臂疲劳.为了过滤无意识手势, DistantHands 使用特殊设计的状态机手势解释模块完成在线手势识别;
- 我们设计了一个新的基于用户的手势集诱导阶段,即基于用户实践的手势评估阶段,借此消除用户想象中和实践之间的经验差距,并得出最适用于大屏交互的手势集;
- 通过 DistantHands 与触摸界面的对比试验结果发现,我们的空中手势交互系统在与大屏幕交互时优于触摸界面,增强了用户使用体验.

2 相关工作

我们的工作与大屏幕的空中手势交互、手势诱导研究、三维人手跟踪、手势识别和无意识手势预防相关.

2.1 用于大屏幕的空中手势交互

大屏幕可以展示丰富的多媒体信息,但用户往往需要远离显示屏一定距离才能整体浏览所有显示的内容[12].近年来,空中手势逐渐成为一种理想的大屏交互方式.它具有如学习成本低,交互体验更加自然等优点[1]; 另外,其作为一种无接触式交互能够避免病菌的传播[5].

2.1.1 代表系统

目前已有许多代表性的手势交互系统(见表 1).为了拥有更自然的交互体验,许多系统都使用视觉传感器捕获和理解手势,同时使用基于用户的方法诱导出合适的与功能对应的手势集合.

与本文最相关的工作是 Sluyters 等人[12]的 LUI ,这是一个基于 Leap Motion 体感控制器的交互系统.用户可以使用可定制的手势浏览该系统中的多媒体内容(如照片、视频、文档和地图等).LUI 提出了一种新的基于多媒体对象的手势诱导方法,并且开发了一个完善的手势交互界面.然而,受到 Leap Motion 工作距离相对较近的限制,该系统的用户必须在控制器很近的距离内才能进行交互.这使得用户在交互过程中下意识的保持一种僵硬的姿态,影响了交互的舒适性和自然性[8].

其他工作给出了许多经验性的结论,他们认为手势交互在大屏幕上的使用体验受多种因素的影响:同时使用的用户数量[8]会影响交互的距离;用户界面的设计和操作逻辑会影响用户的上手速度[6,34];手势集合需要简洁[35].对于手势的选择,研究者们在某些功能上达成了一致,如现有系统中的指点和选择功能大多倾向于使用食指,这也与用户的认知背景一致[6,7,12,34,36,37].总之,目前手势交互的体验仍然不太令人满意,因为它的准确性相对较低[12],同时更容易引起疲劳[36].上述结论成为我们设计手势集合和交互系统的先验知识.

表 1: 与大屏幕交互的代表性空中手势系统的比较

代表系统		系统设计		手势交互设计				用户实验	
作者	年份	工作距离	识别依据	交互应用	使用场景	指点策略	手势设计	考虑疲劳	用户体验结论
本文	2024	**0.5-3.5m**	3D 关节	信息浏览	公共区域	食指	3 阶段	是	比触摸好
Sluyters[12]	2022	<0.6m	3D 关节	多媒体	公共区域	食指	上下文	是	一致,连续,自定义
Huang[5]	2020	<0.6m	3D 关节	FPS 游戏	公共区域	虚拟键	虚拟手	否	好但不如触摸
Malik[6]	2019	<0.6m	3D 关节	-*	-	食指	w/o	否	体验依赖界面设计
Ruiz[36]	2015	-	-	导航地图	公共区域	食指	2 阶段	是	低手臂疲劳
Kim[7]	2018	位于房顶	2D 关节	交互游戏	游乐园	食指	w/o	否	房顶布局有特殊性
Ackad[13]	2015	0.5-2m	-	信息浏览	公共区域	w/o	1 阶段	是	手势集要小且易学
Gentile[8]	2020	0.5-2m	-	广告视频	公共区域	w/o	w/o	否	多人旁观需远距离
Sidd[38]	2017	智能手表	-	-	-	w/o	2 阶段	是	-
Nancel[35]	2011	<2m 手套	-	高分图像	实验室	w/o	w/o	是	简单手势反而高效
Vogel[34]	2005	<5m 手套	视听反馈	信息浏览	公共区域	光投	w/o	否	小目标难操作

* “-”表示其论文中未提及.

2.1.2 与成熟交互方式：触摸的比较

现有的供公众使用的交互式显示器大多都支持触摸使用,同时配套有专门为触摸设计的用户界面.但是触摸界面存在一些缺点:触摸交互界面需要方便人们触及,因此显示器的尺寸选用和放置高度就十分受限;或者在卫生要求极高的使用场景中(如手术室),用户不便直接触摸屏幕.

许多工作比较了基于触摸的用户界面和基于空中手势的用户界面,大多数研究表明,触摸界面在操作准确性、操作速度和自然性方面优于基于空中手势界面[3,5,6,39].然而,需要注意的是,大多数研究[5,6]直接采用工作距离较短的 Leap Motion 和其配套的手势交互系统完成手势信号的识别.总之,在大屏幕交互领域,仍然相对缺乏使用其他体感控制器硬件的空中手势界面同触摸界面的比较.

2.2 手势诱导研究

基于用户的手势集诱导方法在手势交互领域十分重要.尽管存在许多现有研究[4,14,18,19],但仍没有针对特定应用的手势使用标准[4].表 2 比较了代表性工作的手势集合产生的阶段.

表 2: 基于用户的手势集诱导方法的对比

方法阶段	Wobbrock 的方法[14]	Dong 的方法[18,19]	我们的方法
阶段 1：初步问卷收集	✓	✓	✓
阶段 2：二次问卷筛选		✓	✓
阶段 3：用户实机评估			✓

Wobbrock 等人[14]的开创性工作提出了用户手势诱导研究,目前已经成为生成用户定义手势集合的主流方法.其核心思想是探索、调查和分析用户在特定交互场景下对手势及手势对应功能的偏好.该方法首先定义一些必须通过手势执行的操作,然后要求用户提出一些更合理的手势,最后根据收集的数据总结所需的手势集合.这种生成手势集合的方法已经广泛应用在许多手势交互界面中,如智能电视[40]、增强现实[41]和智能手机[42].

然而, 因为用户提出的手势仅基于直觉,上述方法尚未能完全获取用户的偏好,因此需要进行更广泛的调查和深入的用户研究.Dong 等人[18,19]提出了一种两阶段调查方法:第一阶段,用户为指定的功能提出合理的手势;第二阶段,用户从第一阶段中所有人提出的结果中再次选择最喜欢的手势.通过这两个阶段,他们获得了更加用户友好的手势集合.

Dong 等人[19]还引入了外部用户的一致性评估和记忆测试,我们认为有必要引入第三阶段.在这一阶段中,我们将让用户在实际交互场景中体验这些手势,并在体验过程中收集更全面的评估指标.我们希望通过这种方式,能够更准确地了解用户在使用这些手势时的真实感受,从而缩小用户想象与实践之间的差距.

2.3 三维人手跟踪

空中手势交互可以通过手部佩戴各种辅助设备(如特制的数据手套[43]、惯性传感器(如任天堂 Wii 遥控器)、肌电图(EMG)传感器[44]等)或直接裸手[45]完成操作.在公共环境(如我们的问题场景)使用手部设备会影响用户参与交互的热情,使用这些设备并不十分合适,因此我们在本文中关注裸手空中交互.用于裸手空中交互的典型手部跟踪设备包括 Microsoft Kinect Windows[46]、Azure Kinect[47]和 Leap Motion Controller[10].Leap Motion 具有两个红外摄像头,可用于跟踪手部骨骼的关节,从而实现手势控制,而两种 Kinect 仅可跟踪身体骨骼关节而无法跟踪手部骨骼关节,因此只能通过身体关节运动近似大致的手掌运动控制,而在交互过程中身体关节的运动较快,且关节估计误差较大,因此对于精细的交互(如选择或指点)效果较差.

其次,Leap Motion 相对于 Kinect 摄像头的视域范围较小(Leap Motion 建议的工作距离是 60cm,最大 80cm[10],Microsoft Kinect 是 1.8m[46], Azure Kinect 的操作范围可达 0.5m-3.86m[48]),因此 Leap Motion 主要支持与其近距离的交互场景,如与电脑显示屏或 AR/VR 眼镜的交互,对于需要一定操作距离的大屏幕交互并不理想,也不适宜多人同时使用.近年来,基于深度相机的 3D 手部跟踪取得了很大进展[49–51],但这些先进的手部跟踪模型很少被用于人机交互领域的大屏幕空中手势交互.在本工作中,我们选择了一种先进的商用深度摄像头来捕捉人体和人手的深度图像,然后使用一种手部跟踪模型[49]实现高质量的三维人手跟踪.

2.4 空中手势识别

在计算机视觉领域,空中手势识别已经得到了广泛的研究[20].大多数方法将手骨骼关节序列编码为特征向量,并使用 HMM[21]、RNN[22,23]、CNN[24]或 GCN[25–28]提取手势识别所需的分类特征.大多数方法假设输入的手势序列已经被预先分割好,只需要通过分类预测手势类型.然而实际使用空中手势的系统接收的是未处理的输入流,因此理想的手势识别模型应该同时支持手势的识别和分类.为了解决这个问题,Molchanov 等人[22]提出了一种在线手势分割和识别方法,使用滑动窗口技术确定动态手势何时开始.这种方法的问题是计算量大且容易出错,同时因为使用了计算量大的三维卷积神经网络,缺乏关键的手部关节特征,这可能导致识别结果不准确.在人机交互领域,有限状态机(FSM)[52–55]较为常用, 它根据手势变换的连续性来确定手势的输出结果,可以同时分割和识别手势,同时具有良好的实时性[29,31,33].但如表 3 所示,现有基于 FSM 的空中手势识别方法主要直接使用后续计算量很大的图像或视频作为输入,并且这些方法几乎没有利用与手势类型高度相关的精简的手部关节特征.

表 3: 已有的基于 FSM 的空中手势识别方法的对比

方法	年份	FSM 的作用	手势特征输入
本文方法	2024	预防无意识手势	3D 关节点
Tofighi[29]	2017	预防无意识手势	直方图特征
Chen[30]	2010	手势状态转移	Adaboost
Bhuyan[31]	2011	手势序列分割	模糊规则
Tsai[32]	2007	手势序列分割	直方图特征
Graetzel[33]	2004	手势序列分割	纹理特征
Yeasin[54]	2000	建模动态手势	直方图特征

2.5 无意识手势预防

在用户与空中手势交互系统交互的过程中,并非所有手势都需要系统响应.这些不希望产生任何交互效果的手势被称为无意识手势[9,56].例如,用户可能会不可避免地做出与交互目的无关的手势(可能由于疲劳、用户不熟悉手势或注意力不集中导致),这可能会引发系统不必要的响应,从而影响交互的效率和用户体验.因此,预防无意识手势对于空中手势识别和交互至关重要.

