

基于可见光逆反射通信的移动增强现实定向交互

李景聿¹, 陈拓潮², 李明哲³, 徐翔海², 张 斌⁴, 徐子晨⁵,
刘 璠¹, 黄 罡¹, 冯 云⁶, 许辰人^{1*}

(1. 北京大学计算机学院, 北京 100871; 2. 华盛顿大学计算机科学与技术学院, 美国西雅图 98195;
3. 卡内基梅隆大学人机交互研究所, 美国匹兹堡 15213; 4. 康奈尔大学信息科学学院, 美国伊萨卡 14853;
5. 南昌大学数学与计算机学院, 江西南昌 330031; 6. 北京大学第三医院, 北京 100191)

摘 要: 移动增强现实应用中, 用户频繁地与环境中不同的智能物体交互完成任务, 其完成效率和用户体验由所采用的定向交互技术决定。然而, 从交互手段上来说, 现阶段定向交互依赖于 Wi-Fi、BLE 等无线技术, 其信号全向传播的特点使其不能利用用户的空间环境 (即位置和方向) 来缩短互动时间, 带来了不必要的时间成本; 从交互界面上来说, 现阶段基于视觉的界面存在可靠性低、扩展性低等问题, 进一步限制了系统的通用性。本文介绍了 RetroAR——基于可见光逆反射通信的定向交互系统。RetroAR 利用了光的定向传播特性来保留用户的空间环境, 并依靠逆反射通信实现用户与目标设备之间无连接的快速定向交互。系统实验表明, RetroAR 最远可在 4 m 距离支持 100° 视角的交互, 同时实现厘米级的六自由度 (6-Degrees Of Freedom, 6-DoF) 三维跟踪。用户研究表明, 与基于 Wi-Fi 的解决方案相比, RetroAR 将非接触式控制的交互时间减少了 2 倍, 并且具有更好的用户体验。RetroAR 借助可见光逆反射通信来利用用户空间环境, 保持交互过程中的直观性。用户可以“所指即所控”的方式与多个目标进行互动, 实现类似自然交互的快速定向交互。

关键词: 移动增强现实; 定向交互; 所指即所控; 可见光逆反射通信; 聚合物分散液晶

基金项目: 国家自然科学基金 (No.62022005, No.62061146001)

中图分类号: TP399; TN929 **文献标识码:** A **文章编号:** 0372-2112(2024)11-3643-14

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20220677

Directional Interaction for Mobile Augmented Reality Based on Visible Light Backscatter Communication

LI Jing-yu¹, CHEN Tuo-chao², LI Ming-zhe³, XU Xu-hai², ZHANG Cheng⁴, XU Zi-chen⁵, LIU Xuan-zhe¹,
HUANG Gang¹, FENG Yun⁶, XU Chen-ren^{1*}

(1. School of Computer Science, Peking University, Beijing 100871, China;

2. Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA 98195, USA;

3. Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA;

4. Department of Information Science, Cornell University, Ithaca, NY 14853, USA;

5. School of Mathematics and Computer Science, Nanchang University, Nanchang, Jiangxi 330031, China;

6. Peking University Third Hospital, Beijing 100871, China)

Abstract: In mobile augmented reality applications, users interact with smart objects in proximity to finish collaboration or interaction tasks, whose efficiency and user experience are determined by the underlying directional interaction technology. However, current directional interaction technologies at this stage are inefficient. For interaction means, they rely on wireless technologies such as Wi-Fi and BLE, which propagates omni-directionally and thus cannot use the user's spatial context (i.e., location and direction) to shorten the interaction time, bringing unnecessary effort. In terms of interaction interface, current vision-based interfaces suffer from low reliability and low scalability, which further limits the adaptability and efficiency of the system. To address this issue, we developed RetroAR: an optical-sensing solution that leverages visible light backscatter communication to serve for directional interaction with intelligent objects on commodity smartphones.

RetroAR exploits the directional propagation property of light to preserve the user's spatial context, which enables fast connection-free directional interaction between the user and the target devices. RetroAR instruments objects with custom retro-reflective markers called ViTags. When users interact with these smart objects, these ViTags are used to communicate with the camera on the mobile reader by backscattering the flashlight beams. We first conducted a system evaluation which showed that RetroAR could work reliably at a distance up to 4 meters and a view angle up to 100 degrees, and is able to achieve 6-DoF 3D tracking with an error as low as 1 cm on translation and 4.7 degrees on rotation. To evaluate how our system performs in terms of interaction, we then conducted a user study with 12 participants, which demonstrated that RetroAR improves the interaction time of MAR contactless control by at least two times compared to Wi-Fi-based solutions. RetroAR utilizes the user's spatial environment with visible light backscatter communication to maintain the intuitiveness of the interaction process. Users can interact with multiple targets in a point-and-control manner, which reduces interaction costs and brings a natural and intuitive interaction experience.

Key words: mobile augmented reality; directional interaction; point-and-control; visible light backscatter communication; polymer dispersed liquid crystal

Foundation Item(s): National Natural Science Foundation of China (No.62022005, No.62061146001)

1 引言

移动增强现实 (Mobile Augmented Reality, MAR)^[1,2]使用移动设备连接物理和虚拟世界. 移动设备上的摄像头捕捉环境中的目标对象,并将捕捉到的目标对象的增强信息叠加在屏幕上. 在过去的几年里,智能手机已经成为最受欢迎的MAR平台. 因为现代智能手机通常带有丰富的传感器和强大的处理器,能够支持MAR应用,如AR游戏和数字教学^[3]和智能家居^[4]. 在这些MAR应用中,用户经常需要频繁地移动并与环境中的智能物体进行偶发的互动. 这种互动自然伴随着方向性的问题,即如何在视野中诸多候选对象中准确而快速地确定目标对象. 现阶段,这种方向性交互的实现主要是通过使用基于计算机视觉的匹配算法来推测用户的空间环境(即位置和方向)来缩小选择范围,并结合无线射频连接来提供用户身份认证,或者提供基于云端服务的数据拉取^[5-7]. 然而,这种组合式解决方案无法避免地具有建立无线连接以及维护云数据库的额外时间开销,因此降低了用户的交互体验,且增加维护者方面的工作负荷. 例如,机场在新冠肺炎期间为电梯和自助机器提供MAR非接触式控制,运营人员需要在维护Wi-Fi连接和网络服务方面付出巨大的努力;用户则必须扫描多个二维码,并等待数秒钟才能够完成连接.

为了解决这个问题,本文提出RetroAR,如图1所示,其利用了可见光逆反射通信内在的方向性,在商用智能手机上提供了一种“所指即所控”的互动体验. 用户通过简单地将智能手机指向感兴趣的物体即可快速与之互动,而无需云服务或无线连接. 移除无线连接使用户从扫描或选择连接的额外工作量中解放出来,缩短交互时间的同时使用户更加专注于交互任务本身. 该系统还简化了部署与维护步骤,不需要考虑无线冲突或复杂的网络拓扑结构,所有信息交换无须连接即可完成. RetroAR背后的原理符合人类本身的交互

直觉——用户的目标恰好是他视线指向的对象. 本文借助可见光来保留用户与目标对象之间互动的内在方向性. 具体来说,RetroAR将可见光逆反射通信(Visible Light Backscatter Communication, VLBC)^[8,9]这一通信方案改进并迁移到人机交互领域,使用户能够利用它与视线(Line-Of-Sight, LOS)范围内的对象进行双向的定向通信. RetroAR使用一种名为光标签的逆反射标签来改造智能物体,标签可以与随处可见的LCD屏幕整合来最大化部署性. 该系统利用智能手机上的后置摄像头和相应的闪光灯,通过VLBC与物体上的光标签进行双向通信. 从智能手机发送至光标签的信息被编码进闪光灯光束的变化状态中,而响应数据则通过光标签上所搭载的LCD像素编码进反射光的状态中. 用户的位置和方向信息自然地包含在闪光灯光束的方向中,而数据在VLBC信道中实现双向交换. 因此,RetroAR提供了不需要网络连接的快速定向互动,从而降低了用户的工作量和开销,将MAR互动的应用场景拓宽到更广泛的带有商用LCD屏幕的一切智能物体.

