



虚拟现实的人机交互综述

张凤军^{①②*}, 戴国忠^①, 彭晓兰^①

① 中国科学院软件研究所人机交互北京市重点实验室, 北京 100190

② 中国科学院软件研究所计算机科学国家重点实验室, 北京 100190

* 通信作者. E-mail: fengjun@iscas.ac.cn

收稿日期: 2016-10-26; 接受日期: 2016-12-03; 网络出版日期: 2016-12-22

国家自然科学基金 (批准号: 61572479, 61135003)、国家重点研发计划 (批准号: 2016YFB1001403)、国家自然科学基金委员会与新加坡国家研究基金会合作研究 (批准号: 61661146002) 资助项目

摘要 人机交互是虚拟现实的核心技术之一, 对推进虚拟现实广泛应用和提高用户的体验具有重要意义. 由于传感器和其他硬件技术的发展, 目前虚拟现实人机交互有了长足的进步. 本文着重回顾与分析了虚拟现实的人机交互范式; 三维交互、手势交互、手持交互、语音交互、触觉交互和多通道交互等虚拟现实与增强现实交互技术的主要研究成果和发展趋势; 最后给出需要进一步致力研究解决的若干问题.

关键词 虚拟现实 增强现实 人机交互 用户界面 三维 手势识别 触觉界面

1 引言

虚拟现实是一种可以创建和体验虚拟世界的计算机系统, 它由计算机生成, 通过视、听、触、嗅等作用作用于用户, 为用户产生身临其境的感觉的交互式视景仿真^[1]. 沉浸感 (immersion)、交互性 (interaction) 和构想性 (imagination) 是虚拟现实系统的三个基本特征. 其交互性是在交互设备支持下能以简捷、自然的方式与计算机所生成的“虚拟”世界对象进行交互作用, 通过用户与虚拟环境之间的双向感知建立起一个更为自然、和谐的人机环境来保证的. 因此, 人机交互是虚拟现实为用户提供体验、走向应用的核心环节.

人机交互是实现用户与计算机之间进行信息交换的通路, 用户界面则是这一通路传递和交换信息的桥梁. 理想状态下, 人机交互将不再依赖机器语言, 在没有键盘、鼠标以及触摸屏等中间设备的情况下, 随时随地实现人机自由交流, 从而实现人们的物理世界和虚拟世界的最终融合, 但受技术水平的限制, 目前这种理想还达不到. 因此, 人机交互设计的目标是通过适当的隐喻, 将用户的行为和状态 (输入) 转变成一种计算机能够理解和操作的表示, 并把计算机的行为和状态 (输出) 转换为一种人能够理解和操作的表达通过界面反馈给人. 虚拟现实一方面需要感知用户的肌肉运动、姿势、语言和身体跟踪等多个感官通道的输入信息; 另一方面可以从人类的视觉、听觉、触觉、嗅觉等多个感官通道

引用格式: 张凤军, 戴国忠, 彭晓兰. 虚拟现实的人机交互综述. 中国科学: 信息科学, 2016, 46: 1711–1736, doi: 10.1360/N112016-00252

模拟逼真的真实世界的感觉; 还可以根据用户的视点变化和不同输入, 迅速为用户提供实时的眼、耳、身体所能感受到的感觉信息. 从交互的观点看, 虚拟现实是一种新的人机交互界面.

虚拟现实发展至今已经有近 30 年的历史, 然而由于其通常需要使用价格昂贵的特定输入设备, 同时在实际使用中缺少新的交互范式, 使得虚拟现实在显示设备的历史发展中, 没有进入主流应用. 如今, 随着人机交互领域关键的人、事件、思想和范式的变迁以及心理学指导下的用户模型的不断改进, 虚拟现实正逐渐从科幻走入现实, 其重要性已经不仅局限于游戏娱乐领域, 它将从军事、航天航空、计算机辅助设计、外科手术模拟、科学计算可视化、远程控制以及教育等各个领域改变这个世界. 未来可能有这样一种可穿戴产品, 它会观察你日常的所作所为, 学习你为人处事的方法, 逐渐理解人与人之间的行为模式, 迅速提供给你所需要的帮助, 如同一个共生体一样和你生活在一起.

2 虚拟现实中人机交互的界面范式与交互任务

人机交互是虚拟现实的核心之一. 支持桌面环境的图形用户界面 (GUI) 广泛采用的 WIMP (Window, Icon, Menu, Pointer) 范式, 计算机的功能和操作被抽象成 Window、Icon 和 Menu 图形表示, 用户使用 Pointer 操作这些图形实现人机交互, 这些图形表示是二维的, 用户指点设备也是约束在桌面表面上实现二维运动. 虚拟现实则不同, 用户不是被约束在桌面前工作, 支撑或者约束设备的桌面表面不存在了, 用户看到的也不再是显示器里的二维图像, 用户与虚拟环境的交互的输入、输出是在三维空间中进行的, 这使得二维图形界面组件已经不能满足虚拟现实中的交互需求. 因此, 虚拟现实中的交互设计需要考虑新的输入组件、输出显示和交互隐喻, 其研究的基本问题包含界面范式、交互任务分析、主要交互设备以及如何实现交互任务的交互技术等.