根据手势交互的前沿综述[9],确定手势区间可能有助于在时序上将功能手势与其他无意的手部/手臂动作区分开,即防止无意识手势的产生.目前防止无意识手势的方法通常依赖于用户明确的触发指示,例如特殊指定的触发手势、长时间的停留或使用额外的设备[57]来启动和结束手势.这些方法能够提供高质量的手势分割和识别性能.然而,这些额外的要求可能会让用户感到困惑的同时也更会产生疲劳.另一种防止无意识手势的方法是采用后处理策略[22],即利用滑动窗口对每帧的手势进行投票.尽管这些方法能够获得稳定的结果,但它们需要对多帧同时进行计算,因此增加了计算成本.

与现有工作相比,我们接受单帧的输入和有限状态机来处理时间上下文信息,这样避免了复杂且困难的手势序列分割并可以实现实时的手势解析.

3 交互任务定义

在本节中,我们定义了典型的适用于大屏幕的空中手势交互任务.一个理想的用户界面应当简化交互的复杂性,并引导用户聚焦于交互内容本身.为了使用户能够更快地适应新的操作方式,并与大屏幕实现自然交互,我们的交互任务设计受到了智能手机操作逻辑的启发,如图 2 所示,包括"点击"、"左右移动"、"上下移动"、"取消/返回"以及"语音输入".大多数与大屏幕的交互任务(除了搜索输入)都可以通过前四种手势类型完成,而"语音输入"作为一种补充方式,旨在满足用户的搜索需求.

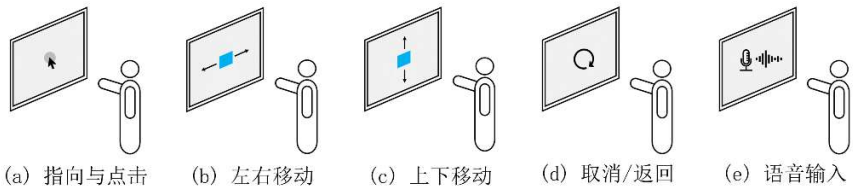


图 2: 我们设计了五种适合用于大屏的空中手势交互任务.

任务 1：点击.该任务是通过保持特定手势在空中移动(通常是食指的指点手势)来移动光标并完成点击.在图形界面中,光标的移动和点击是基本操作,这个任务同时也与智能手机中的触摸操作类似.

任务 2 和 3：左右移动和上下移动.这个任务用于移动导航栏和其他组件,可以提升浏览长幅面的水平和垂直网页的体验.这个任务类似于鼠标滚轮的功能,也类似于用手指在智能手机屏幕上滑动来平移页面的操作.

任务 4：取消/返回."取消/返回"是智能手机用户界面中的一个基本操作.由于需要经常使用,我们将其设计为一个快捷命令手势.因此,用户可以直接使用"取消/返回"功能,而不是使用"点击"来与"取消/返回"按钮交互.

任务 5：语音输入.为了满足用户的搜索需求,摆脱使用手势完成复杂的键盘输入,我们设计了语音输入任务,并将其作为一个快捷命令手势.有了这个功能,用户可以快速提出问题并进行搜索.执行"语音输入"后,系统会关闭麦克风并处理录制的语音片段.

4 空中手势集合设计

本节提出一种三阶段的手势诱导方法,用于筛选更适用于大屏幕交互的空中手势.与先前工作[18,19]中使用的两阶段研究不同,我们设计了一个新的阶段,即用户在实际系统中体验后对手势进行再次评估,这有助于减少

用户主观偏好与实际系统体验之间的差距.

4.1 第一阶段: 问卷收集用户设计的手势

为了给用户提供出色的体验,我们通过问卷调查用户对手势和功能映射之间的偏好.我们首先收集用户的使用习惯,以了解用户认为哪些手势更易于使用.被试需要描述他们将在第 3 节中的四个交互任务上(即图 2 所示的"左右移动"、"上下移动"、"取消/返回"和"语音输入")使用什么手势.

被试情况简介.我们收集了 25 份问卷(15 名男性,10 名女性;大多数被试年龄在 18-30 岁之间;被试有不同的文化、职业和教育背景),根据收集到的典型手势总结了问卷结果,总结如下.

左右移动任务.大多数答案是"摆动手表示左/右"(25 份问卷中有 11 份,占比 44%,图 3(a)).其余的答案是"挥动食指表示左/右"(图 3(b))、"挥动拳头表示左/右"(图 3(c))、"使用拇指方向表示左/右"(图 3(d))、"挥动食指和中指表示左/右"(图 3(e))、"使用手部位置表示左/右"(图 3(f))和"挥动拇指表示左/右"(图 3(g)).

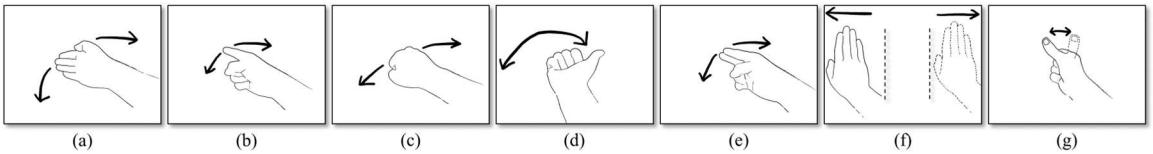


图 3: 用户为"左右移动"任务提出的可能的手势示意图.

上下移动任务."上下移动"的答案与"左右移动"类似.大多数答案是"摆动手表示上/下"(36%,25 份问卷中 9 份,图 4(a))和"挥动食指表示上/下"(16%,25 份问卷中 4 份,图 4(b)).其余的是"使用掌心位置表示上/下"(图 4(c))、"挥动拳头表示上/下"(图 4(d))、"挥动食指和中指表示上/下"(图 4(e))和"挥动拇指表示上/下"(图 4(f)).

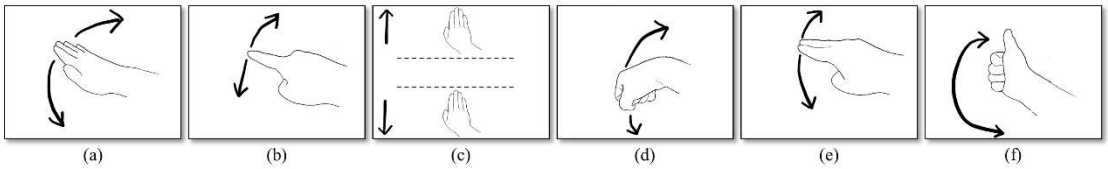


图 4: 用户为"上下移动"任务提出的可能的手势示意图.

取消/返回任务.由于"取消/返回"任务是一个频繁使用的交互任务,我们要求被试使用静态手势,因为静态手势简单且效率更高.本任务答案的分布相对平衡,没有出现有明显共性的手势.答案包括"掌心向前静止"(图 5(a))、"拇指向左"(图 5(c))、"拳头静止"(图 5(f))、"竖起食指"(图 5(g))和"食指指向左"(图 5(d)).

语音输入任务.与"取消/返回"任务类似,我们也要求被试使用静态手势.答案包括"拇指向左"(图 5(c))、"拇指竖起"(图 5(b))、"拇指向右"(图 5(e))、"拳头静止"(图 5(f))、"竖起食指"(图 5(g))、"食指指向左"(图 5(d))、"拇指和小指竖起"(图 5(h))、"拇指触碰食指"(图 5(j))和"张开手掌,手指弯曲"(图 5(i)).

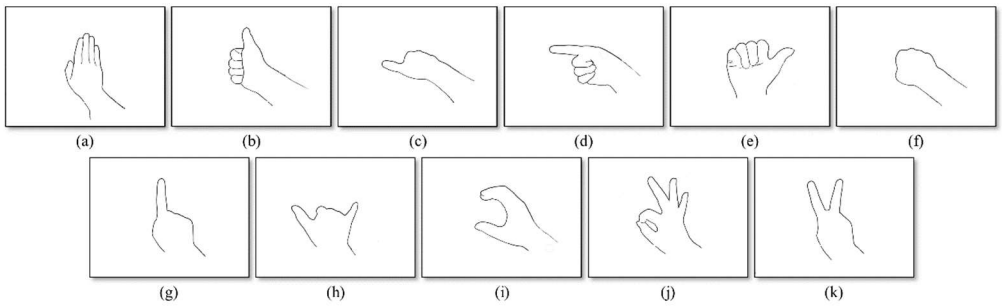


图 5: 用户为"取消/返回"任务或"语音输入"任务提出的可能的静态手势.

4.1.1 "点击"任务

伸出食指是最常见的手势[58],该手势对于各种个人习惯或语言环境的人都很容易理解并执行.在 2.1.1 节中,我们发现食指手势是最常用于指点和选择的手势,所以遵循这个惯例,我们使用食指作为引导和点击鼠标指针的手势.

我们为选定的食指手势设计了两种触发点击功能的动作.第一种是保持食指手势并靠近屏幕一段距离,系统会触发点击动作(图 6 左侧),在此称之为"基于距离的点击"(DISTANCE).第二种是将手指移动引导鼠标移动到要点击的位置并保持一段时间,系统就会触发点击动作(图 6 右侧),在此称之为"基于时间的点击"(TIME).在第一种操作中用户完全主导操作,点击操作可以快速完成;在第二种操作中用户只需手指指针停留一段时间(例如 1 秒),可以获得良好的指点精度.

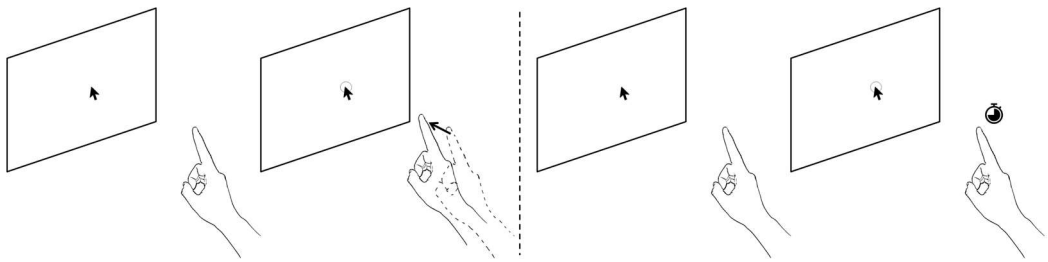


图 6: 保持伸出食指手势完成点击功能的两种触发方式.左侧："基于距离的点击"(DISTANCE),即保持食指手势并靠近屏幕一段距离；右侧："基于时间的点击"(TIME),在需要点击的位置保持静止一段时间.

4.2 第二阶段: 问卷收集用户最喜欢的手势

在第一阶段获得用户提出手势的集合后, 我们执行第二阶段：通过问卷调查用户对集合内的手势的偏好,被试需要重点考虑手势的自然程度.