为了可见光逆反射通信这样纯粹的通信机制迁移到人机交互场景来为定向交互服务,该系统面临3个挑战.(1)如何定制和调整硬件设计,使该系统与市面常见的商用移动设备兼容,并尽可能多地支持MAR应用?(2)如何确保快速选择所带来的开销的优势不会被嘈杂的光学环境(如灯光与反射)所抵消?(3)考虑到移动设备的功率有限和用户手持时的不可预测的运动,如何确保系统能够稳健地进行6-DoF跟踪,以带来流畅的用户体验? 本文后续将阐明,该系统针对上述挑战提出了以下解决方案:(1)提出了一种新型动态光标签设计,其可为设备提供追踪锚点;(2)提出一种适用于上述标签的动态时空混合编码方案,可帮助移动设备在嘈杂环境下实现高效率的抗噪声标签跟踪;(3)提出了两个算法改进——消息级动态兴趣区、快速凸包络检



图1 RetroAR系统

测,使移动设备可以在高效解算标签信息的同时能够稳健地估算实时位姿。

本文对不同条件下的系统性能(视距、视角、闪光灯频率、照明条件)进行系统实验,并召集了12名参与者对MAR非接触式控制场景下本系统的交互效率进行用户研究。结果表明,在不同的室内环境照明条件下,RetroAR可以在最远4 m、最大100°视场角的条件下提供稳健的快速定向交互,相应的6-DoF位姿估计拥有低于1 cm的平移误差,和低至4.7°的旋转误差。用户研究结果表明,在MAR非接触式控制场景中,与基于Wi-Fi的方法相比,RetroAR提供了更好的交互体验,将交互时间提高了至少2倍。本文进一步讨论了RetroAR在现实世界的应用,扩大RetroAR与泛在物体的合作,以及进一步提高通信数据率和范围的可能性。综上,本文作出了以下4点贡献:

其一,提出并实现了RetroAR系统原型,首次将传统的可见光逆反射通信改进并迁移至交互领域,在商用移动设备上利用用户天然的空间环境来提供定向交互,为MAR应用带来一种新的“所指即所控”的快速定向交互机制。

其二,设计并实现了一种名为ViTag的新型多形态动态光标签,其既包含了空间标签的优势(即能够进行6-DoF位姿估计),也包含了单LED标签的优势(即不需要无线连接,并能够利用用户空间环境)。

其三,提出了两个新的算法设计改进,包括消息级动态兴趣区、快速凸包检测,可在不影响系统效率和交互流畅性的情况下,在嘈杂的背景中稳健地跟踪光标签。同时,提出了一套新的时空域混合编码方法,进一步提升了光标签在复杂噪声背景下的工作稳定性。

其四,进行了系统评估和用户研究,将RetroAR与基于无线连接的方法进行了比较。结果显示,对于常见的MAR用户设备互动的场景,互动时间减少了2倍,用户体验也获得了统计学显著的改善。

2 国内外研究现状

2.1 定向交互技术

在MAR应用中,用户需要与附近的特定对象进行

交互来完成特定的任务^[10]。这种交互实质上是利用了用户的空间环境,以用户视线(Line Of Sight, LOS)范围内的物体为感兴趣的对象。因此,如何快速确定视线范围内的候选对象对于流畅的交互体验至关重要。直觉上来讲,人们可以通过外观来过滤出目标对象。例如,学者利用目标设备的外形特征来帮助快速定位和识别^[11]。然而,该系统无法分辨出相同外观的电器。该方法的改进版旨在通过识别目标设备的特征纹理来帮助区分同外观电器^[12],然而当目标设备没有纹理时跟踪性能下降。近年来,人们也开始探索利用无线通信技术在用户和物体之间交换数据来还原空间环境。Bluewave^[7]将BLE与URL整合在一起,在不建立连接的情况下实现了定向信息共享,但没有直接向LOS区域内的设备提供交互界面。文献[13]利用超宽频(Ultra-Wide Band, UWB)技术和同步定位和建图技术(Simultaneous Localization And Mapping, SLAM)来与智能物体互动,但它要求用户事先完成SLAM地图训练。Snap-to-it^[6]通过让用户通过拍摄设备的照片,并在云端识别的方式来确定目标设备。尽管这非常接近自然互动,但仍然需要用背景特征进行预训练,来彻底解决用户的使用负担。RetroAR与之前的工作不同,其利用了可见光的方向性来缩小选择范围。目标设备位于用户的LOS区域内(即用户视线指向的区域),而闪光灯的光束径直地传向目标设备。此外,区别于传统的视觉匹配或者云端检索,设备识别码等数据通过逆反射的光数据链路定向返回(即信息仅来自目标物体)。这一特点有助于快速筛选出感兴趣的设备,以较小的开销实现基于用户环境感知的互动。

2.2 基于视觉的增强现实物体跟踪

对目标实现稳定的跟踪是MAR系统能够正常运作的前提条件。MAR系统将现实世界的锚点(如实体物体、人工视觉标签)与独特的虚拟数据(如三维增强模型、浮动用户界面)绑定在一起^[14,15]。目前,主流的方法是在视频流中用基于计算机视觉的算法识别并跟踪环境中的物体^[16-18]。这种跟踪方式可以进一步分为基于标签和无标签(即基于特征)的方法。无标签法又称自然特征法,通过计算机视觉算法逐帧跟踪纹理特征点,这需

要大量的计算资源^[19]. 市面上也有部分商用 AR 头盔^[20]和开发工具包^[21]实现了基于特征的 AR 跟踪. 相较于无标签方法需要大量算力的特点, 基于标签的方法更便捷轻量. 它使用具有易识别特征的标签, 如二维条码^[22,23]或定制图案^[24,25]等, 来存储并展示物体 ID. 近年来, 发光二极管(LED)由于在视觉上不显眼, 且在现实中部署广泛, 也被学界用来尝试作 MAR 标签^[26,27]. 由于传统的标签只作为空间锚点, 不提供双向通信能力, 因此物体需要无线连接(如 Wi-Fi 和 BLE)来弥补空间环境的丢失. 表 1 中列出了现有方法和 RetroAR 的对比. RetroAR 对基于标签的方法进行了改造和创新. 具体来说, 它将单像素 LED 标签扩展到在空间域上分布的 LCD 矩阵. 用户在识别感兴趣的物体时, 只需将读取器指向该物体, 相关信息就会立刻从光标签中检索出来. 虚拟信息可以通过这种方式与物理对象绑定, 从而实现跟踪.

表 1 主流跟踪方法对比

状态	上行类型	标识类型	下行类型	无线连接	定向交互
静态	外观 ^[6,10,28]	视觉	Wi-Fi 或 BLE	√	×
静态	条形码 ^[22,23]	空间编码	Wi-Fi 或 BLE	√	×
动态	单像素 LED ^[26,27]	闪烁模式	Wi-Fi 或 BLE	√	×
动态	BLE ^[7]	隐藏 URL	BLE	√	×
动态	RetroAR	时空编码	VLBC	×	√

2.3 可见光逆反射通信

可见光逆反射通信是近年来提出的一种新兴的物联网连接解决方案. 其通过反射并调制入射光的方式实现无源通信. 与基于射频的技术相比, 它比有源 VLC^[29]节省 100 倍的能量, 并在安全和抗干扰方面具有天然的优势^[30,31]. RetroVLC^[8]首次提出通过组合逆反射材料和 LCD 调制器, 利用开关键控 (On-Off Keying, OOK) 来实现调制. 随后, 学界通过使用多个小型 LCD 快门来产生多级信号的 PAM 调制方案, 来进一步扩展该技术的速率以应对信道容量限制^[32]. PassiveVLC^[9]用编码协同设计来应对符号长度的限制, 达到了 4 倍于 OOK 的速率. 这些工作使用定制的读取器作为光信号的发射器, 光电二极管作为光信号的接收器. 虽然这种设计有利于建立纯粹的通信系统, 但不利于人机交互系统. 对于交互来说, 发射器应当与常用的手机、AR 眼镜等移动设备整合, 以最大化其适用范围. RetroAR 利用并改进了传统的可见光逆反射通信模型, 将其迁移到人机交互场景, 为定向交互服务. 它利用移动设备摄像头的空间分辨率, 来解码标签上的 LCD 矩阵. 光标签可以很容易地集成到公共服务和基础设施中随处可见的 LCD 屏幕上, 因此具有良好的部署性. 此外, 与传统的 VLBC 设计相比, RetroAR 用闪光灯取代了读写器的发射单元, 并将后置摄像头作为接收单元. 通过这种方

式, RetroAR 可以直接为搭载在商用移动设备上的 MAR 应用, 提供了即时的“所指即所控”的互动体验.

3 系统设计

在 RetroAR 中, 数据在移动设备和光标签之间通过光学媒介进行双向交换. 这种交互方案利用用户的空间环境来消除目标选择中不确定性带来的额外开销, 带来更流畅的用户体验. 该系统包括两个主要部分: (1) 承载无连接 VLBC 数据链路的硬件组合, 如图 2 所示, 包括读取器和光标签; (2) 创新的软件设计, 能够实现高效稳健的标签跟踪、解码, 同时实现 6-DoF 姿态估计. 本文首先介绍 RetroAR 的硬件组合, 揭示底层的物理机制和硬件的设计理念, 接着给出软件算法流水线的工作过程, 同时说明整个软硬件系统如何协调完成交互任务.