2.1 虚拟现实的界面范式

在桌面环境中占有统治地位的图形用户界面基于 WIMP 范式支撑了当前一代包括网页浏览、文档编辑、电子表格、应用制图和桌面游戏等应用, 它提供了一种工作迅速、准确且稳定的交互方式, 为广大用户所接受, 拥有数以亿计的用户. 然而对于虚拟现实环境的应用而言, 这种交互组织方式是不合适的. 例如一个戴头盔的用户使用键盘是不切实际的, 而且头盔式显示器的分辨率往往非常有限, 对于像对话框这样的文本密集型的交互界面来说也不合适; 另外, 在虚拟现实的应用程序中, 通常需要用户在虚拟三维空间的任意位置、任意方位摆放物体, 对于这样的任务, 二维鼠标则难以胜任.

WIMP 范式支持的图形用户界面的主要问题是交互中输入的带宽过窄, 输入侧仅支持精确点取 (每次一个像素) 和离散按键信号, 形成了一种顺序对话式的交互, 这没有充分利用人在现实生活中习得的操作技能. 为突破这种限制, 一些学者提出 Non-WIMP 和 Post-WIMP 的概念^[2,3], 力图突破图形用户界面限制, 提高人机交互的带宽, 使交互过程更为自然. 从 20 世纪 90 年代起, 以三维用户界面、混合和增强现实用户界面、多通道用户界面等为代表的 Post-WIMP 界面迅速成为了国内外研究的热点^[4].

2006 年 ACM CHI 会议组织了题为 “What is the Next Generation of Human-Computer Interaction?” 的 Workshop, 调查了大量风格各异的交互技术, 到 2008 年会议的发起者 Jacob 提出了基于现实的交互 (reality-based interaction, RBI)^[5], 并试图把它作为 Post-WIMP 界面的统一框架来指导界面设计. Jacob 提出的 RBI 框架主要包括 4 个层次:

- 物理学原理: 人类对物理世界的普遍感知, 如地心引力、摩擦力、速度、物体存在和缩放比例;

- 人体感知与技能: 身体意识, 人类对自己身体的本体感受和对肢体的控制与协调能力;
- 环境感知与技能: 人类对周围环境的感知, 及在环境中操作和导航的能力;
- 社会感知与技能: 人类对环境其他人的感知及与其他人交互的能力.

RBI 框架为界面设计者挖掘从现实行为到人机交互的隐喻提供了指南. 虽然 RBI 不是一种统一的交互范式或模型, 但它给虚拟现实的用户界面范式设计提供了有意义的指导.

从 Post-WIMP 到 RBI, 我们看到并没有形成像 WIMP 那样的固定的、具有普遍意义的范式. 这并不奇怪, 因为在像虚拟现实、增强现实这类交互中, 采用的输入方式不再是单一、固定的, 输出呈现给用户的也不再只是显示其上的二维图像, 用户和虚拟现实系统的双向感知呈现出多通道的特点. 总结这些不同的用于交互的通道方式可以用中医术语“望闻问切”来概括, 用户通过自身的感知通道用“望闻问切”四种方式感知虚拟现实环境的状态, 系统通过各类传感器也是用“望闻问切”四种方式感知用户的输入. 在用户侧这四种方式含义如下:

- 望: 通过人的视觉通道观察虚拟场景, 包括 2D 图像或立体显示图像;
- 闻: 通过听觉 (耳朵) 通道感知虚拟声音;
- 问: 用户可以通过语音与系统对话;
- 切: 用户利用触觉感受器感知系统的力触觉呈现.

在虚拟现实系统侧, 这四种方式含义如下:

- 望: 通过计算机视觉技术以非接触的方式捕获用户的运动、动作或表情, 如手势、姿势、眼动;
- 闻: 利用麦克风阵列捕获用户的语音;
- 问: 系统向用户询问其意图;
- 切: 系统通过用户佩戴或触碰不同类型的传感器感知用户肢体的空间方向或位置、加速度、触点、肌肉伸缩扭转甚至脑神经区电信号的变化, 如操作杆、触控、数据手套、惯性跟踪等.

通过“望闻问切”这四种方式实现是自然人机交互, 这就是虚拟现实的交互范式. 在图形用户界面中, 需要学习鼠标的使用、记忆各种图标所代表的操作意义; 在自然用户界面中, 只需要以最自然的交流方式 (如眼动、语言、表情、手势和肢体) 在交互环境 (移动、桌面、空间环境) 中与机器进行互动. 当然发展中的脑机接口有可能带来更大的变化, 即获取用户的意图来操控虚拟现实场景, 但场景的反馈输出还是需要通过传统的通道为用户所感知.

2.2 虚拟现实的交互任务

虚拟现实强调沉浸感、交互性和构想性, 这决定了它不同于传统的二维人机对话方式的交互. 三维交互能够增强用户对虚拟环境的感知, 提供自然的交互, 是虚拟现实最重要的交互方式. 人们是生存在三维空间中的, 在日常生活中围绕着三维世界习得了许多操纵三维对象和在三维空间运动的技能, 并能很好地理解三维空间关系, 因此采用三维交互方式能够充分发挥人类这些固有的技能, 使虚拟现实人机交互更为自然和谐、为用户所理解. 因此, 面向虚拟环境的三维交互是虚拟现实人机交互研究重点.