如 4.1.1 节所述,我们为"点击"任务保留伸出食指的手势.因此为了防止出现冲突,我们在第二阶段的问卷中删除了其他 4 个任务中的伸出食指手势(图 3(b)、图 4(b)、图 5(d)、图 5(g)).因此,在"左右移动"任务中有 6 个手势,在"上下移动"任务中有 5 个手势(见 4.1 节描述).由于"取消/返回"和"语音输入"任务都采用静态手势,我们将用户设计的手势和其他常见手势合并,总共有 9 个手势,如图 5 所示.

在总共收集的 32 份问卷中(17 名男性,13 名女性,2 名匿名;大多数年龄在 18-25 岁之间;参与者有不同的文化、职业和教育背景),结果如图 7 所示.手势的顺序首先依据"不自然"的评分由低到高排序,然后依据"自然"评分高到低排序.结果说明手势之间用户体验存在明显差异.我们从排序结果中挑选出几个得分排名显著较高的手势进行下一步研究.

左右移动.我们选择了前 3 个手势:"挥动掌心表示左/右"(PALM-S)、"使用掌心位置表示左/右"(PALM-M)和"挥动食指和中指表示左/右"(INDEX-MID).

上下移动.我们选择了前 3 个手势:"挥动掌心表示上/下"(PALM-S)、"使用掌心位置表示上/下"(PALM-M)和"挥动食指和中指表示上/下"(INDEX-MID).

取消/返回.我们选择了前 5 个手势:"拇指向左"(THUMB-L)、"掌心向前静止"(PALM)、"拇指向右"(THUMB-R)、"拳头静止"(FIST)和"拇指竖起"(THUMB-U).

语音输入.我们选择了前 4 个手势:"拳头静止"(FIST)、"掌心向前静止"(PALM)、"拇指竖起"(THUMB-U)和"拇指触碰食指"(THUMB-INDEX).

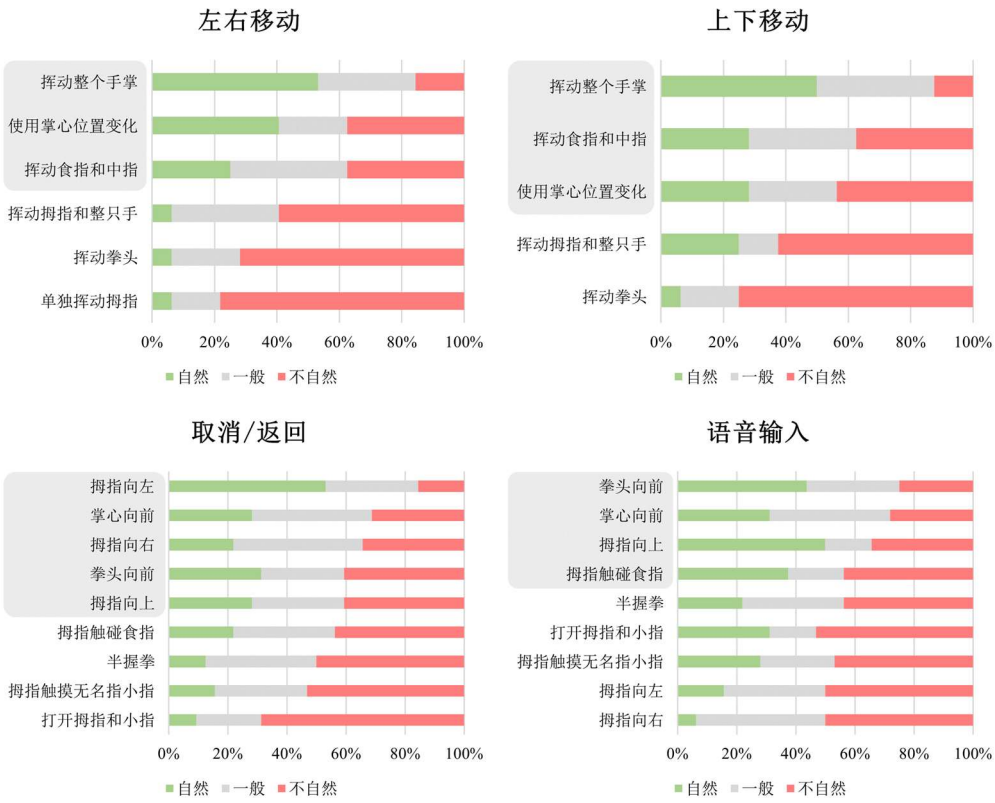


图 7: 第二阶段问卷的用户偏好分析.

4.3 第三阶段: 候选手势的评估

我们发现,被试在体验实际交互系统前提出的手势与他们想象中的使用体验可能存在很大差异.因此在本节将描述如何评估前两个阶段结束后候选的手势.我们首先将所有任务抽象成几个简单的交互任务,同时设计一个系统来实现上述空中手势任务的检测和识别,然后让被试通过该系统体验每个任务所有可能的手势,并使用主观评估和客观指标对其进行评估,选出手势系统中最终使用的手势集合.

本阶段包含五个具体任务,对应于我们手势系统的五个功能.这些任务旨在测试每个手势的准确性和效率.对于每个手势和相应的任务,参与者将有 1 分钟的时间学习操作,随后完成 10 次任务.之后,参与者将从 7 个方面对该手势进行评估(准确性、易学性、易用性、眼疲劳、手疲劳、速度和综合体验).本阶段被试总共测试 15 个手势(包括伸出食指),每位被试将测试$15 \times 10 = 150$次.对于相同任务中不同的手势,我们使用拉丁方设计(Latin square design)改变顺序,以避免数据的偏差.

4.3.1 任务设计

点击任务.对于"点击"功能,我们设计了一个随机在全屏幕上任意位置生成固定大小圆形的页面.被试需要使用不同的指点策略点击圆形.如果成功在圆形内点击,则记此次点击动作成功,并刷新页面生成一个新的随机位置的圆形.如果点击在圆形外,则该次点击动作失败,参与者需要重试,直到点击在圆形内.每位被试完成 10 次成功点击后,记录整个过程的时间.

左右移动和上下移动.对于"左右移动"功能,我们设计了,可以生成 2 个圆形(一个大,一个小,大小固定)的页面.在像素坐标系内,这两个圆形的 y 坐标值相同,x 坐标值有差异.被试可以使用不同的手势左右移动小圆形.任务的目标是将小圆形移动到大圆形内.为了测试手势的灵敏度和准确性,被试可以自主确定小圆形的位置直到满意.本任务将记录完成任务消耗的时间和两个圆心之间的距离.

"上下移动"任务与此类似,不同之处在于两个圆形的 x 坐标值相同,y 坐标值有差异.被试可以使用不同的手势上下移动小圆形.

取消/返回任务.对于"取消/返回"功能,我们设计了一个中间显示"取消/返回"字样的页面.被试需要使用不同的手势来响应.当系统检测到正确的手势时,页面会显示"成功"字样,同时被试需要放下手臂并放松手部.然后界面会显示一个随机的时间间隔(3 到 5 秒),这个过程总共重复 10 次.我们记录了每次操作的消耗时间.

语音输入任务."语音输入"功能包括 3 个步骤:"开始语音输入"、"按住说话"和"结束语音输入".我们设计了一个中间显示"开始说话"字样的页面.被试需要做出不同的手势来模拟语音输入.检测到正确手势后,手势系统会显示倒计时(随机 3 到 5 秒).被试需要一直保持手势,直到倒计时结束,屏幕也会提示停止手势.如果被试提前停止手势或摆出错误的手势,则该操作将被视为失败.本任务记录每次操作的消耗时间(扣除要求的随机保持时间).

4.3.2 结果

该实验邀请了 16 位被试(其中包含 10 男 6 女,平均年龄 24.7 岁,标准差 3.96)参与实验.对于每个任务中的不同手势,被试有 1 分钟的时间学习并熟悉该手势,随后完成任务.

表 4: 5 项交互任务中的 17 种手势的主观评分结果(均值±SEM).

	左右移动			上下移动			点击	
	PALM-S	PALM-M	INDEX-MID	PALM-S	PALM-M	INDEX-MID	DISTANCE	TIME
1 准确性	**4.8 ± .1**	4.2 ± .3	4.6 ± .2	**4.6 ± .2**	4.4 ± .2	**4.6 ± .2**	4.5 ± .2	**4.8 ± .1**
2 易学性	**4.6 ± .1**	3.9 ± .3	4.3 ± .2	**4.3 ± .2**	3.6 ± .4	4.2 ± .2	4.2 ± .3	**4.6 ± .1**
3 易用性	**4.7 ± .1**	4.4 ± .2	4.6 ± .1	4.4 ± .2	4.3 ± .3	**4.5 ± .2**	**4.6 ± .2**	**4.6 ± .1**
4 眼疲劳	**4.2 ± .2**	3.8 ± .3	3.5 ± .2	**3.7 ± .3**	3.5 ± .3	3.6 ± .3	4.2 ± .2	**4.4 ± .2**
5 手疲劳	**4.2 ± .2**	3.2 ± .3	3.8 ± .2	**4.1 ± .2**	3.6 ± .3	**4.1 ± .2**	3.9 ± .3	**4.5 ± .1**
6 速度	**4.1 ± .2**	3.5 ± .2	4.0 ± .2	4.1 ± .2	3.7 ± .3	**4.2 ± .2**	4.1 ± .2	**4.3 ± .2**
7 综合	**4.3 ± .2**	3.6 ± .2	4.0 ± .2	4.1 ± .2	3.7 ± .3	**4.2 ± .2**	4.1 ± .2	**4.3 ± .1**
8 总和	**4.4 ± .1**	3.8 ± .1	4.1 ± .1	4.1 ± .1	3.8 ± .1	**4.2 ± .1**	4.2 ± .1	**4.5 ± .1**