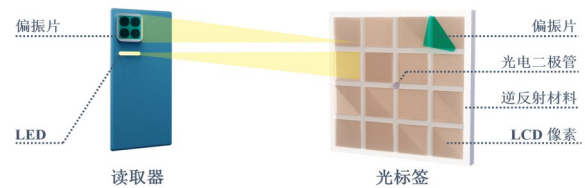


图 2 RetroAR 硬件组合概况

3.1 硬件设计

硬件组合包含两个单元, 即移动读取器和光标签 (也称 ViTag). 二者中每个都是独立的光收发器, 包含完整的光发射单元和光接收单元. 表 2 对各个单元的详细构成进行了总结.

表 2 系统硬件组成

模块类型	发射单元	接收单元
读取器	LED	摄像头
光标签	LCD 矩阵	光电二极管

3.1.1 读取器

读取器可以是任何常见的带有发光二极管和摄像头的移动设备, 如平板电脑、智能手机等. 二极管发出的光束既可用于携带数据位信息, 同时又可用于照亮光标签. 在本文所搭建的原型系统中, 决定采用智能手机作读取器. 背面闪光灯中的 LED 以高频率开启和关闭, 用于向光标签传输数据位, 而背面的摄像头则用于捕捉光标签并从中提取返回的数据位. 为了减少如眩光、反射光等环境噪声, 本文利用摄影中偏振片的概念, 在相机的顶部粘附一层偏振片. 由于反射光和眩光等常常具有单一的偏振方向, 故偏振片可以一定程度上消除这些噪声的影响. 这层偏振片也有助于增加相机视野中标签周围的局部对比度, 从而更有利于标签的检测和跟踪.

这种读取器设计利用了智能手机、平板电脑和智能眼镜等商用设备上的常见传感器.使用时只需要进行简单的软件更新即可,不需要对硬件进行任何修改.这保证了本系统可以最大限度与当前 MAR 应用兼容.

3.1.2 光标签

光标签自身采用刚性 LCD 实现,由于自身刚性表面平整的特性,可以用于跟踪和位姿估计.标签针对 MAR 中场景的动态跟踪场景,分别在通信单元设计和空间布局设计上作出了创新和优化.

(1)通信单元设计:刚性光标签用光电二极管作接收单元,LCD 矩阵作发送单元.LCD 矩阵具有多个像素,每个像素可以独立地打开和关闭,借此嵌入数据位.整个通信单元与 3.2 节中的软件算法协同工作,实现无连接的定向光学数据共享.该通信单元设计区别于传统的静态条形码或者单像素 LED 的设计,引入了多 LCD 像素来进行空间编码,能够带来更稳健的追踪性能以及位姿估计能力.

(2)布局设计:为了让光标签更加适应 MAR 场景,本文作出了 3 个创新的空间布局设计.首先,与单个 LED 像素标签相比,光标签上的 LCD 像素数量是可变的,可以根据应用场景进行定制.像素总数是一个完全平方数,如 1、4、9 等,以便保持矩阵的整体形状是矩形来帮助三维位姿估计.相应地,平方根(1、2、3 等)称为矩阵的等级.第二,在整体布局的最外环留出了空白,这部分空白可以暴露出来 LCD 矩阵下面所附着的逆反射材料,这一圈逆反射材料会形成明亮的外边框,帮助光标签更容易被识别,从而提高跟踪算法的效率.第三,剥离了 LCD 像素上表层偏振片,并将其黏附在镜头表面.这样,LCD 像素的闪烁不可被人眼感知,消除了裸眼时闪烁的影响;同时,由于反射和眩光噪声往往是单方向偏振,因此偏振片的引入也消除了正交方向的噪声,进一步保证了追踪算法的稳定性.

如图 3 所示,光标签有两个部分:接收单元,接收来自读取器的信息(下行链路);发送单元,向读取器发送信息(上行链路).刚性光标签采用 LCD 像素矩阵,单像素价 9.9 元.两种矩阵下方均附着 3M Scotchlite 8910 逆反射材料.接收器采用 Hamamatsu S6036 接收信号,后接一个带高频滤波器的跨阻放大器电路以及一个电压比较电路,放大器型号为 TI OPA344.共设计标签 3 种,包括一个 15.5 cm × 8.7 cm 的与商用 LCD 屏幕集成的 8 级标签,一个 5.8 cm × 6.6 cm 的分体 4 级标签,一个 12.8 cm × 12.8 cm 的分体 4 级标签;其中,集成光标签所用的商用液晶屏是 ALIENTEK ATK-7,其前偏振片被剥离以增强抗噪性能.

3.2 软件设计

如同上文,为简单起见,使用智能手机作为读取器



图 3 RetroAR 所使用的光标签

来演示系统的软件设计.同时使用 4 级光标签,因为像素总数是 8 的倍数,适合嵌入字节数据.如图 4 所示,手机和光标签之间的双向通信可以分为六个步骤.(1)智能手机按照时域编码打开和关闭闪光灯来唤醒光标签;(2)光标签被唤醒后,将准备返回信息,并将信息以时空域混合的方式编码到各个 LCD 像素;(3)摄像头捕捉 LCD 矩阵的多帧图像序列;(4)跟踪算法从单帧图像中识别并分割出光标签,同时确定光标签的等级以及 6-DoF 位置;(5)算法持续跟踪光标签的位置变化并实时解调返回信息;(6)信息向上层传递并生成相应的结果用于不同的用途,例如叠加虚拟菜单到用户 UI.接下来详细介绍系统的软硬件如何配合完成每个步骤.

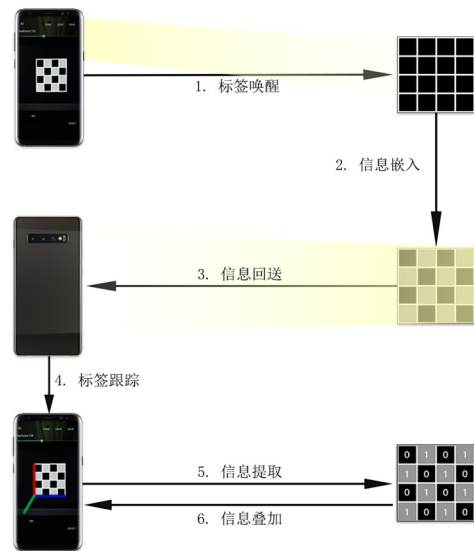


图 4 手机和光标签之间的双向通信

3.2.1 标签唤醒

为了触发通信过程,手机通过闪烁闪光灯发送唤醒请求,这条数据链路称为下行链路,其帧格式如图 5 所示.高电平代表闪光灯开启,低电平代表闪光灯关闭.“前导码”字段用于识别下行链路帧,“类型”字段帮助光标签区分收到的请求类型(即写或读)。“ID”字段帮助光标签过滤掉不发给自己的请求.默认情况下,“类型”被设置为读取模式.一旦光标签检测到前导码,它将按照帧格式提取信息,并按后续介绍的步骤向手机返回其身份.需要注意的是,为了防止环境光造成光标签接收单元饱和,接收单元附加了高通滤波器.因此,下行链路使用了差分编码来帮助消除直流分量.



图5 下行链路帧结构

3.2.2 强抗噪时空混合编码

收到请求后,光标将准备返回信息.首先需要对其信息进行编码,将数据位嵌入到LCD矩阵中.可见光信道由于复杂的背景和无法预测的环境光,充满着大量的噪声.由于光标采用了具有一定空间面积的动态矩阵形式,比起传统静态条形码标签多了时间域上的动态信息,比起传统单LED标签多出了空间域上的额外信息.因此,如何利用光标的这种时空域额外信息来增强编码的抗噪声性能成为一个设计难题.

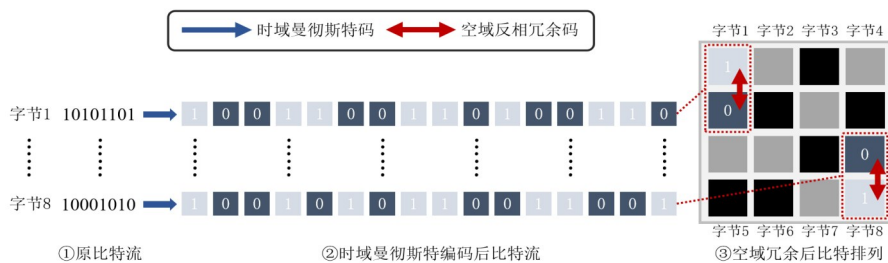


图6 时空混合编码

数据编码流程如下:8个字节可以并行传输.每个字节的8位比特数据首先经过时域曼彻斯特编码,生成16位曼彻斯特码比特流.随后,16 bit的每1 bit按照上述位置规定放在空间帧中,并在其上方相邻像素放置相反的比特以进行冗余编码.每一个空间帧实际上是8个字节各自比特流中的1 bit组成的,而每一组像素对经过冗余编码,实际上等效于在空间上进行了第二次曼彻斯特编码.