三维用户界面是虚拟现实人机交互最主要的界面形式, 三维交互任务在三维用户界面设计中处于核心地位, Bowman 等^[6]从任务、子任务和交互技术三个层面将通用的三维交互任务分为三类, 分别是选择/操纵、导航和系统控制. 选择/操纵的基本任务包含选择、定位和旋转. 选择任务指的是到达目标的距离、相对目标的方位、目标的尺寸, 目标周围的对象密度, 要选择目标的数目, 目标遮挡; 定位任务指的是相对初始位置的距离和相对方位、相对目标位置的距离和方位、平移距离, 要求的定位精度; 旋转任务指的是相对目标的距离、初始方位、终止方位、旋转量和要求的旋转精度. 导航任务主

要分为漫游和路径查找任务。其中, 漫游是导航的动力构件, 是用户控制视点位置和方向的一种低层活动, 是最普通和通用的任务, 例如脚的移动、旋转方向盘、松开制动踏板等; 路径查找是导航的认知构件, 是与用户运动有关的高层思考、规划和决策。它包含空间理解和规划任务, 如确定在环境中的当前位置、确定从当前位置到目标位置的路径、建立环境的意念地图等。系统控制任务指的是改变系统交互模式和系统状态的那些系统命令, 例如请求系统执行一个特定功能、更改交互模式和更改系统状态等。

虚拟现实交互的效果很大程度上依赖于所应用的交互任务, 规范任务方法把操作任务简化到它们最本质的性质, 这种规范任务的分类方法考虑了任务和实现任务的方法, 对虚拟现实人机交互的设计具有重要指导意义。当然, 对于一些复杂的特殊任务, 一般的操作任务就没有意义了, 就需要针对特定的应用进行更详细的交互任务设计。例如在虚拟现实医学训练应用中, 医学探针相对于内部器官的虚拟三维模型的定位, 飞行仿真器中虚拟飞机控制手柄的移动等。

3 虚拟现实的交互设备

用户如何输入信息以及系统如何显示输出信号, 对于虚拟现实人机交互设计来说非常重要。传统计算机视觉显示器能显示二维图像, 但虚拟现实需要使用更高级的显示器提供立体观察, 另外, 很多虚拟现实还用非视觉显示器, 需要向其他感觉器官呈现信息。传统计算机使用文本输入 (键盘)、二自由度的鼠标。虚拟现实往往需要为用户提供同时操作的更多自由度, 因为用户需要给出位置、方向、手势等三维空间中的行为来完成复杂的交互任务。

3.1 输出显示设备

虚拟现实系统通过输出硬件向用户的一种或几种感官器官提供信息, 其中的大部分设备主要用来刺激人的视觉、听觉和触觉, 在极少数情况下, 也可以将信息传给嗅觉或味觉。在虚拟现实人机交互中, 设计、开发和使用不同的交互技术时, 必须认真考虑选择显示设备, 因为对某一种特定的显示器来说, 某些交互技术可能比其他的交互技术更适用。虚拟现实交互输出设备主要包含视觉显示设备、听觉输出设备和力/触觉输出设备。

- 视觉显示器。视觉显示设备的属性包含观察区域和可视区域、空间分辨率、显示屏形状、光线传播、刷新率和功效学等。虚拟现实交互视觉显示设备的设计根据一般包含单眼的静态线索、眼球的运动线索、运动视差、双眼视觉不一致性和立体成像等。虚拟现实的视觉显示设备种类一般有终端显示器、环屏显示器、工作台显示器、半球形显示器、头盔显示器 (HMD) 和悬臂式显示器 (AMD) 等。

- 声音显示器。声音显示器的一个重要作用是通过产生和显示空间三维声音, 使参与者可以利用他的听觉定位 (判断声源的位置和方向) 能力的优势。三维声音定位线索包含双耳线索 (两耳时差、声强差)、声强、声谱 (高频率有用)、动态线索 (很弱)、回声 (很弱)、头部相关传递功能 (描述了声波与人外耳的交互, HRTF (head-related transfer function) 会根据声源的位置对声波进行更改, 这种更改后的声波向听者提供了定位线索)。声音输出的主要类型有简单音素、图标式音素和自然声音 (录制或语音合成)。声音在虚拟现实人机交互中往往作为第二种反馈形式输出, 或作为其他感应通道 (如触觉) 的替代效果, 一般使用耳机和外部扬声器为用户输出立体声、环绕立体声和 3D 音频等。

- 力/触觉输出显示器。虚拟现实中心力/触觉输出设备主要根据触摸线索、肌肉运动知觉线索和运动神经子系统而设计。触觉输出设备可以分为以地面为参考系、以身体为参考系和直接刺激神经产生

三种主要类型. 常用的力/触觉输出设备有数据手套、力反馈鼠标、力反馈操纵杆、力反馈方向盘和力反馈手臂等. 这些输出设备能有效增强交互的真实性, 但是模拟真实世界触觉感受非常困难、有时候还可能使用户产生恐惧感, 而且往往需要制作特定的机械设备.

3.2 输入设备

虚拟现实人机交互设计中的一个同样重要的部分是选取一组合适的输入设备, 实现用户和应用任务的通信. 就像输出设备一样, 当开发虚拟现实的人机交互时, 也有很多不同类型的输入设备可供选择, 某些设备对特定的任务比其他设备更合适. 虚拟现实三维交互输入设备可以分为离散输入设备、连续输入设备和直接人体输入设备.