续上	取消/返回					语音输入			
	THUMB-L	PALM	THUMB-R	FIST	THUMB-U	FIST	PALM	THUMB-U	INDEX-MID
1	**4.9 ± .1**	4.8 ± .1	4.7 ± .1	4.7 ± .1	**4.9 ± .1**	**4.8 ± .1**	4.6 ± .3	**4.8 ± .1**	**4.8 ± .1**
2	**4.8 ± .1**	4.7 ± .1	4.6 ± .1	4.7 ± .1	**4.8 ± .1**	4.6 ± .1	4.6 ± .1	**4.8 ± .1**	4.7 ± .1
3	**4.7 ± .1**	4.6 ± .2	4.6 ± .1	**4.7 ± .1**	**4.7 ± .2**	**4.8 ± .1**	4.6 ± .2	**4.8 ± .1**	4.7 ± .2
4	4.6 ± .1	4.6 ± .1	4.6 ± .1	4.6 ± .1	**4.7 ± .1**	4.3 ± .2	4.5 ± .1	**4.7 ± .1**	4.3 ± .2
5	4.8 ± .1	4.8 ± .1	4.7 ± .1	4.7 ± .1	**4.9 ± .1**	4.4 ± .2	4.2 ± .3	**4.8 ± .1**	4.7 ± .1
6	4.7 ± .1	**4.8 ± .1**	4.7 ± .1	4.7 ± .1	**4.8 ± .1**	4.6 ± .1	4.7 ± .1	4.8 ± .1	**4.9 ± .1**
7	**4.6 ± .1**	4.5 ± .1	4.3 ± .2	4.4 ± .1	4.5 ± .2	4.2 ± .2	4.1 ± .3	**4.7 ± .1**	4.3 ± .2
8	**4.7 ± .04**	**4.7 ± .05**	4.6 ± .1	**4.7 ± .04**	**4.7 ± .04**	4.5 ± .1	4.5 ± .1	**4.8 ± .04**	4.6 ± .1

客观指标.我们为五个任务选择了 3 个客观指标(时间、精度和错误率),每个指标都取重复 10 次的平均值并数据进行了统计学显著性分析.结果表明,大多数客观指标没有统计学显著性.但特别的是在"上下移动"操作中,PALM-M 显著慢于 PALM-S 和 INDEX-MID($p < .005$),PALM-S 比 PALM-M 显著更准确($p < .002$),所以我们将通过后续实验结果选择手势.在"点击"操作中,TIME 显著慢于 DISTANCE($p < .002$),但 TIME 显著比 DISTANCE 更准确($p < .001$).我们认为时间快和精度高在这项任务中确实不能同时满足,差异是合理的.在"语音输入"操作中,FIST 显著慢于所有其他语音操作($p < .005$),但这并不十分重要.总之,本阶段随后将主要利用主观评分的结果来选择合适的手势.实验的统计结果见表 4.

主观评分.我们遵循 Surale 等人提出[59]的评分设计,并添加了一个"综合得分"项,以表达被试的总体偏好和感受.参与者需要从易学性、易用性、准确性、速度、眼睛疲劳、手部疲劳和综合得分这 7 个方面对每个手势进行评分.被试将填写 5 分量表进行评估,1 分为最低分(如准确性低、很难学习、非常疲劳),5 分为最高分(如准确性高、容易学习、不疲劳).表 4 展示了所有主观评分的结果.

结果分析.表 5 展示了不同手势的客观指标(时间、精度、准确性).可见大多数手势之间的差异都不显著.由于原始数据并非正态分布,我们使用 Aligned Rank Transform[60]对数据进行转换,然后对转换后的数据进行重复测量方差分析(ANOVAs).在"左右移动"任务的评分中,易学性、手部疲劳、准确性、速度和综合得分的主效应达到显著性($p < .05$).成对详细比较表明,与 PALM-S 相比,PALM-M 的易学性较差($p = .05$)、准确性较低($p < .01$)、速度较慢($p < .03$)且手部疲劳较大($p < .02$),综合得分也比 PALM-S 低($p < .02$),这表明 PALM-M 的整体使用体验比 PALM-S 差.对于"点击"任务,TIME 除了眼疲劳得分外,其他得分都高于 DISTANCE,易用性和准确性的差异显著($p < .04$).对于"语音输入"任务,易用性的主效应达到显著性($p < .05$);THUMB-U 比 FIST 更易使用($p < .04$).对于"上下移动"和"取消/返回"功能,所有手势的主观评分没有显著差异.

表 5: 5 项任务中不同手势的客观指标(均值±SEM).我们可以观察到大多数手势之间的区别并不显著.

	左右移动			上下移动		
	PALM-S	PALM-M	INDEX-MID	PALM-S	PALM-M	INDEX-MID
用时(ms)	6089 ± 757	6479 ± 641	6004 ± 445	4560 ± 160	5446 ± 159	4680 ± 174
错误距离(像素)	21.1 ± 2.2	22.1 ± 2.6	20.8 ± 2.7	18.1 ± 2.5	27.8 ± 3.0	19.8 ± 1.5

	点击		语音输入			
	DISTANCE	TIME	FIST	PALM	THUMB-U	THUMB-INDEX
用时(ms)	34771 ± 2005	41887 ± 1983	2342 ± 76	2175 ± 64	2113 ± 84	2127 ± 48
正确率	0.9 ± .03	1 ± 0	0.93 ± .03	0.89 ± .03	0.97± .02	0.98 ± .02

	返回/取消				
	THUMB-L	PALM	FIST	THUMB-INDEX	THUMB-U
用时(ms)	1659 ± 45	1670 ± 71	1639 ± 43	1767 ± 50	1623 ± 48

最终手势的选择.对于不同任务使用同一个手势会产生逻辑上和系统中的冲突.例如,THUMB-U 是"取消/返回"和"语音输入"功能的最佳手势.为了构建没有冲突的手势系统,我们构建了在不同任务中没有使用相同手势的所有组合,随后根据选定手势的"综合"得分之和对组合进行排序.得分最高的组合是[PALM-S 用于"左右移动"、INDEX-MID 用于"上下移动"、TIME 用于"点击"、THUMB-L 用于"取消/返回"和 THUMB-U 用于"语音输入"].我们将使用这 5 个手势构建手势系统,并进行下一步研究.图 8 展示了手势系统中最终为每个任务选择的手势.

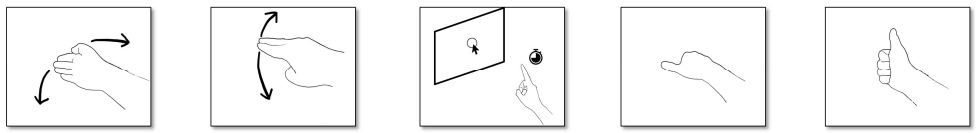


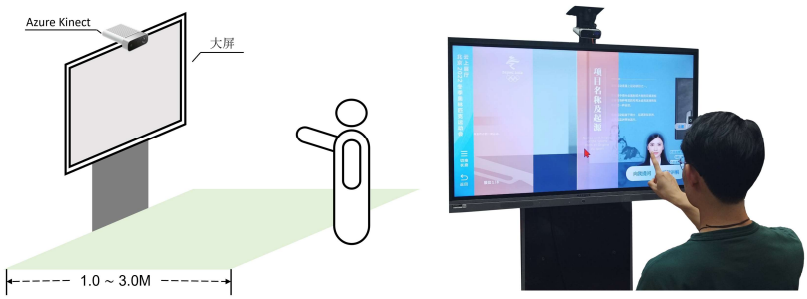
图 8: 完成三阶段实验后我们为手势系统中的每个任务最终选择的手势.

所选手势的优势.基于在实际系统中的手势体验,被试可以获得手势交互的真实体验,同时可以在系统中体验不同的手势集合.在被试明确评价指标后,他们会真实感受到操作中每个手势在交互中的特点并给出建设性意见.正是在第三阶段实验中,我们发现被试在第二阶段认为用于"上下移动"最自然的 PALM-S 并不能带来最佳的用户体验,这肯定了我们提出的第三阶段的价值.

5 手势交互系统设计

5.1 概述

本文提出的 DistantHands 是一个基于空中手势的系统,用户可以使用该系统在一定操作距离范围内与大屏幕进行交互.DistantHands 的设计原则是广泛的支持各种应用,并且只需要简单的学习即可拥有出色的用户体验.第 4 节得出了几个对远距离大屏幕交互友好且有效的空中手势.DistantHands 接受 Azure Kinect 相机[47]同时获取的彩色和深度图像作为输入(图 9(b)),并使用深度学习模型进行手部跟踪,手势被映射为鼠标和键盘功能.我们设计了有效的手部分割模块来提取可靠的手部掩模,同时提出了带有投票模块和负采样策略的无意识手势预防模块,在有效地对手势进行分类的同时避免无意识手势.最后,本系统通过有限状态机利用手势的时序上下文来细化手势预测.DistantHands 可以实现实时手势识别,并降低无意识手势的影响.



（a）硬件设备设置　　（b）真实交互场景

图 9: 系统硬件设备与用户操作的图示.

5.2 硬件设置

DistantHands 使用了一台 55 英寸的商用大屏幕,并在大屏幕上安装 Azure Kinect 来捕捉彩色和深度图像流.系统的硬件设置如图 9 所示.Azure Kinect 可以在室内的各种照明条件下捕捉高质量的彩色和深度图像,其深度传感器的工作距离约为 3 米,足以支持较广的交互距离范围.为了测试手势系统在信息浏览任务中的有效性,我们在大屏幕上设计了一个信息浏览应用程序(在第 6.2 节的"设计和过程"和图 17 中描述相关信息),该应用程序支持用户在交互过程中执行第 3 节中定义的典型交互任务.

5.3 在线手势识别系统

在线手势识别系统的完整工作流程如图 10 所示,其包含人手分割、手势投票和在线手势解释器模块.手部分割模块以 RGB-D 图像为输入,输出手部区域的深度信息.手势投票模块将使用先前的识别结果来平滑当前结果.在线手势解释器模块使用有限状态机处理时序上下文信息,并最终判断手势动作.

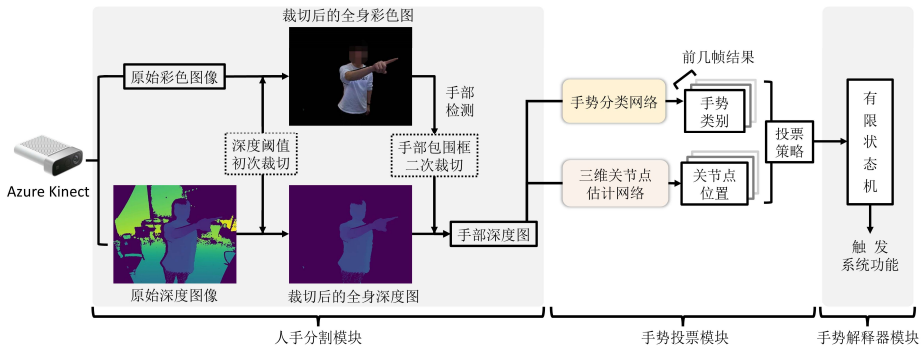


图 10: 我们的在线手势识别系统的工作流程.

5.3.1 人手分割模块

在人手分割模块中,我们使用颜色和深度图像获取交互用手的掩码,并将分割的手部深度输入到后续的手势识别模块.我们首先在自收集的手势数据集(119,274 个样本)上训练的 YOLO-v5[61]目标检测器检测彩色图像中的手部区域.随后为了减少环境背景或人群的影响,使用深度阈值去除与手不相关的图像区域.由于捕获的彩色和深度图像对齐良好,我们使用阈值遮蔽后的深度图像和彩色图像中的手部边界框一起获取初始的手部分割结果.最后遵循 DeepPrior++[62]的方法,即从深度图中提取以手部中心点为中心的固定大小三维立方体来获取手部掩码.除此之外,我们发现用户的两只手通常都会在初步分割中被截取出来,因此还会使用身体骨骼关节的信息来识别主要的交互用手.