3.2.3 信息回送

LCD矩阵之下附着有一层逆反射材料,它将闪光灯光束反射回智能手机的后置摄像头.LCD像素自身依照编码规则改变光学性质,从而改变回送的信息比特.需要注意的是,整个过程都是以被动而非主动发光的方式形成可被摄像头感知的明暗图案,因此返回光束很好地保持了用户自身的方向性.以这种方式发送数据帧的链路称为上行链路.

3.2.4 基于凸包络的光标签快速跟踪算法

为了提供流畅的交互体验,目标标签的跟踪算法需要既稳健(即准确地确定光标每个焦点的坐标),又高效(即每帧计算时间足够地来保证帧率大于60帧).为了实现这一目标,本文必须解决两个挑战——(1)如何在背景嘈杂图像序列中检测和定位光标; (2)如何减少检测和定位的时间,以便在计算能力

针对使用LCD的刚性光标,本文提出“时空域混合编码”的概念,来实现利用光标额外的时空域信息.如图6所示,信息流依照比特拆分并均匀排布到整个LCD面板上.每个字节信息覆盖1个像素在时间轴上的16个帧,帧间亮度变化采用曼彻斯特编码^[33,34].原始数据中的“0”由0到1的电平转换表示,反之亦然.曼彻斯特码可以保证单像素的连续在时间域上不会产生直流分量从而漂移;而不同字节依次向后排布到空间轴的不同像素上.同时,考虑到摄像头自身动态范围以及环境噪声的影响,每个字节按照约定的空间排布顺序固定放置在一个空间帧中的一组像素对上.如图6最右侧中字节1安排在标签左上角的像素对上.每个像素利用自身上侧的相邻像素进行冗余编码,即该相邻像素与原像素之间数值相反.

有限的移动读取器上实现高帧率跟踪.本系统提出两个新颖的设计来应对这些挑战,包括(a)消息级动态兴趣区;(b)快速凸包络检测.

(1)光标二维跟踪:如图7所示,当拍摄图像到达读取器缓冲区后,第一步需在整个图像中找到光标.本文观察到,当用户手持手机时,每一帧中的光标的位置都与前一帧中的位置紧密相关.光标的这种时间上的连续行为带来一点启发——可以裁剪不必要的区域,只保留最后已知的光标的位置的附近感兴趣的区域(Region Of Interest, ROI).将能够成功提取光标位置的帧称之为有效帧,而因噪声等原因未能识别光标的帧称为存续帧.当有效帧到达时,首先裁剪最后已知的ROI以消除额外的噪声,然后尝试在裁剪的区域内定位光标.如果找到光标,就把ROI更新为当帧中标签轮廓外扩矩形.外扩矩形的大小动态得根据当前有效帧和上一次有效帧之间光标移动的幅度进行改变,如果幅度过大则适当扩大外扩矩形;如果没有找到,则继续使用最后一个已知的ROI,并将当前帧置为存续帧.存续帧本身并不能提供最新的光标位置信息,只可以按照预测的ROI位置寻找光标.总的来说,ROI在消息传输过程中是动态更新的,一旦完成一次消息传递就会重置.这种动态更新的消息级ROI可以有效地跟踪视野中光标的轨迹,同时最大限度地减少计算开销.

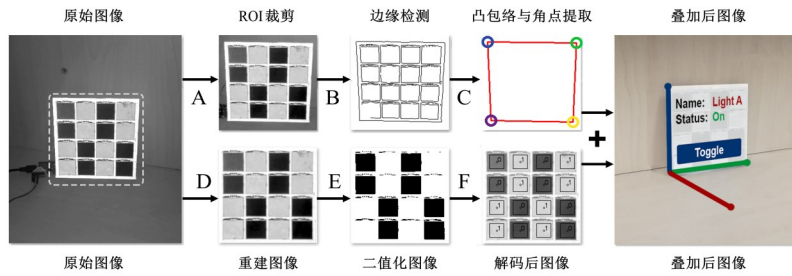


图7 跟踪和解码算法

下一个步骤是从 ROI 中重建出光标签的区域. 具体来说, 需要找到光标签的 4 个焦点的坐标, 并对视角进行变形, 重建出一个正方形网格来提取数据. 首先使用 Canny 边缘检测算法^[35]从图像中提取边缘信息, 明亮外环的存在使得该检测可以找到光标签的矩形轮廓. 随后, 筛选出其中合法的封闭轮廓^[36], 不合法的封闭轮廓直接被排除. 为了进一步缩小选择范围, 用 Hough 变换^[37]排除边线扭曲过大的封闭轮廓, 同时离最后已知位置太远的轮廓或者非矩形, 或者在矩形内部的轮廓会被排除. 经过上述多步排除之后, 寻找凸包络 (Convex hull^[38]) 来找到矩形的 4 个顶点. 最后编写快速映射算法, 根据 4 个顶点之间的相对位置进行划分, 来将提取的 4 个角点与标签自身的顶点相对应. 最后, 对所提取的区域图像进行变形得到重建的光标签区域.

(2) 光标签三维跟踪与位姿估计: 在上一步中, 通过确定 4 个角的二维坐标来定位所捕获图像中的光标签. AR 锚点需要能够帮助读取器进行 3D 位姿估计. 因此, 在上述信息的基础上本文进一步去对标签进行位姿估计. 这个问题可以转化为给定摄像头的内参数矩阵作为先验知识, 根据 2D 画面中物体的角点来推算其三维位姿. 在计算视觉中, 这被称为 Perspective-n-Point 问题^[39,40]. 考虑到 RetroAR 中所使用的刚性标签以及移动读取器摄像头问题, 该问题可以重新描述如下:

给定世界坐标系下 n 个三维点坐标 $C_1^3, C_2^3, \dots, C_n^3$, 及他们所对应的摄像头图像中二维投影的二维坐标 $C_1^2, C_2^2, \dots, C_n^2$. 同时, 给出读取器摄像机自身的内参数矩阵 \mathbf{K} , 根据上述映射关系和内参数矩阵来估计出摄像机自身相对于世界坐标系的 6-DoF 三维位姿, 包括 3 个轴 (x, y, z) 的平移量和 3 个方向 (滚动角、偏航角及俯仰角) 的旋转量.

读取器的摄像头可以通过透视投影模型^[41]来分析. 首先, 本文对真实世界三维点与摄像头图像中二维投影, 根据透视投影模型可以得到:

$$sC^2 = \mathbf{K}[\mathbf{R}|\mathbf{T}]C^3 \quad (1)$$

其中, s 为标量放大系数, C^2 为二维投影点, \mathbf{K} 为相机内矩阵, \mathbf{R} 和 \mathbf{T} 分别为旋转矩阵与平移矩阵, C^3 为世界坐标系下三维点坐标. 展开后可以进一步得到:

$$s \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} f_x & \gamma & \mu_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad (2)$$

RetroAR 场景下, 摄像头内参数是未知的. 通常来讲, 摄像头需要特制的棋盘图形来进行校正. 但是, 由于光标签特殊的编码特性, 其单帧中的空间编码本身就可以呈现出特殊的棋盘样式. 因此, 初始状态下本文直接拍摄光标签, 并利用亚像素焦点检测法提取摄像头的内参数矩阵^[42]. 获得摄像头的内参数矩阵后, 进一步使用 EPnP 算法^[43]来确定每一帧中光标签的 6-DoF 三维姿态. 具体的, 对于三维空间中的 n 个点 $C_1^3, C_2^3, \dots, C_n^3$ 其都可以表示为 4 个预定义点 $P_1^3, P_2^3, \dots, P_4^3$ 的加权和, 即

$$C_i^3 = \sum_{j=1}^4 \alpha_j P_j^3, \quad \sum_{j=1}^4 \alpha_j = 1, \quad i = 1, 2, \dots, n \quad (3)$$

同理, 其所对应的二维投影点, 亦可表示为 4 个二维参考点的加权和:

$$C_i^2 = \sum_{j=1}^4 \alpha_j P_j^2, \quad \sum_{j=1}^4 \alpha_j = 1, \quad i = 1, 2, \dots, n \quad (4)$$

投影点经过摄像头畸变后得到的画面点位

$$s_i C_i^{\text{img}} = \mathbf{K} \sum_{j=1}^4 \alpha_j P_j^2 \quad (5)$$

上个步骤中提取出了 4 个角点, 每个角点坐标对应到上述投影模型实际上就是 $C_i^{\text{img}} = (u_i \ v_i \ 1)$, 而参考点的其次坐标是 $P_j^2 = (x_j^2 \ y_j^2 \ z_j^2)$. 将预定义点坐标代入上式后得到两个方程:

$$\sum_{j=1}^4 \alpha_j f_x x_j^2 + \alpha_j (\mu_0 - \mu_i) z_j^2 = 0 \quad (6)$$

$$\sum_{j=1}^4 \alpha_j f_x y_j^2 + \alpha_j (v_0 - v_i) y_j^2 = 0 \quad (7)$$

由于光标签 4 个角点可以提取出 4 对映射关系, 代入方程后得到线性方程 $\mathbf{M}\mathbf{X} = \mathbf{0}$. 其中矩阵 \mathbf{X} 即是 4 个参考点的坐标, 且 4 对关系可以保证具有唯一解. 将解算出的参考点代入上述畸变方程后, 即可解算出矩阵 \mathbf{R} 和 \mathbf{T} . 解算完成以后, 本文可以得到摄像头相对于光标签的三轴位移以及三自由度旋转角度. 对此关系进行

逆变化即可解算出标签相对于摄像头坐标系的位置,在此基础上可以将相关信息进行透视变换并以正确的视角叠加到光标签上,实现 AR 全息图效果,同时可以实现利用光标签对用户进行定位.