- 离散输入设备. 这类设备根据用户动作一次产生一个事件, 生成一个简单的数值 (布尔值或一个集合中的元素), 常用于改变模式或者开始某个动作, 用于离散型的命令界面, 例如键盘、pinch glove^[7] 等.

- 连续输入设备. 这类设备利用包括力、热、光、电和声等不同传感器跟踪用户的连续动作, 跟踪点的数据包括位置、方向或者加速度等信息. 典型的设备如三维鼠标、电磁跟踪器、力反馈手套、数据手套、内置传感器的手柄 Wii Remote¹⁾、摄像机、深度相机 Kinect²⁾、Leap Motion³⁾、3D 摄像头 RealSense^[8] 等.

- 语音和生理信号感知设备, 收集语音或者其他生理信号. 这类设备主要包含语音输入、生物电输入和脑电波输入等, 代表性的语音输入产品有智能音箱 Google Home⁴⁾和语音助手 Amazon Echo⁵⁾等. Google Home 是配有内置扬声器的语音激活设备, 可以在智能家居中通过语音控制家庭设备. Amazon Echo 是一款实时联网的圆筒状设备, 可以在生活中充当助手, 用户用普通语音提出问题, Amazon Echo 也会用语音作答, 还能列出购物清单. 生物电输入设备主要通过生物电传感器读取人的肌肉或神经信号变化而交互, 主要用于智能可穿戴设备中.

- 脑电波输入设备通过脑电图 EEG 信号等监视脑电波的活动, 代表性产品有 NeuroSky MindWave 意念耳机⁶⁾、能操控无人机的 EmotivInsight⁷⁾、能锻炼专注力的意念头箍 BrainLink⁸⁾和能模仿表情的 EmotivEpoC⁹⁾等. 这些直接人体输入设备可以作为其他输入通道的补充, 是理想的虚拟现实交互必不可少的部分.

在虚拟现实人机交互设计中, 输出/输入设备的选择需要考虑具体交互技术的需求、输入设备和输出设备之间的相互约束和多通道交互之间的互补等. 在实践中, 费用往往是最大的因素, 还需要考虑交互技术的设计是否受给定设备的限制、是否需要为实现交互技术购买先进的设备, 以及是否需要为交互技术制作新的交互设备等.

1) Wii Remote. https://en.wikipedia.org/wiki/Wii_Remote.

2) Kinect. <https://dev.windows.com/en-us/kinect>.

3) LeapMotion. <https://www.leapmotion.com>.

4) Google Home. <https://home.google.com>.

5) Amazon Echo. https://en.wikipedia.org/wiki/Amazon_Echo.

6) NeuroSky Mindwave. <http://www.neurosky.com/mindset/mindset.html>.

7) <http://emotiv.com/product/emotiv-insight-5-channel-mobile-eeeg>.

8) <https://www.brainlink.com>.

9) <http://emotiv.com>.

4 虚拟现实中的人机交互技术

人机交互技术指的是通过设备和界面完成给定交互任务的具体方法. 为了给用户提供一种临境和沉浸式的体验, 虚拟现实需要同时具有高真实感的环境表达特征和高效的用户和环境信息交换特征, 它涉及到的人机交互技术不仅包含以三维交互、姿势交互和手持移动设备交互, 也包含语音交互技术、力/触觉交互技术和多通道交互技术等.

4.1 三维交互技术

三维交互是虚拟现实最重要的交互方式, 相对于二维交互, 虚拟现实三维交互提供了更多的操作自由度, 交互任务更为复杂, 交互界面设计有着更大的设计空间, 因而需要全新的交互隐喻和技术. 交互隐喻是通过对现实世界存在的一些机制进行比拟或抽象, 并借用到交互过程中. 在三维交互中, 交互隐喻把输入设备的空间方向/位置信息或离散的按钮状态映射为虚拟空间的操作, 完成特定的交互任务, 不同的交互隐喻提供了不同的交互方法. 根据交互系统对输入的不同隐喻, 可以把已有的三维交互方式分为直接映射式和间接映射式两大类.

直接映射式输入是指直接映射设备输入的位置/空间信息为虚拟空间中手或设备的操作动作. 主要的隐喻方法包括:

- 光线投射 (ray-casting) ^[9,10] 隐喻. 光线投射技术使用虚拟光线投射到要抓取的虚拟对象上, 而光线的发射方向由用户的手控制. Liang 等 ^[10] 为了解决这种技术对远程小对象的选取, 提出了用锥光柱代替圆柱光柱的方法.

- 虚拟手隐喻. 虚拟手技术是在虚拟环境中构造用户手的虚拟化身, 虚拟手的位置/方向由用户手控制 (通过跟踪器). 但由于人手臂工作范围的限制, 难以对远距离对象进行操作, 为此, Poupyrev 等 ^[11] 基于非线性映射原理提出 GO-GO 技术, 来扩大手在虚拟空间的可达范围. 其他一些研究者也基于非线性映射思想提出了不同映射方式 ^[12].

此外, Wloka 等 ^[13] 提出的在视觉上复制六自由度 DOF 输入设备到虚拟环境中的 Virtual Tricoder 隐喻, 以及 Bowman 等 ^[12] 提出的组合虚拟手延伸和光线投射的交互方式 HOMER (hand-centered object manipulation extending ray-casting) 也属于直接映射式交互方法. 关于 3D 交互点取选择的绩效研究也在逐步展开 ^[14].