5.3.2 无意识手势预防模块

在这个模块中,手势投票模块和负采样策略能够有效地对手势进行分类,并避免无意识手势.

手势投票模块.手势投票模块使用上一个模块分割后的手部深度区域和三维手部关节点同时对初始手势进行分类.手势类别包括"INDEX"、"PALM-S"、"INDEX-MID"、"THUMB"和"NONE",其中"NONE"表示无意识手势,"THUMB"同时表示 THUMB-L 和 THUMB-U(在进行手势预测后通过手掌方向来区分 THUMB-L 和 THUMB-U).我们首先将交互用手的手部深度图输入到基于 ResNet-18[63]的轻量级手势分类网络.图 11 展示了手势分类网络的混淆矩阵.混淆矩阵展示该网络能够有效地对手势进行分类,并区分无意识手势(即矩阵中的"NONE"类).对于像指点这样的细粒度手势任务,本模块通过三维手部关节点的坐标变化识别并达成用户意图.因为 A2J[49]在准确性和效率之间达到了良好的平衡,我们使用 A2J 从分割的深度图像中获取三维手部关节点.为了从"THUMB"中区分 THUMB-U 和 THUMB-L,本模块使用骨骼的角度来判断手的朝向,以区分执行"语音输入"或"取消/返回"任务.

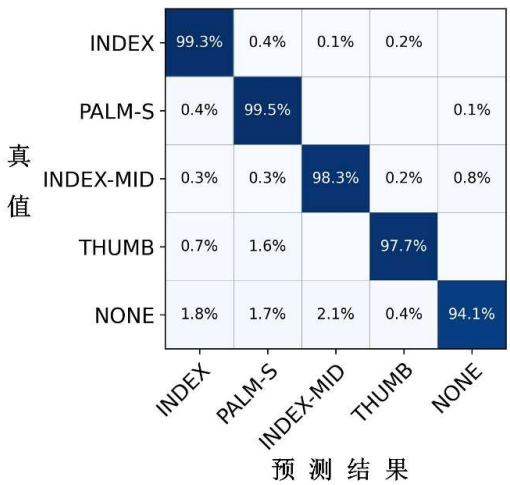


图 11: 手势分类网络分类结果的混淆矩阵.

无意识手势预防模块.为了提高手势系统的稳定性,我们收集了大量不同于系统中已使用手势的负样本手势,并训练了一个基于 ResNet-18 的轻量级手势分类网络以将无意识手势与"有意手势"分开.如果用户操作产生无意识手势,系统将不会响应.投票模块可以通过最近几帧组成的时间窗口中的手势进行投票来实现稳定的手势识别.如果预测的手势类型发生变化,系统将不会立即给出手势预测,而是采用最近几帧(例如 3 帧,对于 20FPS 的识别速度约为 0.15 秒)的手势进行投票,这种策略可以明显减少连续输入手势的交互操作中意外发生的中断(如"语音输入").例如"语音输入"任务需要用户在说话时保持一个特定的手势.如果手势系统此时获取到其他手势,将被视为"语音输入"任务完成,系统会关闭麦克风并处理录制的语音片段,如果分类网络在这个特定手势序列期间出现错误,就会中断"语音输入"任务.

不失一般性的,以只包含一种特定手势的手势序列为例,我们用"R"表示正确分类结果(概率为q),"W"表示错

误分类结果(概率为$1-q$).在我们的手势投票系统中需要前两帧的预测结果进行分类.所有之前帧的结果都是分类网络的初始手势类型,投票后的结果将被用作输出.表 6 列出了所有预测及其概率.基于该表可以将正确手势投票结果的概率相加来计算新的召回率:

$$q^* = q^3 + 3q^2(1-q) = q^2(3-2q) \tag{1}$$

这也就是说,对于一个手势(如"语音输入")的分类网络召回率为q,投票系统可以将其提升到一个新的召回率$q^* = q^2(3-2q)$.对于$0.5 < q < 1$,我们有$q^* > q$.由于我们手势分类的召回率为q=0.9817,投票系统可以将手势的召回率提高到q^* = 0.9990.即投票系统可以将错误率降低到十八分之一(从$1-q$ = 1.83%降到$1-q^*$ = 0.1%).图 12 展示了在不同手势召回率下,使用和不使用投票系统的比较结果.

表 6: 投票系统产生不同的情况和对应概率. "不同情况"栏的三项是分类网络对前两帧及当前帧的预测结果.

不同情况	投票后改为	概率
RR**R**	RR**R**	q^3
RR**W**	RR**R**	$q^2(1-q)$
RW**R**	RW**R**	$q^2(1-q)$
RW**W**	RW**W**	$q(1-q)^2$
WR**R**	WR**R**	$q^2(1-q)$
WR**W**	WR**W**	$q(1-q)^2$
WW**R**	WW**W**	$q(1-q)^2$
WW**W**	WW**W**	$(1-q)^3$

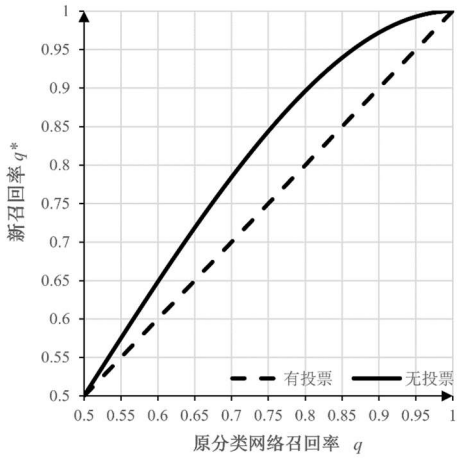


图 12: 投票系统对手势分类网络分类性能的提升.我们发现分类结果的召回率得到了提升.

5.3.3 手势解释器模块

在手势解释器模块中,手部位置、手势类型和手关节位置会一起输入有限状态机,以解释手势动作的类别.为了能轻松控制网页和其他程序,我们将每个手势动作映射到特定的鼠标/键盘事件,因此只需要对支持鼠标/键盘的 UI(如网页和应用程序) 少量修改即可部署我们的手势系统.

状态机可以确定手势动作的触发、停止和切换,从而实时解释手势.系统使用状态机的状态来存储历史信息,而不是将过去的手势序列一同输入.一旦用户做出新的手势或当前手势的保持时间超过时间阈值,状态就会改变并解释响应的动作.在图 13 中"返回"手势有两个时间阈值:第一次"返回"动作为 0.5 秒,后续的连续"返回"动作为 1 秒,后者的更长以防止误判.当"返回"手势第一次出现时,状态机将切换到 RETURN-INIT 状态,并记录

进入时间,如果用户继续保持手势,状态机将保持在 RETURN-INIT 状态.经过 0.5 秒后,状态机将执行"返回"动作,同时切换到 RETURN-MORE 状态并重置进入时间.如果用户继续保持手势 1 秒,状态机将再次执行"返回"动作并重置进入时间.如果用户做出其他手势,状态将发生变化,所有参数都将被重置.这种状态机机制确保了系统可以实时解释和执行手势.图 14 展示了状态机在用户连续保持"返回"(即 THUMB-L)手势时的系统情况.

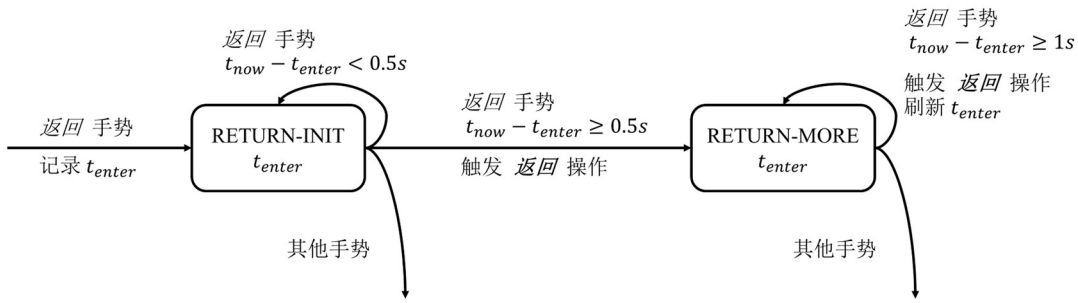


图 13: 通过状态机执行"返回"操作的过程示例.

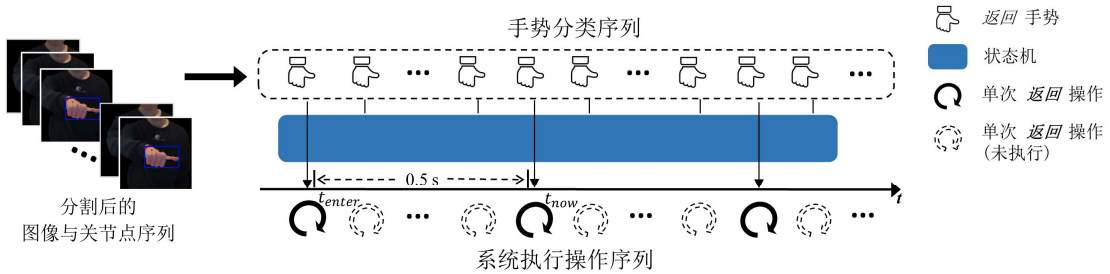


图 14: 状态机的效用实例.分类后的手势序列经过状态机被转换成系统执行的操作序列.

5.3.4 手势识别系统性能

手势解释模块的效果.我们设计了三种不同的手势操作序列,其中含有所有手势类别(也包含无意识手势),每个手势保持较短的一段随机时间.为了研究系统的稳定性,我们选择了一位有经验的用户执行这些复杂的手势序列.实验结束后,我们标注了每个操作序列的手势类别标签和持续时间.有效手势区间是从举起手执行功能手势开始到操作者结束功能手势为止;除了有效手势之外的其他手势被标记为无意识手势"NONE".

图 14 展示了系统中状态机处理"返回"动作的示例,通过状态机,"返回"手势序列每 0.5 秒就会生成一个"返回"操作,如果没有状态机,每帧都将触发一次返回操作.此外,图 15 展示了手势解释模块在手势识别中的实验验证结果.得益于手势解释模块,系统在动作的持续时间内不容易被打断,并且能够稳定地触发操作.图中颜色连续的区域是较好的识别结果.基于负采样策略,系统可以很好地识别各种无意义的无意识手势(如灰线上方的图片所示,比如触摸脸部、触摸腹部、抓挠头部).基于投票模块,系统可以有效地过滤掉手势识别结果中的单帧错误,如灰线下方的图片所示.尽管我们的系统拥有很好的识别能力,但仍会偶尔出现一些识别错误(见 7.3 节).总之,实验结果显示大多数时候系统可以很好地完成手势识别,并避免无意识手势.