3.2.5 自适应时空混合解码

图像重建完成以后,本文需要获取嵌入到 LCD 像素中的比特信息.因此,本文需要适当的阈值来对图像进行二值化分割,以获得 LCD 的对应状态.然而,多变的环境光条件和各样的闪光灯功率导致所拍摄的图像性质不可预测,反光和环境中的灯光等噪声也会一定程度上影响像素的实际亮度,导致固定的阈值无法适用于所有情况.需要寻找一个合适的自适应阈值算法,因此,本文提出一种新的自适应解调算法来适应这种编码机制,如算法 1 所示.

算法 1 自适应时空混合解码

输入:一个消息中的多个帧 f_1, f_2, \dots, f_n

输出:每帧中每个像素的解码结果 $D_n(i, j)$

FOR $i = 1, 2, \dots, n$:

运行凸包跟踪算法获得重建区域 R_i

各坐标像素块亮度,取中心区域 σ 半径内的高斯均值记为 $G_\sigma(i, j)$

按照棋盘排列,将各像素和其对应的冗余反相像素组成像素对

$(G_\sigma, G_\sigma)_{1,2,\dots,m}$

将所有像素对 $(G_\sigma, G_\sigma)_{1,2,\dots,m}$ 为直方图 H

对直方图运行 K -means 聚类算法获取该帧图像的最佳阈值
依据上部所求最佳阈值对该帧像素块进行分割和信息解码,得到结果 S_i

对位置 (i, j) 像素构成的亮度序列 $G_\sigma(i, j)_1, G_\sigma(i, j)_2, \dots, G_\sigma(i, j)_n$ 进行解码

将时域单像素解码结果与空域单帧解码结果进行比较,给出最终解码结果

单帧情况下,可以对像素的亮度信息进行统计获得其亮度直方图.由于冗余对的存在和棋盘排布,直方图会呈现良好的分界线.在此基础上,为了获得像素间最佳分割阈值,对它的直方图应用 K -means 算法进行聚类,得到最佳的二值化阈值.如图 8 所示,该阈值可以保证对该单帧图像的所有像素块进行最佳的阈值分割.得到阈值后,利用高斯算子计算每个 LCD 像素中心点周围的方形区域的平均亮度.平均亮度高于阈值的像素将被解调为“1”,反之亦然.单帧的解码结果暂存不使用.

一个消息中包含多个帧.首先对各个帧进行上述空域解码.在得到各个帧自身的空间解码信息后,按照像素位置将序列帧中各像素的亮度变化按照像素位置串成多个序列 $G_\sigma(i, j)_{1,2,\dots,n}$.随后,每一个序列,实际上都是一个时间轴上的曼彻斯特编码序列,因此相应地采用动态阈值法进行曼彻斯特解码.时域和空域解码完成后,对于位于 (i, j) 的像素,其在时间 m 时分别出现

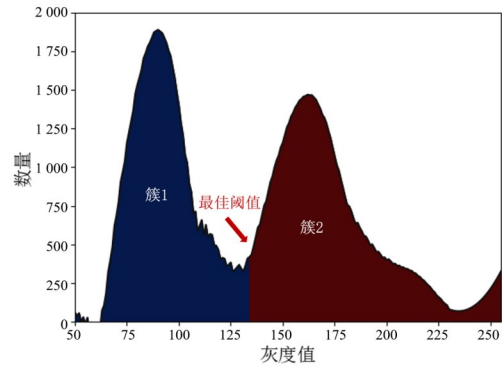


图 8 K -means 聚类算法获取像素间最佳分割阈值

两个解码结果,即空域解码 $S_s(i, j)^m$,以及时域解码 $S_t(i, j)^m$.本文分别对空域和时域所使用的分割阈值进行记录,并将每个像素自身亮度与阈值的距离 $d(i, j)$ 进行归一化,得到一个权值 w 作为该像素解码的置信度,亦可理解为该像素解码结果正确的可信度.距离越大,说明分割结果越可信.最终解码结果使用对上述两个结果综合得到,即

$$S_{\text{fin}}(i, j)^m = \begin{cases} S_s(i, j)^m, w_s > w_t \\ S_t(i, j)^m, w_s \leq w_t \end{cases} \quad (8)$$

4 系统实验

4.1 实验设置与步骤

在系统实验中,使用华为 Mate 30 Pro 作为首选读取器.使用闪光灯将信息发送到光标签(下行链路),并使用后置摄像头来捕捉返回信息(上行链路).相机设定为 60 帧,分辨率为 720 p.光标签垂直放置在深绿色背景前,智能手机安装在面向光标签的三脚架上.本文使用了两个分体刚性 4 级光标签来评估标签尺寸的硬响.实验器具设置如图 9 所示.



图 9 系统实验的部署以及所使用的设备

本文评估了不同条件下,包括距离、角度、光线条件和闪光灯频率等指标.每种条件下,向光标签发送 10 次预定义信息,测量平均比特错误率 (Bit Error Rate, BER).本文还测量了两种光标签在传输过程中的平均能量消耗.随后,分别在两个链路上传输 32 字节的信息来测量下行链路和上行链路的最大数据速率,取 10 s 的平均数据率作为结果.最后,为了评估 6-DoF 位姿估计性能,使用等尺寸的 ArUco 标签^[23],结合 OpenCV 的 ArUco 库^[44]来获取真实的平移和旋转数

据,并与算法的估计结果相对比。

4.2 实验结果

图 10(a)显示,最大通信范围随着所选摄像头分辨率的提高几乎呈线性增长。这是因为更高的分辨率能够捕捉到更多的像素级细节,从而使摄像头视图中更小的光标能够成功解码。摄像头的分辨率与最大上行距离呈现出相关性。当使用较高的分辨率如 2K 时,大刚性光标通信距离可以达到 7 m。然而,这势必会过度消耗移动设备的计算资源,导致更长的处理时间。对于 1080P 这一目前商用智能手机和平板电脑上大多数摄像头的标准分辨率,距离可达 4 m,能够满足房间规模的互动应用。因此,RetroAR 的设计保证了在大多数移动设备上都能保持足够的通信距离,支持大多数房间规模的应用。

图 10(b)显示,在估计光标的 X 轴和 Y 轴平移时,RetroAR 的误差中值为 1.0 cm 和 1.3 cm。对于 Z 轴,中位数明显更高,达到了 2.5 cm。原因是当光标沿 Z 轴移动时,它更容易受到眩光和垂直方向环境光源的反射

影响,因此降低了焦点定位的精度,进一步影响了最终的坐标输出。本文可以通过沿着 Z 轴调整光标签的布局,例如添加盖子阻挡来对抗噪声。

图 10(c)显示,RetroAR 的滚动、偏航和俯仰角度的中值误差分别为 4.7° 、 5.1° 和 7.5° 。同样,俯仰角的性能下降是由于倾斜的光标签引入了更多环境光噪声源的反射。总的来说,该系统能够实现稳健和准确的 6-DoF 跟踪,这对基于位置的 MAR 应用非常重要。因为这些应用基于用户的空间环境选择性地提供和传递信息,因此定位的准确性直接影响到了最终的用户体验。

表 3 显示了尺寸、能源消耗和传输距离之间的权衡。对于光标来说,能耗与光标的大小成正比,因为驱动具有更大像素面积和更多像素的大型光标需要消耗更多的能量。在传输状态下,两种刚性光标的最大消耗量分别为 2.82 mW 和 12.17 mW。与基于 Wi-Fi 的连接相比,降低了 2 倍的功耗^[45]。小标签可以在标准的 1 740 mAh 锂离子电池上运行 3 个月以上,这足以应对许多需要长期部署的交互式应用,如实景 MAR 游戏等。