间接映射式交互是指把设备输入的信息映射为手势, 通过手势控制场景空间的比例, 进而在新的比例空间完成交互任务. 在这种映射方式中, 通常需要采用双手交互的方式, 利用双手操作的非对称分工, 即利用人的非利手作为参考坐标框架, 有利手相对非利手进行精细的操作. 主要的隐喻方法包括:

- WIM (worlds in miniature). Stoakley 等 ^[15] 提出的 WIM 交互隐喻使用双手操作, 两只手分别持有带跟踪器的面板和按钮道具, 面板道具在虚拟空间被隐喻为整个虚拟场景的缩微拷贝, 缩微拷贝中的物体与场景中的物体是相关的, 在缩微空间实现对场景对象的操作.

- 图像平面. Pierce 等 ^[16] 提出了基于图像平面交互方法, 主要思想是构造三维对象的一个二维图像投影平面, 基于该图像平面采用与二维鼠标类似的交互.

- Voodoo Doll. Pierce 等 ^[17] 的另一项工作是基于木偶隐喻的交互方法, 即在虚拟手上动态地创建一个被操作对象的拷贝 (称为 Doll), 并放缩到一个适合操作的比例. 用户针对 Doll 操作, 场景中的对象随 Doll 改变而改变, 当操作完成后 Doll 自动消失. Voodoo Doll 采用双手交互, 左手作为右手工作的参考框架 (假定用户的右手是习惯手). Voodoo Doll 交互需要双手手套和跟踪器.

- Scaled-world grab. Mine 等^[18]建议用本体感觉来补偿沉浸式虚拟环境中缺乏的触觉反馈, 并提出 Scaled-world grab 交互技术. 其思想是用户抓取虚拟对象时, 场景自动地相对于用户的头部放缩比例, 当用户释放对象, 则比例自动回退.

- Tablet-Pen 模式. 用户非利手持有 Tablet, 有利手持 Pen, Tablet 和 Pen 附有空间位置跟踪器. 这样 Tablet 不但作为 Pen 运动的参考基准, 而且还为 Pen 提供了二维平面物理约束, 因此可以在三维环境中嵌入二维桌面的交互机制. 这类模式的系统主要有 Szalavari 等^[19]的 PIP (personal interaction panel) 系统, Schmalsteig 等^[20]介绍的透明 Pad-Pen 系统, Billinghamurst 等^[21]介绍的 3D Palette 系统和 Poupyrev 等^[22]介绍的 Virtual Notepad 等. Virtual Notepad 系统还提供了在沉浸式虚拟环境中写、画和注释文档 (采用笔迹识别在虚拟环境中输入文本) 的功能. 近年来, 随着手持智能设备的普及, 使用 Pad 和手指触控的方式逐步在取代传统的 Tablet-Pen 模式, 手持设备具有显示能力, 比只用触控板增加了更靠近用户的显示反馈, 并且手指可直接操作触摸屏上的虚拟对象, 也增强了操作远距离屏幕时的手眼协调性, 这方面代表性工作可参见文献 [23~26].

- 触控手套模式. Mapes 等^[27]开发的仿真和训练 PolyShop 系统使用了一种称为 ChordGloves 或 PinchGloves 的手套, 这种手套在手掌和每个手指端部装有电子接触器, 用于侦测手指间的接触. 由于这种手套没有关节传感, 成本低. Mapes 认为侦测手指间的接触比通过手指关节或者手的运动来识别手势更为可靠, 更为重要的是手指接触的方式比手势语言更易于学习, 尤其对于频繁使用的命令更为有效. 基于此, PolyShop 系统使用指、掌间的接触作为离散切换命令, 利用双手在虚拟世界中实现比例变换、旋转或平移对象等操作, 该系统包含一些功能强大的交互技术, 如手持 Palettes、双手飞行、双手动态放缩比例世界和 ModelTime 行为等.

此外, 基于三维器件 (widget)^[28], 如三维悬浮菜单、三维图标等进行交互的方式, 并不是一种直接操作技术, 因而多用于系统控制类交互任务.

前面简要地介绍了一些主要的交互隐喻, 这些三维交互技术已经在一些领域如视频游戏、互动体验等得到应用, 对三维用户界面也开始有了一些评估工作^[29]. 总体上看, 目前, 三维交互方法主要存在这样一些问题:

- 三维操作的空间范围问题. 直接映射式中的手臂延伸技术主要是解决虚拟手操作范围的问题, 如 GO-GO 技术采用非线性映射的策略扩展手臂的工作范围, 但对远程、小对象的抓取并不方便, 因为虚拟手很难定位到要抓取的物体上.

- 此外, 非线性映射的方式有时会使虚拟手的延伸长度难以控制. 间接映射式中的 WIM 虽然给出了场景空间的缩微全景, 但在缩微全景中对场景中的小对象操作存在严重问题. 因为在缩微全景中, 这些对象变得更小. 另外, 在缩微全景中微小的移动有可能导致场景中发生较大的变化, 这对漫游/导航或移动场景对象是不利的.