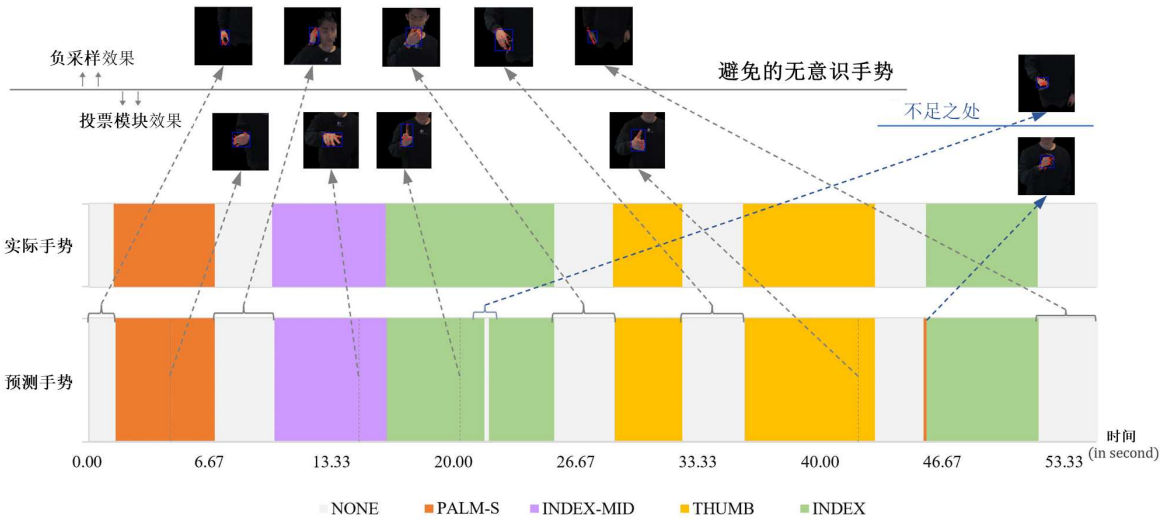


图 15: 典型手势序列的分类结果.灰色箭头指向的图片是有代表性的无意识手势,蓝色箭头指向的图片是出错的识别结果.

系统运行时间.系统使用一张 Nvidia RTX 3060Ti GPU 运行手势模型.在该环境下,我们的手势系统推理帧率为 20FPS(延迟 50ms),这足以支持大屏幕上的实时手势交互.

系统的操作距离.为了评估系统在不同操作距离下的性能,我们在 Azure Kinect 的可用交互距离内（0.5m-3.86m）选择了四个距屏幕不同距离的交互点位进行用户实验并测试系统中五种任务手势的性能.我们选择了 4.3 节中设计的交互任务,邀请了相同的 16 位被试,并使用如图 8 展示的手势完成交互任务.在不同交互距离下系统的性能如表 7 所示,被试站立于 0.5m,1.5m 和 2.5m 处操作时五项交互任务的性能没有显著差距,在 3.5m 这一最远的交互距离下由于接近 Kinect 的可用交互距离极限,五项任务的性能都有轻微下降,但被试仍可顺利完成各项任务.实验证明 DistantHands 足以支持在较远的距离交互.

表 7：手势系统在不同交互距离下的性能.

交互距离	左右移动		上下移动	
	用时(ms)	错误距离(像素)	用时(ms)	错误距离(像素)
0.5m	6194 ± 866	**20 ± 2.1**	4709 ± 149	18.5 ± 1.1
1.5m	**6131 ± 521**	20.9 ± 2.3	4827 ± 383	**18.4 ± 1.9**
2.5m	6479 ± 641	22.1 ± 2.6	**4680 ± 174**	19.8 ± 1.5
3.5m	6993 ± 912	24.4 ± 3.2	5004 ± 192	20.3 ± 1.1

交互距离	点击		语音输入		返回/取消
	用时(ms)	正确率	用时(ms)	正确率	用时(ms)
0.5m	43710 ± 3378	0.94 ± 0.07	2150 ± 121	0.96 ± 0.05	**1602 ± 60**
1.5m	**41450 ± 1763**	0.96 ± 0.06	2126 ± 66	**0.98 ± 0.04**	1650 ± 58
2.5m	41887 ± 1983	**1 ± 0**	**2113 ± 84**	0.93 ± 0.09	1659 ± 45
3.5m	51976 ± 5107	0.85 ± 0.1	2266 ± 182	0.97 ± 0.02	1619 ± 107

5.3.5 针对大屏幕的手势实现细节

为了减少空中手势可能产生的交互疲劳,DistantHands 还设计了高效的指点策略和快捷手势.

指点策略.为了减少参与者在执行"点击"手势时产生的手臂疲劳,我们设计了一个平行于显示屏的虚拟交互区域,并使用单应性变换[64]将虚拟平面区域上手势的位置映射到显示屏里的像素坐标.这个虚拟平面区域以

我们收集数据集中所有手势的平均位置为中心(见 5.3.1 节),宽度和高度约为手势可达范围的一半(远小于大屏幕的尺寸).图 16 图示了我们的映射方法.得益于这种映射方法,与光线投射方法相比,DistantHands 的操作可以与显示屏尺寸和交互距离完全解耦.我们只需关注手势在虚拟区域的位置,系统将映射到显示器上的相应位置,因此 DistantHands 所需的手势移动量少于直接触摸交互,从而导致较少的手部疲劳.

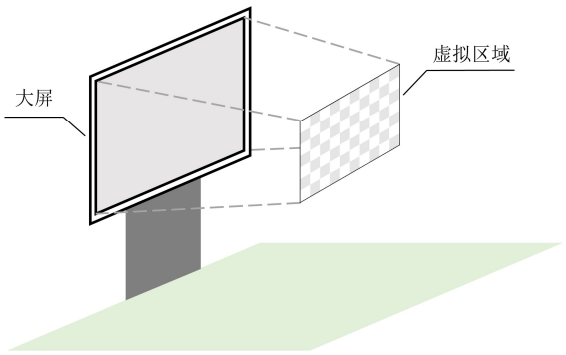


图 16: 将物理空间中指点的虚拟交互区域映射到显示器像素空间示例.

快捷手势.为了减少使用"取消/返回"和"语音输入"操作导致的手势疲劳,我们设计了全局快捷手势直接触发系统的"取消/返回"或"语音输入"的功能,而不是点击固定位置的返回按钮,因为这可能会产生手臂大幅度的移动并可能远离当前的指点位置.此外,为了执行连续多次返回的手势序列,我们设计了状态机机制,允许用户保持手势静止达成目标,这也有效减少了手势疲劳.这种快捷手势的机制如图 13 所示.**多人交互策略.**当多人同时出现在相机视野中,需要选择交互的用户.DistantHands 首先进行多人关节点检测,并选择相机视野最中心的人作为潜在用户.随后选择该人的右手作为交互用手并提取手部深度图以此完成后续流程中手部关节点的提取和手势的分类.

6 大屏幕交互方式的比较:触摸与空中手势

现有许多手势交互与触摸交互的对比研究,普遍的结论是手势交互的用户体验不如触摸交互.因此本实验的目的是比较 DistantHands 和触摸屏交互在大屏幕上的信息浏览中的用户体验.我们主要关注用户在使用过程中的,如易用性,易学性,使用效率等的参与体验.

6.1 研究原型系统设计和实现

本实验设计了一个名为 OlympicHall2022 的信息浏览应用程序并用其评估我们的手势交互系统[65],这是一个典型的,包含了大屏幕常见的交互任务的交互原型,同时也是 2022 年北京冬奥会的在线展览馆.我们希望评估手势交互在公共大屏幕信息浏览中的用户体验,因此有成熟的系统知识并适合所有人阅读学习的冬奥会相关信息就非常适合使用.该系统的操作与本文上述交互操作的定义相匹配,能够很好地代表公共环境中大屏幕交互式信息浏览的各种用途.我们收集大量多媒体素材来组织 OlympicHall2022 的内容,系统内集成了四大主题展览(即体育、场馆、人物和城市),每个主题展览都展示了相关的简介、历史、文化、运动员、回忆、亮点等.图 17 中展示了 OlympicHall2022 的用户界面示例.

（a）主要浏览界面　　（b）项目选择长廊

图 17: OlympicHall2022 的操作页面实例. (a)展示了用户在 OlympicHall2022 中主要的浏览界面,主要由展览展示区(1)、主题展览选择按钮(2)和展览项目选择栏(2A)、返回按钮(3)、设置按钮(4)和虚拟人问答[65]按钮(5-6)组成. (b)展示了 OlympicHall2022 的导航页面,其中包含返回按钮(3)和每个子页面的条目(7).

6.2 实验简介

被试和设备.16 名被试(11 男 5 女,平均年龄 25.8 岁,标准差 4.32)被随机分为两组以平衡测试顺序带来的影响.实验前,被试详细了解在线展厅的用法和用户实验过程.对于手势交互系统,硬件设置与第 5.2 节相同.用于手势交互系统的大屏幕也是触摸屏,因此天然的支持多点触控交互.

设计和过程.我们在 OlympicHall2022 上比较 DistantHands 和触摸屏.两种交互系统使用完全相同的在线展厅布局和展示内容.每个被试需要先后使用两种不同的交互方式体验 OlympicHall2022,被试之间手势交互和触摸屏交互的顺序是平衡的,一半被试先体验手势交互后体验触摸屏交互,另一半则相反.

在用户实验开始之前,被试首先学习如何通过手势或触摸与 OlympicHall2022 交互.然后,被试将通过其中一种交互方式随意浏览在线展厅 3 分钟,并在完成后填写用户体验问卷.休息 2 分钟后,被试将完成另一种交互方式下的体验测试,测试过程与之前完全相同.

测量指标.我们沿用第 4.3 节主观评估中的所有指标,并添加了几个指标来衡量在线展厅的用户体验,包括:有效性、易用性、易学性、效率、舒适性、可访问性、连贯性、满意度和专注度.我们使用 7 分连续量表进行测量,1 分为最低分,7 分为最高分.表 7 展示了 DistantHands 和多点触控系统的对比试验结果.

6.3 结果

由于数据分布不符合正态分布的假设,我们使用 Wilcoxon 符号秩检验(双尾检验)作为非参数检验来分析成对样本.对于用户体验指标,分析结果表明可访问性指标达到显著性($T = 75, Z = -2.12, p = .03$);连贯性和满意度的比较则边缘显著($p$ <.08).如表 8 所示,被试的数据在大多数指标上都表明,除了使用效率之外,使用手势浏览大屏幕在线展厅的体验优于使用触摸屏.