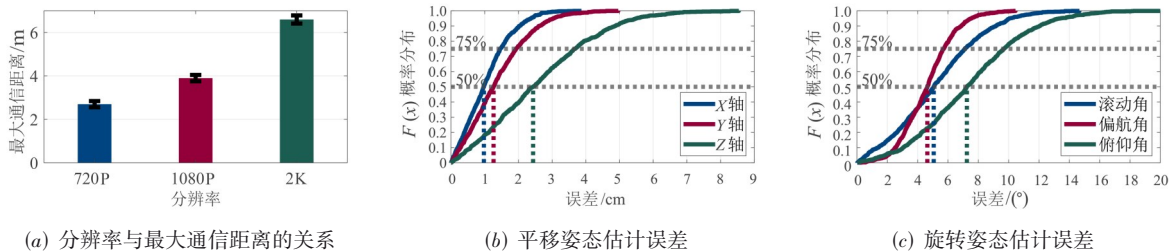


图 10 其他实验结果

表 3 光标签的能耗和通信距离

种类	面积/cm ²	能耗/mW	最大上行距离/cm
小刚性光标	38.00	2.82	140.00
大刚性光标	163.84	12.17	275.00

5 用户研究

用户研究旨在找出 RetroAR 与目前基于 Wi-Fi、BLE 的非接触式控制相比,在总交互时间和整体交互体验方面的表现差异。针对交互时间和体验问题,本文设置了两个场景:(1)MAR 智能家居控制;(2)基于 MAR 的非接触式电器互动。

5.1 用户选择与实验设置

本文招募了 12 名年龄在 20 岁到 47 岁之间的参与者(6 名女性),中位数为 24 岁。每位参与者得到了 50 元的补偿。实验使用的光标签:智能灯和智能音箱采用分体式的光标签连接,废弃电梯面板屏和自动售货机屏采用整合式的光标签连接,机器人采用分体式光标

连接。场景测试点相距 2 m。光标都配备了额外的 Esp32 MCU 电路板来提供无密码的 Wi-Fi 连接,并在近处通过二维码提供 UUID 来帮助连接;同时,光标也预先通过 BLE 与实验用读取器建立连接,并唯一指定了名称方便语音控制区分指令对象。若参与者需要在 VLBC 链接上发送数据,需要将手机指向光标并点击安卓 APP 中的按钮;若参与者要通过 Wi-Fi 发送数据,需扫描二维码以连接到 Wi-Fi,然后在安卓 APP 中选择目标对象。若参与者需要通过 BLE 发送数据,需按下 APP 中的语音键说出指令并松开发送。本文为参与者提供了预先装好应用的两部智能手机和一部平板电脑作为移动读取器,包括华为 Mate 30 Pro(安卓版本 10.0)、谷歌 Nexus 6P(安卓版本 8.0)和 LENOVO Tab P11(安卓版本 11.0)。参与者可以任选其一来完成互动任务。摄像头帧率设置为 60 帧,闪光灯频率为 90 Hz。在整个实验过程中,用户站在光标签约 2 m 的位置。

5.2 试验步骤

两个场景目的为定量研究 RetroAR 与无线方式的对比,参与者被要求依次参与这两个场景,每个场景重复 6 轮. 对于智能家居场景,他们需要按照随机生成的 3 个控制动作序列(例如打开或关闭智能灯、调高或调低智能音箱的音量),向现场的智能灯和音箱发送控制命令. 对于无接触电器互动场景,他们需要依次完成两个任务:(1)在自动售货机的 10 个候选物品中购买特定的物品;(2)在电梯面板上选择特定的楼层. 在每一轮开始之前,参与者从软件生成器中得到两个数字的随机数,每个数字从 1 到 10 不等. 第一个数字代表要在自动售货机里买哪种物品,第二个数字代表要在电梯里去哪一层. 这些随机数剔除了用户选择物品带来的时间影响.

对于 Wi-Fi 控制,参与者应在扫描二维码连接到 Wi-Fi 后,点击界面中的相应按钮,通过本地网络服务器控制每个智能物体;对于 BLE 控制,由于无线通信时延与 Wi-Fi 相当,故主要探究与其常搭配的语音控制所带来的时间开销与 RetroAR“所指即所控”的对比,因此参与者需要在按下语音按钮后说出对应指令完成相应动作;对于 RetroAR,参与者需将读取器指向光标签,并点击 APP 上的按钮来完成交互.

本文对参与者在每一轮中使用的方法顺序(Wi-Fi、BLE、RetroAR)进行了平衡化处理. 参与者在实验开始前有 5 min 的辅导时间以熟悉如何使用应用程序完成操作. 本文定义互动时间为从参与者按下开始按钮到所有动作成功完成的时间. 每个参与者参与 6 轮并计算平均完成时间结果. 在完成整个实验后,参与者被要求完成一份包含 NASA-TLX 任务负荷指数的调查问卷. 问卷包含 6 道 7 分李克特量表问题,用于测量每种方法的工作量. 鉴于 Wi-Fi 与 BLE 在交互方式和数据传输上的相似性,问卷中仅针对 Wi-Fi 和 RetroAR 两种方法进行对比,用来探究 RetroAR 的“所指即所控”的交互方式带来的主观影响. 对于任务表现问题,“1”表示“失败”,“7”表示“完美”;对于其他问题,“1”表示“非常低”,“7”表示“非常高”. 此外,试验结束后还设计了 3 个开放性问题邀请参与者回答:(1)与基于 Wi-Fi 的方法相比,你认为 RetroAR 表现如何?(2)与基于 BLE 的语音方式相比,你认为 RetroAR 表现如何?(3)你认为 RetroAR 有什么可改进之处?

5.3 实验结果

5.3.1 场景完成结果

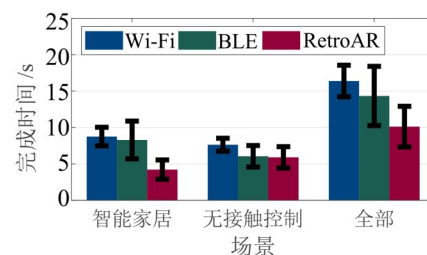
图 11(a)显示完成两个场景的单场景时间和总时间. 利用 Welch's *t* 检验来检验同一场景下不同方法间完成时间的分布差异. 以智能家居场景为例,首先作出假设 u_0 : Wi-Fi 与 RetroAR 完成时间的均值相等. 对两

种方法的完成时间数据进行 Welch's *t* 检验,规定显著性水平为 0.05. 若结果所得 p 值大于等于 0.05,则接受该假设,即认为二者均值相等,否则拒绝该假设. 文章后续检验步骤均相同,故仅报告 p 值来显示检验结果. 结果显示,智能家居场景的 p 值分别为 0.002(Wi-Fi)与 0.002(BLE),无接触控制场景 p 值为 0.004(Wi-Fi)和 0.003(BLE). 两个场景的结果中 RetroAR 与其他两类方法均存在统计学显著的差异,表明使用 RetroAR 时用户单场景完成时间较短. 这意味着 RetroAR 实现的定向互动可以显著减少总互动时间,在多动作切换的场景尤为明显. 相比较之下,RetroAR 的时空域混合编码和光标签的光学特性可以让用户进行快速实现“所指即所控”式的通信.

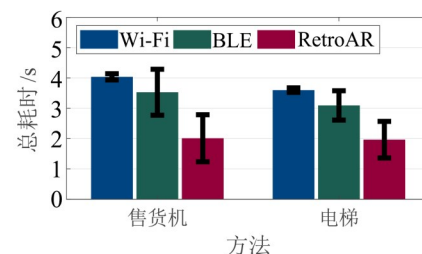
图 11(b)显示了在非接触式控制场景中排除系统延迟(即由安卓系统和底层通信逻辑造成的延迟)后的交互时间,该事件反映了用户在发送控制命令之前单纯由于互动造成的开销. 可以看到在每个任务中,基于 Wi-Fi 方法的开销是 RetroAR 的 2 倍以上,而基于 BLE 语音的方案也有较大的时间开销以及更大的标准差. 此外本文还发现,与基于 Wi-Fi 的方法相比,RetroAR 的开销具有更大的标准差. 这是由于在 RetroAR 中,集成的光标签不像二维码那样明显,用户需要花费更多的精力来调整相机的位置,使闪光灯光束与光标签上的接收器对齐. 这启发在未来的设计中寻找让接收器区域更加显眼的方法.

5.3.2 NASA-TLX 问卷结果

图 12 展示了 NASA-TLX 任务负荷调查表的结果.



(a) 场景一至二完成时间



(b) 场景二排除系统延迟后的耗时

图 11 场景一至二结果

Wilcoxon Signed-rank 检验显示,相较于 Wi-Fi, RetroAR 的物理工作量更少, p 值为 0.03. 这符合 RetroAR 的设计理念,因为与 Wi-Fi 方法相比 RetroAR 不需要用户建立无线连接和扫码,直接指向目标光标签即可实现交互. 此外,一个令人惊讶的结果是,虽然本文预计两种方法完成非接触式控制任务时候给用户带来的挫败感不会有太大的差别,但实际上 RetroAR 的挫败感实际上显著低于 Wi-Fi 方法 (p 值为 0.04). 本文向部分参与者询问了原因. P2 评论道:“虽然扫描二维码或直接连接到 Wi-Fi 非常普遍,但这种服务的公共 Wi-Fi 往往需要额外的步骤进行用户认证,如果只是想简单地用一部电梯的话真的很烦人. 认证过程需要输入很多字符还需要等待短信接收,这就与本身快速交互的初衷背离了.”由此可见,基于 Wi-Fi 的方法容易受连接设置的影响,其带来的额外开销会降低交互体验,因此与 RetroAR 相比用户挫折感更强.