- 三维输入设备提供了更多的自由度, 在虚拟交互中如何有效地控制这些自由度, 目前的交互方法并没有很好地解决. 以光线投射法为例, 该方法采取忽略一些自由度进行交互的策略, 虽然方便了对虚拟对象的选取, 但难以实现对对象进行旋转或沿射线方向移动的操作. 锥光柱方法会选取到过多的对象, 如何从这些备选对象中选择所需要的对象, 对单手交互来说是个问题.

- 间接映射式隐喻为虚拟对象 (或场景) 的操作提供了不同比例的空间, 但如何攫取这些比例空间以及比例空间的呈现形式还存在问题. 如 Scaled-world grab 虽然能够自动设置空间放缩的比例, 但由于只工作在一个比例空间, 所以只能操作比例空间中的可见对象, 另外, 场景的放大也使操作范围相应地缩小了. 图像平面隐喻虽然可工作在多个比例空间, 但比例空间的攫取 (也为 Voodoo Doll 所采用) 需要“目光”的配合, 而且必须使用数据手套才能实现, 因而限制了它的适用范围, 其操作精度也是一

个问题.

- 对虚拟环境下复杂、交叉交互任务缺乏理解, 而且没有合理的交互方法以及实验研究.
- 交互过程中, 交互模式的转换以及虚拟对象与设备的离合机制, 还没有自然和有效的解决方案.

三维交互技术已成为虚拟现实技术在各种领域进一步应用的瓶颈问题, 亟待解决. 我们认为解决现有三维交互中存在的问题要从以下两个方面着手:

- 融合现有交互隐喻, 开发新的交互方法;
- 研究能充分发挥人类固有技能的输入范型及其配套的交互方法, 如采用双手交互.

4.2 手势与姿势交互

在虚拟现实, 用户身体或肢体的运动可以作为一种重要的输入通道. 通过跟踪器或计算机视觉的方法跟踪人体相关部位 (如头、手、臂或腿), 获得人在物理世界的运动姿态信息作为虚拟现实系统的输入, 通过识别算法解释为手势 (gesture) 或姿势 (posture, 以下统称为手势). 这是目前最为主要的虚拟现实输入方式之一.

关于手势, Cerney 等^[30] 给出了一个广义的定义: 手势是指人的手、臂、面部或身体的物理运动, 用以传达信息或意图的目的. 而手势识别不仅包含人的运动跟踪, 也包括把运动解释为具有语义的命令, 一些常用的人体手势包括手部姿势、臂部姿势、头部姿势和面部表情等. 姿势分为动态的和静态的两种.

Cerney 认为手势可以根据它的功能、语言学以及在通讯中的角色进行分类. 从功能的角度看, 手势分为 3 类^[31]: 记号姿势、表示活动的手势和认知手势. 记号姿势通过增强或推动交流来传递有意义的信息; 表示活动的手势包括对物理环境的操作, 通常涉及到工作的概念; 认知手势包括通过触觉体验或触觉探查来探索环境的过程. 记号姿势可以进一步按它在通讯中的角色分类为图标的、隐喻的、指代的和节奏性姿势^[32]. 图标姿势是活动、对象或事件的表示; 隐喻姿势表示通用的隐喻, 而不是直接指向对象或事件; 指代手势是用指点的方式指向人、对象或方向; 节奏性姿势是一些用手或头完成小幅度的用力的动作. 这些姿势分类对虚拟现实的交互设计具有极大的指导作用. 一些已有的虚拟现实系统也使用了这些姿势, 如 Brooks 用活动性姿势进行科学可视化的应用^[33] 等. 使用记号姿势的如指代式导航^[34]、记号语言解释^[35] 等.

手势是人类众多身体姿势中最常用的一种, 在虚拟现实系统中最早使用手势交互主要使用符号手势 (预定义、含义明确的手势) 来触发系统命令的工作. Bolt 等^[36] 在 1992 年把他们以前的工作扩展到 3D 对象的操作. Cavazza 等^[37] 分析了虚拟现实中的指点手势, 他们定义了扩展指点用于处理一个或多个对象的选择操作, 从运动跟踪中捕获到的人体运动的原始数据.

主要的手势、姿势输入技术按照使用的输入设备可以划分为:

- 基于手套的手势识别. 从手套设备中获得的原始数据可以采用如隐马尔可夫模型、神经网络等识别算法进行分析, Granit 等^[38] 采用基于滑动窗口的方法识别手套手势, 达到较高的精度. 手可以被用作一个按钮、计算器、定位器和拾取设备. 掐捏手套可以用来做有限的手形, 数据手套也可以通过使用关节角测量提供手形和手势. 尽管由于手势识别、数据手套校准的困难以及类似的缺点, 致使基于手势的交互在 3D 交互界不怎么受欢迎, 但它仍然是一种仅需要用户的手进行输入的强有力方法.