表 8: 我们的手势交互系统与触屏交互之间的用户实验对比结果.(均值±SEM)

	有效性	易用性	易学性	效率	舒适性	可用性	连贯性	满意度	专注度
手势	**5.8 ± .4**	**5.4 ± .3**	**5.8 ± .3**	3.8 ± .4	**3.9 ± .4**	**4.8 ± .5**	**4.8 ± .5**	**4.6 ± .3**	**5.1 ± .4**
触摸	5.5 ± .5	**5.4 ± .5**	**5.8 ± .4**	**4.2 ± .5**	3.4 ± .3	3.9 ± .5	4.0 ± .5	3.8 ± .5	4.9 ± .4
p	.164	.394	.416	.300	.309	.039	.048	.075	.743

7 讨论

7.1 系统优势

DistantHands 可以给用户带来良好体验,相比与触屏交互有更好的主观评分,有效增强了用户体验.下面我

们将分析该系统的优势.

更自然.首先,DistantHands 使用了 Azure Kinect,能够捕捉较广距离范围内的高质量手部图像,同时获取准确的手部姿态,这允许用户能够在与大屏幕进行交互时可以前后走动,站的位置更自由.用户不必站在某处才能交互,困惑于在不合适的观看距离上寻找交互对象,大大减轻了用户在交互过程中产生的疲劳和心理压力.其次,我们采用了一种新颖的三阶段手势诱导方法,通过这种方法我们获得了最适合 DistantHands 的自然手势集合.

更准确. DistantHands 采用了先进的人手分割、手势分类和识别方法;对于不可避免的无意识手势问题,我们设计了系统性的策略,并实现了出色的手势识别精度.

7.2 关键发现

"取消/返回"任务中用户更偏好手指向左的手势.对于"取消/返回"任务,很多被试提出的手势都与左侧相结合(如拇指向左).我们认为这是因为所有被试的阅读和书写习惯都是从左到右[66].在后续实验中,相同手势的不同朝向确实产生了不同的用户体验,这也支持本结论.

手势系统的用户体验优于触摸系统.触摸系统通常需要非常近的交互距离.使用手势系统,用户可以从更合适的距离浏览.我们发现一些被试在使用触摸系统导航到所需页面时,仍会首先退后浏览以获得更好的视野.对于手势系统而言,执行"取消/返回"操作并不需要找到"取消/返回"按钮; 但对于触摸系统来说,找到并触发"取消/返回"按钮往往会引起疲劳和困惑,因为不同应用程序或页面上"取消/返回"按钮的位置可能不同.因此,DistantHands 使用快捷手势统一"取消/返回"功能能够减轻用户疲劳,同时对于基于手势的应用程序,UI 设计师甚至可以删除"取消/返回"按钮,以获得更美观的页面.

没有物理反馈的点击可能会令用户感到困惑.当比较"点击"功能的平均消耗时间时,"基于距离的点击"的效果更好("基于距离的点击"平均 10 次点击耗时 34.77 秒,而"基于时间的点击"为 41.89 秒)."基于距离的点击"的成功率为 90%,而"基于时间的点击"为 100%.这两个客观指标表明这两种点击方式的性能是相当的.然而,当评估用户的主观得分时,"基于时间的点击优于"基于距离的点击".在本文的实验中,我们发现没有物理反馈的情况下,用户更倾向于将手移动更远的距离以防止点击失败,这导致更剧烈、更大幅度、更疲劳的手势.一种可能的解决方案是在点击过程中添加明显的视觉反馈,这样用户就可以明确他们的点击过程,并使用更小幅度的动作触发大屏幕上的点击功能.

7.3 系统局限性

硬件自身缺陷. Azure Kinect 可以采集精确的视觉数据并且耐用且成本较低,但它仍然存在一些局限性,如在户外场景下强光会严重影响深度相机的采集准确度.

缺乏自适应滚动速度.传统的鼠标和触摸系统有"抓取和拖动"的操作逻辑,允许用户根据需要快速拖动(滚动)页面,然而在我们的手势系统中没有包含类似的操作.如"左右移动"操作在映射到键盘后,系统以固定的时间间隔触发滚动,这让我们的系统只能以固定的速度滚动页面.但如果滚动速度太快可能会导致较差的用户体验,如果速度太慢,用户需要更长时间才能达成目的.一种可能的实现自适应滚动速度的方法是将手势保持时间映射到滚动速度,这需要额外的界面设计,但可能会增加用户的学习难度并降低了手势系统的普适性.

难以避免的无意识手势.尽管 DistantHands 拥有很好的识别性能,但仍有几个需要改进的地方.如图 15 所示,首先,当使用食指手势(INDEX)快速执行点击操作时会产生动态模糊,此时系统很难识别出该手势.因此,连续指点序列被判断为无意识手势,并中断了 11 帧.其次,在 46 秒的位置,即食指手势开始之前,因为食指手势还没有完全形成,系统会短暂地将该手势序列识别为掌心(PALM).尽管可以通过更多无意识手势负样本训练来改善这个问题,但由于无意识手势是无穷无尽的,我们很难完全解决这个问题.

用户疲劳.即使我们已经选择了理想的手势集合,用户在长时间交互后仍会感到手臂疲劳.我们注意到,如果不把手臂举到眼前操作,而是让手臂自然下垂可能是解决手臂疲劳[36]的一个较好的方案.此外,由于在实验中参与的用户数量较少,可能无法准确代表来自不同背景的用户的交互习惯[67].因此,在不同文化背景的不同区域设计不同的手势操作可能会缓解这个问题.

8 结论

我们提出了一种新的空中手势交互系统 DistantHands,实现了较广距离范围内与大屏幕的交互.本文引入了一个新的手势诱导阶段,根据用户的实机体验和评估选择合适的空中手势集.此外,本文提出了一种空中手势在线识别方法,该方法避免了大量的无意识手势,从而提高了用户体验.最后,本文比较了 DistantHands 和触摸交互在大屏幕交互中的用户体验,发现我们的新手势交互系统在舒适性、连贯性、满意度、专注度等指标上都优于触摸界面.我们的工作也可以为大屏之外的空中手势交互系统和应用(如 AR/VR 交互和车载座舱人机交互)提供借鉴意义.

References:

[1] Ardito C, Buono P, Costabile MF, Desolda G. Interaction with large displays: a survey. ACM Computing Surveys (CSUR), ACM New York, NY, USA, 2015, 47(3): 1–38.

[2] Koutsabasis P, Vogiatzidakis P. Empirical research in mid-air interaction: a systematic review. International Journal of Human-Computer Interaction, Taylor & Francis, 2019, 35(18): 1–22.

[3] Waugh K, Robertson J. Don't touch me! a comparison of usability on touch and non-touch inputs. Human-Computer Interaction – INTERACT 2021, Springer International Publishing, 2021: 400–404.

[4] Vuletic T, Duffy A, Hay L, McTeague C, Campbell G, Grealy M. Systematic literature review of hand gestures used in human computer interaction interfaces. International Journal of Human-Computer Studies, Elsevier, 2019, 129: 74–94.

[5] Huang S, Ranganathan SPB, Parsons I. To touch or not to touch? comparing touch, mid-air gesture, mid-air haptics for public display in post covid-19 society. SIGGRAPH Asia 2020 Posters. New York, NY, USA: Association for Computing Machinery, 2020.

[6] Malik ZH, Arfan M. Evaluation of accuracy: a comparative study between touch screen and midair gesture input. Arai K, Bhatia R. Advances in Information and Communication. Cham: Springer International Publishing, 2020: 448–462.

[7] Kim D, Kim H, Kim H-K, Park S-R, Lee K-S, Kim K-H. ThunderPunch: a bare-hand, gesture-based, large interactive display interface with upper-body-part detection in a top view. IEEE Computer Graphics and Applications, IEEE, 2018, 38(5): 100–111.

[8] Gentile V, Khamis M, Milazzo F, Sorce S, Malizia A, Alt F. Predicting mid-air gestural interaction with public displays based on audience behaviour. International Journal of Human-Computer Studies, Elsevier, 2020, 144: 102497.

[9] Pavlovic VI, Sharma R, Huang TS. Visual interpretation of hand gestures for human-computer interaction: a review. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 677–695.

[10] Ultraleap. How hand tracking works. 2020.

[11] Vuletic T, Duffy A, McTeague C, Hay L, Brisco R, Campbell G, Grealy M. A novel user-based gesture vocabulary for conceptual design. International Journal of Human-Computer Studies, Elsevier, 2021, 150: 102609.

[12] Sluÿters A, Sellier Q, Vanderdonckt J, Parthiban V, Maes P. Consistent, continuous, and customizable mid-air gesture interaction for browsing multimedia objects on large displays. International Journal of Human–Computer Interaction, Taylor & Francis, 2023, 39(12): 2492–2523.

[13] Ackad C, Clayphan A, Tomitsch M, Kay J. An in-the-wild study of learning mid-air gestures to browse hierarchical information at a large interactive public display. Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015: 1227–1238.

[14] Wobbrock JO, Morris MR, Wilson AD. User-defined gestures for surface computing. Proceedings of the SIGCHI conference on human factors in computing systems. 2009: 1083–1092.

[15] Seyed T, Burns C, Costa Sousa M, Maurer F, Tang A. Eliciting usable gestures for multi-display environments. Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces. 2012: 41–50.

[16] Connell S, Kuo P-Y, Liu L, Piper AM. A wizard-of-oz elicitation study examining child-defined gestures with a whole-body interface. Proceedings of the 12th International Conference on Interaction Design and Children. 2013: 277–280.

[17] Piumsomboon T, Clark A, Billinghurst M, Cockburn A. User-defined gestures for augmented reality. CHI'13 Extended Abstracts on Human Factors in Computing Systems. 2013: 955–960.

[18] Dong H, Danesh A, Figueroa N, El Saddik A. An elicitation study on gesture preferences and memorability toward a practical hand-gesture vocabulary for smart televisions. IEEE access, IEEE, 2015, 3: 543–555.