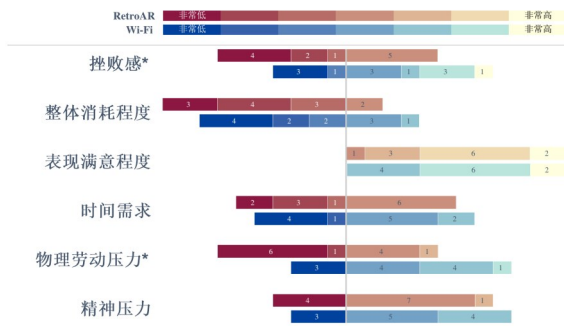


图 12 问卷结果

注:标*号意味着统计学显著

此外,问卷调查显示,在心理需求和努力方面,Wi-Fi 控制和 RetroAR 之间没有明显的区别. P4 说:“我认为 RetroAR 确实减少了完成任务所需的努力,但综合来讲比起 Wi-Fi 区别没有那么大. 对我来说身边的一切都需要扫码,所以这个事情并不难. 就个人而言,RetroAR 的点对点互动方案对老年人确实有帮助,因为它更直观,更接近本文理想中的自然互动方式.” P7 提到:“我已经习惯了在手机列表中选择目标设备来与物联网设备互动. 与此相比,扫描二维码连接不是什么问题. 我认为操作上来讲心理负担并不重,但是想象一下充满二维码的世界确实很碍眼很不美观.”这些评论表明,鉴于 Wi-Fi 连接和扫码已经成为生活常态,尽管 MAR 交互有其独特的简单性和便利性,但目前还不能完全取代无线解决方案. 基于无线连接和扫码的方式已经成为人们心中习以为常的交互方式,虽然实验结果已经验证了在交互速度和交互体验方面 RetroAR 有着更多的天然优势,但实际使用中仍然需要一段过渡时间来使人们适应这种更贴合直觉的交互方式,其部署也需

要更多推广上的努力.

6 讨论及展望

当前版本的 RetroAR 对 MAR 场景中的定向互动有良好的帮助. 然而,系统仍然有较大的局限性和可改进的空间. 目前原型系统通信范围有限,最远达到 4 m,支持房间规模的应用. 但是这个通信距离使用户无法控制更远处的设备,无法做到建筑规模的应用. 未来的工作可以探索如何重新设计光学接收单元和发射单元,绕过闪光灯自身和安卓操作系统的软硬件限制,延长下行通信距离至大于 5 m;同时进一步改良跟踪算法,使其在高分辨率(4K)下实时运行,保证上行链路也具有较大的通信范围. 此外,目前的数据速率约为 400 bps,足以满足 MAR 非接触式控制等应用. 对于一些具有高达 720 fps 的超高速视频拍摄模式的高端智能手机,这一数据速率可以提升到几 kbps. 在这种数据率下,RetroAR 可以实现定向文件传输共享等新应用. 此外,在手机闪光灯上添加特殊的硬件控制单元来绕过软时钟,有可能将下行数据率提高到几 Mbps 的水平,支持实时的视频流传输. 相信 RetroAR 可以在这类高端智能手机或智能眼镜上用于新颖的 MAR 应用. 例如,第一人称实时 AR 对抗游戏中,玩家使用即时的方向控制来与多个对手进行复杂的互动操作,所有操作所需要的增强模型和附加信息都通过 VLBC 信道实时生成和共享.

7 结论

本文提出了基于可见光逆反射通信的定向交互系统 RetroAR,它为商用移动设备上的增强现实应用带来了自然、快速的定向交互模式,实现了“所指即所控”的快速互动. RetroAR 通过定制的光标签,借助可见光束的方向性利用用户空间环境,在商用移动设备上实现了快速选择目标. 系统实验显示,RetroAR 最大支持在 4 m 的距离上、100° 的视角下进行通信. 12 人参与的用户研究表明,与基于 Wi-Fi 的控制方法相比,RetroAR 在智能家居、非接触控制、多目标人机协作等 MAR 应用中明显改善了交互时间和体验. 本文进一步讨论了 RetroAR 的实际应用与除定向交互外可能的应用, RetroAR 现阶段自身的限制与不足,以及高端智能手机带来更高数据率时潜在的新应用. 本系统证明了利用可见光逆反射通信为 MAR 应用提供即时方向控制的可行性,旨在启发未来探索使用光信号的方向性解决现实世界互动效率和体验问题的研究.

参考文献

- [1] 林惊, 杨珂, 王涌天, 等. 移动增强现实系统的关键技术研究[J]. 中国图象图形学报, 2009, 14(3): 560-564.

- LIN L, YANG K, WANG Y T, et al. Key issues study for mobile augmented reality system[J]. *Journal of Image and Graphics*, 2009, 14(3): 560-564. (in Chinese)
- [2] 陈靖, 王涌天, 郭俊伟, 等. 增强现实技术在智能手机上的应用[J]. *电子科技大学学报*, 2010, 39(S1): 80-84.
CHEN J, WANG Y T, GUO J W, et al. Application of augmented reality technology in smart phone[J]. *Journal of University of Electronic Science and Technology of China*, 2010, 39(S1): 80-84. (in Chinese)
- [3] 王同聚. 虚拟和增强现实(VR/AR)技术在教学中的应用与前景展望[J]. *数字教育*, 2017, 3(1): 1-10.
WANG T J. Application and prospect forecast of VR/AR technology in education[J]. *Digital Education*, 2017, 3(1): 1-10. (in Chinese)
- [4] 汤晓颖, 区雪兰, 谢丹妮. 基于AR技术的广府家具“纸上博物馆”设计应用与研究[J]. *包装工程*, 2018, 39(4): 115-122.
TANG X Y, OU X L, XIE D N. Design application and research of "museum on paper" of Cantonese furniture based on AR technology[J]. *Packaging Engineering*, 2018, 39(4): 115-122. (in Chinese)
- [5] PERLI S D, AHMED N, KATABI D. PixNet: Interference-free wireless links using LCD-camera pairs[C]//*Proceedings of the Sixteenth Annual International Conference on Mobile Computing and Networking*. New York: ACM, 2010: 137-148.
- [6] DE FREITAS A A, NEBELING M, CHEN X', et al. Snap-to-it: A user-inspired platform for opportunistic device interactions[C]//*Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. New York: ACM, 2016: 5909-5920.
- [7] DE FREITAS A A, NEBELING M, RANITHANGAM A S K K, et al. Bluewave: Enabling opportunistic context sharing via bluetooth device names[C]//*Proceedings of the 8th ACM SIGCHI Symposium on Engineering Interactive Computing Systems*. New York: ACM, 2016: 38-49.
- [8] LI J T, LIU A L, SHEN G B, et al. Retro-VLC: Enabling battery-free duplex visible light communication for mobile and IoT applications[C]//*Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications*. New York: ACM, 2015: 21-26.
- [9] XU X Y, SHEN Y, YANG J R, et al. PassiveVLC: Enabling practical visible light backscatter communication for battery-free IoT applications[C]//*Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*. New York: ACM, 2017: 180-192.
- [10] HEUN V, KASAHARA S, MAES P. Smarter objects: Using AR technology to program physical objects and their interactions[C]//*CHI '13 Extended Abstracts on Human Factors in Computing Systems*. New York: ACM, 2013: 961-966.
- [11] TANG L Z W, ANG K S, AMIRUL M, et al. Augmented reality control home (ARCH) for disabled and elderlies [C]//*2015 IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*. Piscataway: IEEE, 2015: 1-2.
- [12] JO D, KIM G J. ARIoT: Scalable augmented reality framework for interacting with Internet of Things appliances everywhere[J]. *IEEE Transactions on Consumer Electronics*, 2016, 62(3): 334-340.
- [13] HUO K, CAO Y Z, YOON S H, et al. Scenariot: Spatially mapping smart things within augmented reality scenes [C]//*Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. New York: ACM, 2018: 1-13.
- [14] 韩玉仁, 李铁军, 杨冬. 增强现实中三维跟踪注册技术概述[J]. *计算机工程与应用*, 2019, 55(21): 26-35.
HAN Y R, LI T J, YANG D. Overview of 3D tracking registration technology in augmented reality[J]. *Computer Engineering and Applications*, 2019, 55(21): 26-35. (in Chinese)
- [15] REKIMOTO J. Matrix: A realtime object identification and registration method for augmented reality[C]//*Proceedings. 3rd Asia Pacific Computer Human Interaction*. Piscataway: IEEE, 1998: 63-68.
- [16] CHOY C B, STARK M, CORBETT-DAVIES S, et al. Enriching object detection with 2D-3D registration and continuous viewpoint estimation[C]//*2015 IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2015: 2512-2520.
- [17] GAO Q H, WAN T R, TANG W, et al. Object registration in semi-cluttered and partial-occluded scenes for augmented reality[J]. *Multimedia Tools and Applications*, 2019, 78(11): 15079-15099.
- [18] 王笛, 胡辽林. 基于双目视觉的改进特征立体匹配方法[J]. *电子学报*, 2022, 50(1): 157-166.
WANG D, HU L L. Improved feature stereo matching method based on binocular vision[J]. *Acta Electronica Sinica*, 2022, 50(1): 157-166. (in Chinese)
- [19] GAO Q H, WAN T R, TANG W, et al. A stable and accurate marker-less augmented reality registration method [C]//*2017 International Conference on Cyberworlds*. Pis-