- 基于视频的识别. 手或手指姿势以及肢体或头部的视频图像, 可以通过计算机视觉算法来识别出特定的姿势. 微软发布的 Kinect^[39] 深度相机为基于视频的手势交互带来了革命性的进展. Kinect 主要利用一个普通的 RGB 摄像头和一个检测深度的红外摄像头来获取人的 RGB 图像和深度图像, 基于深度图像数据跟踪人体的骨骼、手、头等各个部位的精度、速度都有大幅的提升, 为手势交互走

向实用提供了基础. 另一个基于视觉原理的跟踪设备是 Leap Motion^[40], 它运用双目视觉的手势识别技术, 通过双目摄像头采集操作者手的左右视觉图像, 通过立体视觉算法生成深度图像, 再使用跟踪算法对手运动进行跟踪. Leap Motion 能实现对手和手指运动的精细跟踪. 此外, 由于 Leap Motion 尺寸小, 因此可附着在头盔上用于跟踪手部动作. Alejandro 等^[41] 对多通道中使用视频技术捕捉肢体姿势、手势、面部表情以及视线跟踪作了全面的综述, 对 3D 手势较新的综述参见文献 [42, 43], 更早一些的综述见文献 [44, 45].

- 基于表面的识别. 显示屏、触摸屏或者其他的表面可以用来传感手指的接触状态和轨迹, 进而识别为手势. 这里手势由笔划构成.

- 基于惯性传感单元的手势识别技术. 通过手上佩戴的惯性传感器单元 (IMU) 获得手或关节运动的角度、加速度和方位势等信息, 进一步通过模式识别的方法获取用户的手势. 典型的如任天堂的 Wii Remote、SONY 的 PS4 手柄以及大多数智能手机. 此外, 在智能穿戴领域, 通常把惯性传感单元附着在人体的关节、头部等进而解算出人体的运动, 代表性的惯性穿戴设备如诺亦腾公司的产品¹⁰⁾以及 Shahidul 等的工作^[46].

- 雷达技术. 通过持续发射和接受手部反射的电磁信号, 测量精细、复杂的动作变化, 而后转码分析、识别. Google Project Soli¹¹⁾ 是这一技术的典型代表, 他们设想四类手指的隔空动作指令: 转盘手势、刻度手势、按手势和位移手势. 用户可以通过搓一下手指来调手表的时间, 捏一捏空气来浏览地图, 不需要实体按键就可以隔空选音乐、调音量等.

- 肌肉电技术. 典型产品是加拿大创业公司 Thalmic Labs 推出的手势臂环 MYO¹²⁾, 通过感知手臂上肌肉运动的变形信号, 利用模式识别的方法可识别出近 20 种手势.

- 电磁波技术. 最具代表性的电磁波方案则是一项由卡耐基梅隆大学团队所研发的 SkinTrack^[47] 技术, 该系统由一个能连续发射高频信号的环和内置多个感测电极的腕带组成, 通过测量与分析高频信号在皮肤上传播的相位差, 实现在皮肤上连续跟踪手指的触摸位置.

手势交互是一种重要的自然交互方式, 也是一种理想的虚拟现实交互技术. 以体感交互为代表, 手势交互目前可以用于一些不特别复杂的交互场景, 如浏览网页、阅读文章、翻看照片, 还有播放音乐、三维造型, 或者像微软 Kinect 游戏那样控制虚拟化身的动作. 但手势交互的广泛使用仍面临一些重要问题需要解决:

- 金手指问题. 金手指 (midas touch) 问题是指手势识别系统不能有效判别人的连续运动中, 哪些动作是有意图的交互或下意识的动作, 或者不能明确判别一个手势的发起或结束. 这对使用手势交互的系统造成极大的困扰, 解决这一问题通常有 2 种手段, 一是通过延时, 即先把手放在手势起点一小段时间, 告诉系统用户即将开始手势; 二是在做手势之前向系统做出一个特定的姿势, 或者激发一个开关, 或者用语音告知系统手势动作即将开始. 无论哪种方法, 都需要一个明确的“离合”切换, 造成连续交互的中断.

- 手势交互需要识别, 必须在一个相对完整的运动轨迹之后系统才能判别出手势的语义, 这将造成一定的延时.

- 使用手势交互需要根据应用场景的特点来选取手势, 用户必须需要经过一个或长或短的学习和记忆的过程. 此外, 当一个系统使用手势过多或者不同的应用使用不同的手势, 将给用户带来记忆和

10) 诺亦腾官网. <http://www.noitom.com.cn>.

11) Google Project Soli. <https://www.google.com/atap/project-soli>.

12) Nuwer R. Armband adds a twitch to gesture control. New Scientist, 2013, 217 (2906): 21. <https://www.thalmic.com>.

认知的负担.

- 限于肢体姿态和动作获取技术, 目前只有记号手势取得较好的应用, 因为这类手势从识别的角度看已经达到了很高的准确程度, 而且它所代表的语义也是相对静态的或者是全局意义上的状态转换, 比如用 3D 手势进行 3D 菜单的选择^[48]. 而活动性手势只是应用在一些化身驱动性的应用中, 一些对虚拟场景对象的精细操作目前还不能实现, 这需要进一步提升动作捕获和识别的精度.

4.3 手持移动设备交互

随着电子信息技术的进步, 以智能手机为代表的手持移动设备的计算能力、图形显示能力得到持续的增强, 同时也集成了相机、GPS、加速度计、陀螺仪、电子罗盘等不同种类的传感器, 提升了设备对环境 and 用户行为的感知能力^[49]. 人们除了关注用户与这些智能移动设备本身的交互之外, 也已经注意到了智能手持移动设备的三维交互能力, 把它用于虚拟/增强现实^[50]和普适计算领域^[51]. 从 1993 年 Fitzmaurice 发表 Chameleon 系统^[52]以来, 在虚拟/增强现实和普适计算领域涌现了 20 余种不同的基于手持设备交互的原型系统^[53], 它们各有特色, 但一个共同的理念就是摆脱传统的桌面交互, 实现用户与环境的交互, 其核心就是基于手持移动设备的三维交互的设计问题. 基于手持移动设备的交互隐喻主要有两类: 信息透镜隐喻和直接指点控制器隐喻.