[19] Dong H, Figueroa N, Saddik AE. An elicitation study on gesture attitudes and preferences towards an interactive

hand-gesture vocabulary. Proceedings of the 23rd ACM international conference on Multimedia. 2015: 999–1002.
[20] Cheng H, Yang L, Liu Z. Survey on 3d hand gesture recognition. IEEE transactions on circuits and systems for video technology, IEEE, 2015, 26(9): 1659–1673.
[21] Ng CW, Ranganath S. Real-time gesture recognition system and application. Image and Vision Computing, 2002, 20(13): 993–1007.
[22] Molchanov P, Yang X, Gupta S, Kim K, Tyree S, Kautz J. Online detection and classification of dynamic hand gestures with recurrent 3d convolutional neural networks. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016: 4207–4215.
[23] Min Y, Zhang Y, Chai X, Chen X. An efficient pointlstm for point clouds based gesture recognition. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020: 5761–5770.
[24] Hou J, Wang G, Chen X, Xue J-H, Zhu R, Yang H. Spatial-temporal attention res-tcn for skeleton-based dynamic hand gesture recognition. Proceedings of the European Conference on Computer Vision (ECCV) Workshops. 2018: 0–0.
[25] Yan S, Xiong Y, Lin D. Spatial temporal graph convolutional networks for skeleton-based action recognition. Proceedings of the AAAI conference on artificial intelligence. 2018, 32.
[26] Shi L, Zhang Y, Cheng J, Lu H. Two-stream adaptive graph convolutional networks for skeleton-based action recognition. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 12026–12035.
[27] Zhang W, Lin Z, Cheng J, Ma C, Deng X, Wang H. Sta-gcn: two-stream graph convolutional network with spatial–temporal attention for hand gesture recognition. The Visual Computer, Springer, 2020, 36: 2433–2444.
[28] Kwon T, Tekin B, Stühmer J, Bogo F, Pollefeys M. H2o: two hands manipulating objects for first person interaction recognition. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 10138–10148.
[29] Tofighi G. Engagement detection framework for hand gesture and posture recognition. Ryerson University, 2017.
[30] Chen W-H, Lin Y-H, Yang S-J. A generic framework for the design of visual-based gesture control interface. 2010 5th IEEE Conference on Industrial Electronics and Applications. IEEE, 2010: 1522–1525.
[31] Bhuyan MK, Bora PK, Ghosh D. An integrated approach to the recognition of a wide class of continuous hand gestures. International Journal of Pattern Recognition and Artificial Intelligence, World Scientific, 2011, 25(02): 227–252.
[32] Tsai T-H, Lin C-Y. Visual hand gesture segmentation using signer model for real-time human-computer interaction application. 2007 IEEE Workshop on Signal Processing Systems. IEEE, 2007: 567–572.
[33] Graetzel C, Fong T, Grange S, Baur C. A non-contact mouse for surgeon-computer interaction. Technology and Health Care, IOS Press, 2004, 12(3): 245–257.
[34] Vogel D, Balakrishnan R. Distant freehand pointing and clicking on very large, high resolution displays. Proceedings of the 18th annual ACM symposium on User interface software and technology. 2005: 33–42.
[35] Nancel M, Wagner J, Pietriga E, Chapuis O, Mackay W. Mid-air pan-and-zoom on wall-sized displays. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2011: 177–186.
[36] Ruiz J, Vogel D. Soft-constraints to reduce legacy and performance bias to elicit whole-body gestures with low arm fatigue. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 2015: 3347–3350.
[37] Bolt RA. “Put-that-there” voice and gesture at the graphics interface. Proceedings of the 7th annual conference on Computer graphics and interactive techniques. 1980: 262–270.
[38] Siddhpuria S, Katsuragawa K, Wallace JR, Lank E. Exploring at-your-side gestural interaction for ubiquitous environments. Proceedings of the 2017 Conference on Designing Interactive Systems. 2017: 1111–1122.

[39] Müller J, Bailly G, Bossuyt T, Hillgren N. MirrorTouch: combining touch and mid-air gestures for public displays. Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices and Services. New York, NY, USA: Association for Computing Machinery, 2014: 319–328.
[40] Vatavu R-D, Zaiti I-A. Leap gestures for tv: insights from an elicitation study. Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video. 2014: 131–138.
[41] Pham T, Vermeulen J, Tang A, MacDonald Vermeulen L. Scale impacts elicited gestures for manipulating holograms: implications for ar gesture design. Proceedings of the 2018 Designing Interactive Systems Conference. 2018: 227–240.
[42] Ruiz J, Li Y, Lank E. User-defined motion gestures for mobile interaction. Proceedings of the SIGCHI conference on human factors in computing systems. 2011: 197–206.
[43] Wang RY, Popović J. Real-time hand-tracking with a color glove. ACM transactions on graphics (TOG), ACM New York, NY, USA, 2009, 28(3): 1–8.
[44] Fan J, Fan X, Tian F, Li Y, Liu Z, Sun W, Wang H. What is that in your hand? recognizing grasped objects via forearm electromyography sensing. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, ACM New York, NY, USA, 2018, 2(4): 1–24.
[45] Vogiatzidakis P, Koutsabasis P. Gesture elicitation studies for mid-air interaction: a review. Multimodal Technologies and Interaction, MDPI, 2018, 2(4): 65.
[46] Microsoft. Kinect for windows. 2023.
[47] Microsoft. Azure kinect dk hardware specifications. 2023.
[48] Microsoft. Azure kinect dk: build for mixed reality using ai sensors. 2023.
[49] Xiong F, Zhang B, Xiao Y, Cao Z, Yu T, Zhou Tianyi J, Yuan J. A2J: anchor-to-joint regression network for 3d articulated pose estimation from a single depth image. Proceedings of the IEEE Conference on International Conference on Computer Vision (ICCV). 2019: 793–802.
[50] Armagan A, Garcia-Hernando G, Baek S, Hampali S, Rad M, Zhang Z, Xie S, Chen M, Zhang B, Xiong F, Xiao Y, Cao Z, Yuan J, Ren P, Huang W, Sun H, Hrúz M, Kanis J, Krňoul Z, Wan Q, Li S, Yang L, Lee D, Yao A, Zhou W, Mei S, Liu Y, Spurr A, Iqbal U, Molchanov P, Weinzaepfel P, Brégier R, Rogez G, Lepetit V, Kim T-K. Measuring generalisation to unseen viewpoints, articulations, shapes and objects for 3d hand pose estimation under hand-object interaction. Proceedings of European Conference on Computer Vision. Berlin, Heidelberg: Springer-Verlag, 2020: 85–101.
[51] Tompson J, Stein M, Lecun Y, Perlin K. Real-time continuous pose recovery of human hands using convolutional networks. ACM Transactions on Graphics (ToG), ACM New York, NY, USA, 2014, 33(5): 1–10.
[52] Wigdor D, Wixon D. Brave nui world: designing natural user interfaces for touch and gesture. Elsevier, 2011.
[53] Wachs JP, Kölsch M, Stern H, Edan Y. Vision-based hand-gesture applications. Communications of the ACM, ACM New York, NY, USA, 2011, 54(2): 60–71.
[54] Yeasin M, Chaudhuri S. Visual understanding of dynamic hand gestures. Pattern Recognition, 2000, 33(11): 1805–1817.
[55] Wu Y, Huang TS. Vision-based gesture recognition: a review. Gesture-Based Communication in Human-Computer Interaction: International GestureWorkshop, GW'99 Gif-sur-Yvette, France, March 17-19, 1999 Proceedings, Springer, 1999: 103–115.
[56] Gu Y, Yu C, Chen X, LI Z, Shi Y. TypeBoard: identifying unintentional touch on pressure-sensitive touchscreen keyboards. The 34th Annual ACM Symposium on User Interface Software and Technology. Association for Computing Machinery, 2021: 568–581.
[57] LaViola Jr JJ. An introduction to 3d gestural interfaces. ACM SIGGRAPH 2014 Courses. 2014: 1–42.
[58] Cooperrider K. Fifteen ways of looking at a pointing gesture. PsyArXiv, 2020.

[59] Surale HB, Matulic F, Vogel D. Experimental analysis of barehand mid-air mode-switching techniques in virtual reality. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. New York, NY, USA: Association for Computing Machinery, 2019: 1–14.
[60] Wobbrock JO, Findlater L, Gergle D, Higgins JJ. The aligned rank transform for nonparametric factorial analyses using only anova procedures. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY, USA: Association for Computing Machinery, 2011: 143–146.
[61] Jocher G, Stoken A, Chaurasia A, Borovec J, Kwon Y, Michael K, Liu C, Fang J, Abhiram V, Skalski S, Others. Ultralytics/yolov5: v6. 0—yolov5n 'nano'models, roboflow integration, tensorflow export, opencv dnn support. Zenodo Tech. Rep., 2021.
[62] Oberweger M, Lepetit V. Deepprior++: improving fast and accurate 3d hand pose estimation. Proceedings of the IEEE international conference on computer vision Workshops. 2017: 585–594.
[63] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016: 770–778.
[64] Hartley RI, Zisserman A. Multiple view geometry in computer vision. 第 2 版. Cambridge University Press, 2004.
[66] Morikawa K, McBeath MK. Lateral motion bias associated with reading direction. Vision Research, Elsevier, 1992, 32(6): 1137–1141.
[67] Archer D. Unspoken diversity: cultural differences in gestures. Qualitative sociology, Springer Nature BV, 1997, 20(1): 79.

附中文参考文献:
[65] 刘舫, 吕天, 刘心阁, 叶盛, 郭锐, 张烈, 马翠霞, 王庆伟, 刘永进. 引入智能虚拟讲解员的云展厅构建与交互反馈研究. 软件学报, , 2024(35(3)): 1534–1551.

张永浩(2000－), 男, 硕士生, 主要研究领域为计算机视觉, 人机交互.

万炎广(1999－), 男, 硕士, 主要研究领域为计算机视觉, 人机交互.

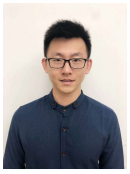
刘舫(1993－), 女, 博士, 主要研究领域为计算机视觉, 深度学习, 草图交互, 情绪计算.

程坚(1996－), 男, 硕士, 主要研究领域为计算机视觉, 人机交互.

刘心阁(1991－), 女, 博士, 助理研究员, 主要研究领域为面孔加工, 情感计算, 人机交互.

张维(1982－), 男, 博士, 工程师, 主要研究领域为人机交互, 手势识别.

马翠霞(1975－), 女, 博士, 研究员, CCF高级会员, 主要研究领域为人机交互, 媒体大数据可视分析.



卞玉龙(1988－), 男, 博士, 副教授, 硕士生导师, ACM SIGCHI China Chapter 执委, 主要研究领域为人机交互, 虚拟现实, 严肃游戏.

周晓(1986－), 男, 博士, 副研究员, 硕士生导师, 主要研究领域为空间信息处理系统体系架构, 多源传感器数据智能处理与应用, 空间信息系统仿真技术.

邓小明(1980－), 男, 博士, 研究员, CCF高级会员, 主要研究领域为计算机视觉, 人机交互.

刘永进(1977－), 男, 博士, 教授, 博士生导师, CCF 杰出会员, 主要研究领域为计算几何, 计算机图形学, 计算机视觉, 认知计算, 模式识别.

王宏安(1963－), 男, 博士, 研究员, CCF高级会员, 主要研究领域为自然人机交互, 实时智能计算.