- cataway: IEEE, 2017: 41-47.
- [20] MICROSOFT. Microsoft hololens I mixed reality technology for business[EB/OL]. (2020)[2022]. <https://www.microsoft.com/en-us/hololens/>.
- [21] GOOGLE. ARCore - google developers[EB/OL]. (2020)[2022]. <https://developers.google.com/ar>.
- [22] GARRIDO-JURADO S, MUÑOZ-SALINAS R, MADRID-CUEVAS F J, et al. Automatic generation and detection of highly reliable fiducial markers under occlusion [J]. *Pattern Recognition*, 2014, 47(6): 2280-2292.
- [23] KAWSAR F, RUKZIO E, KORTUEM G. An explorative comparison of magic lens and personal projection for interacting with smart objects[C]//*Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services*. New York: ACM, 2010: 157-160.
- [24] JUNG J D A, VOGEL D. Methods for intentional encoding of high capacity human-designable visual markers [C]//*Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. New York: ACM, 2018: 1-10.
- [25] COSTANZA E, HUANG J. Designable visual markers [C]//*Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York: ACM, 2009: 1879-1888.
- [26] AHUJA K, PAREDDY S, XIAO R, et al. LightAnchors: Appropriating point lights for spatially-anchored augmented reality interfaces[C]//*Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. New York: ACM, 2019: 189-196.
- [27] YANG J, LANDAY J A. InfoLED: Augmenting LED indicator lights for device positioning and communication [C]//*Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. New York: ACM, 2019: 175-187.
- [28] SUN Y B, ARMENGOL-URPI A, REDDY KANTAR-EDDY S N, et al. Magic hand: Interact with IOT devices in augmented reality environment[C]//*2019 IEEE Conference on Virtual Reality and 3D User Interfaces*. Piscataway: IEEE, 2019: 1738-1743.
- [29] 卢霆威, 王泽平, 刘梦, 等. 基于可见光通信技术的全双工以太网通信系统设计[J]. *电子学报*, 2022, 50(1): 45-53.
- LU T W, WANG Z P, LIU M, et al. Full-duplex indoor ethernet communication system based on visible light communication[J]. *Acta Electronica Sinica*, 2022, 50(1): 45-53. (in Chinese)
- [30] GUMMADI R, WETHERALL D, GREENSTEIN B, et al. Understanding and mitigating the impact of RF interference on 802.11 networks[J]. *ACM SIGCOMM Computer Communication Review*, 2007, 37(4): 385-396.
- [31] YANG J J. Apparatus and Method for Suppressing Interference Caused by Coexistence of WiMAX and Wi-Fi [M]. New York: Google Patents, 2013.
- [32] SHAO S H, KHREISHAH A, ELGALA H. Pixelated VLC-backscattering for self-charging indoor IOT devices [J]. *IEEE Photonics Technology Letters*, 2017, 29(2): 177-180.
- [33] AMIRI I S, ALI J. Data signal processing via Manchester coding-decoding method using chaotic signals generated by PANDA ring resonator[J]. *Chinese Optics Letters*, 2013, 11(4): 41901-41904.
- [34] TAO Q, ZHONG C J, LIN H, et al. Symbol detection of ambient backscatter systems with Manchester coding[J]. *IEEE Transactions on Wireless Communications*, 2018, 17(6): 4028-4038.
- [35] CANNY J. A Computational approach to edge detection [M]//*Readings in computer Vision*. Amsterdam: Elsevier, 1987: 184-203.
- [36] SUZUKI S, BE K. Topological structural analysis of digitized binary images by border following[J]. *Computer Vision, Graphics, and Image Processing*, 1985, 30(1): 32-46.
- [37] DUDA R O, HART P E. Use of the Hough transformation to detect lines and curves in pictures[J]. *Communications of the ACM*, 1972, 15(1): 11-15.
- [38] GRAHAM R L. An efficient algorithm for determining the convex hull of a finite planar set[J]. *Information Processing Letters*, 1972, 1(4): 132-133.
- [39] LEPETIT V, MORENO-NOGUER F, FUA P. EPnP: An accurate O(n) solution to the PnP problem[J]. *International Journal of Computer Vision*, 2009, 81(2): 155-166.
- [40] PERSSON M, NORDBERG K. Lambda twist: An accurate fast robust perspective three point (P3P) solver[C]//*Proceedings of the European Conference on Computer Vision*. New York: ACM, 2018: 318-332.
- [41] WILLSON R G, SHAFER S A. Perspective projection camera model for zoom lenses[EB/OL]. (2022)[2022]. <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/2252/0000/Perspective-projection-camera-model-for-zoom-lenses/10.1117/12.169831.full>.
- [42] YANG X F, HUANG Y M, GAO F. A simple camera calibration method based on sub-pixel corner extraction of

the chessboard image[C]//2010 IEEE International Conference on Intelligent Computing and Intelligent Systems. Piscataway: IEEE, 2010: 688-692.

- [43] LEPETIT V, MORENO-NOGUER F, FUA P. EPnP: An accurate $O(n)$ solution to the PnP problem[EB/OL]. (2022) [2022]. <https://link.springer.com/article/10.1007/s11263-008-0152->.
- [44] OPENCV. Detection of ArUco markers[EB/OL]. (2022) [2022]. https://docs.opencv.org/4.x/d5/dae/tutorial_aruco_detection.html.
- [45] MAHMOUD M S, MOHAMAD A A H. A study of efficient power consumption wireless communication techniques/modules for Internet of Things (IoT) applications [J]. Advances in Internet of Things, 2016, 6(2): 19-29.

作者简介



李景聿 男, 1996 年出生, 山东烟台人。本科毕业于北京邮电大学, 现为北京大学计算机学院博士一年级。主要研究方向为移动增强现实系统、光逆反射通信。

E-mail: motren909@pku.edu.cn



陈拓潮 男, 1999 年出生, 浙江温州人。本科毕业于北京大学, 现为华盛顿大学计算机科学与技术学院博士三年级。主要研究方向为无线通信系统。

E-mail: ctc1998@uw.edu



李明哲 男, 1995 年出生。本科毕业于多伦多大学, 现为卡内基梅隆大学人机交互研究所博士四年级。主要研究方向为人机交互。

E-mail: mingzhe2@andrew.cmu.edu



徐栩海 男, 1995 年出生。本科毕业于清华大学, 现为华盛顿大学信息学院博士五年级。主要研究方向为基于传感器和机器学习的行为感知。

E-mail: xuhaixu@uw.edu



张 铖 男, 1985 年出生。现为康奈尔大学计算与信息学院助理教授。主要研究方向为人机交互和感知。

E-mail: chengzhang@cornell.edu



徐子晨 男, 1985 年出生。现为现任南昌大学数学与计算机学院副院长、计算中心主任。主要从事包括数据密集计算、智能计算、高性能计算及分布式数据存储等数据系统软件可靠性及可持续性等方面的教研工作。

E-mail: xuz@ncu.edu.cn



刘讚哲 男, 北京大学计算机学院副教授、博士生导师。主要研究方向为面向移动计算、大数据、服务计算等网络化应用模式的新型系统软件。中国电子学会会员编号: E190015296M。

E-mail: liuxuanzhe@pku.edu.cn



黄 翌 男, 1975 年 10 月生, 湖南人。北京大学计算机学院副教授、博士生导师。主要研究方向为自适应软件计算反射原理、运行时模型驱动方法、自适应系统的软件定义方法等。中国电子学会会员编号: E190013031S。

E-mail: hg@pku.edu.cn



冯 云 女, 北京大学第三医院主任医师、博士生导师。主要研究方向为角膜病、角膜内皮移植、眼外伤前节重建、白内障玻璃体术后各类眼表疾病。

E-mail: fengyun@bjmu.edu.cn



许辰人 男, 1985 年 11 月生, 上海人。北京大学计算机科学技术系“博雅青年学者”长聘副教授、软硬件协同体系结构(SOAR)实验室主任。主要研究方向为高速无线网络体系架构、低功耗可见光通信系统和情感计算应用。中国电子学会会员编号: E190026855S。

E-mail: chenren@pku.edu.cn