(1) 信息透镜隐喻

信息透镜隐喻是目前使用最多的一种方法, 尤其在增强现实领域, 其核心思想是实时计算手持设备相对于被观察场景 (已标住数字信息) 的相对位置 (通过位置传感器或者视觉的方法), 把已标注的数字信息实时叠加到由相机捕捉到的图像上, 其后的操作将基于手持设备上的按钮、笔或触摸在 2D 下操作. 信息透镜隐喻发端于 Fitzmaurice 开发 Chameleon 系统, 这个系统提出了空间感知显示器的概念^[54], 即用带有位置跟踪器的手持设备去感知叠加在空间的数字信息.

随着手持移动设备逐渐集成了摄像头并且自身的计算能力和图形能力的逐步增强, 信息透镜隐喻也不断发生变化, 其中基于内置相机的光学跟踪得到最大的关注. 1995 年, 索尼公司的 Rekimoto 的 NaviCam 系统^[55], 使用安装在手持设备上的内置相机, 通过光学跟踪的方法检测颜色标签, 在手持显示设备上以透视的风格直接显示上下文敏感信息. 值得注意的是, NaviCam 系统并不感知相机的相对位置. Wagner 等^[56]的室内 AR 导航系统, 基于标志卡光学定位跟踪技术计算 PDA 相对标志卡的位置, 实现室内三维增强现实导航, 是第一个自包含的手持移动增强现实系统. Rohs 提出了一套视觉编码^[57,58], 以实现基于增强现实的 Widgets 系统, 该系统使用手机内置相机作为透视工具来恢复物理对象的相关信息和功能. 把基于标志卡片定位的方法加以扩充, 通过计算相机相对于大幅面纸制实物 (如地图、海报或图纸) 的相对位置, 可实现具有更好空间上下文的魔镜技术 (Magic Lens). Peephole 技术^[59,60]是另一类信息透镜隐喻. Peephole 在用户周围构造出虚拟工作空间, 通过机械式传感器感知手持设备在这个虚拟工作空间的位置, 进而采用基于笔的双手交互技术与虚拟工作空间的数字内容交互. 由于在手持显示器之外没有与数字信息相结合的物理上下文的视觉反馈给用户, 这个技术只能用于查看和操作数字信息.

一些研究者也对信息透镜隐喻作了各种实验, 来对比其性能和可用性等. Rohs 等^[61]设计了使用带有内置相机的手机结合实物地图进行导航的测试任务, 对比了使用手机操纵杆、没有实物上下文的 Peephole 和魔镜隐喻在目标获取时的操作绩效, 并对 Fitts 法则在魔镜隐喻下进行了修正. 实验结果表明, 相比于操纵杆, Peephole 和魔镜隐喻有着更好的性能. Morrison 等^[62]把魔镜技术用于地图导航, 提出了 MapLens 技术, 并通过一个基于位置的室外游戏, 探讨了用户在协同讨论完成游戏任务时使用 MapLens 技术的优势.

信息透镜隐喻主要用于浏览虚实叠加信息, 如在一些位置导航^[63,64]、协同游戏^[65,66] (利用手持设备的通讯能力和键盘/触摸输入)、广告以及城市管道检查等领域已有一些应用. Schmalstieg 等^[67]开发了一个手持 AR 平台 Studierstube ES, 主要面向 Windows CE 移动设备. 信息透镜隐喻的优势是, 为用户浏览查看叠加在真实空间或物体上的数字信息提供了方便的手段, 但由于使用手持移动设备时, 总有一只手被占用, 同时交互的背景来自于相机抓取的场景, 因此, 对进一步操作虚拟信息, 如批注、选择、编辑或移动等任务还面临巨大的挑战.

(2) 直接指点隐喻

直接指点隐喻是捕捉手持移动设备相对于空间显示器的位置, 进而通过手持设备来操纵空间显示器. 近年来随着显示技术的进步, 面向公众的大尺寸显示设备被用于各种场合, 但与其交互存在的问题: 传统使用鼠标键盘的模式要求用户必须在固定位置, 这一问题限制了大屏幕作为一种公共交互空间的利用率. 而利用手持移动设备机动性强的特点, 用户就可以不必在固定位置操纵远距离屏幕, 方便多用户、移动使用, 将大大提高大屏幕的利用效率.

使用手持设备直接操作大屏幕的一种直接的方法是使用类似远程桌面的方式, 把远程屏幕的内容直接绘制到手持显示器上, 并使用手持设备的笔或触摸进行控制, 但受限于手持移动设备屏幕尺寸和比例限制, 这种方法并不能精确地进行交互操作. 为了更精确地操作大屏幕, 研究者多采用基于视觉跟踪的方法来定位手持设备与被操作屏幕之间的关系, 把设备的运动直接映射为光标的运动, 从而实现直接操作 (其他操作通过手持设备上的按键或操纵杆来实现). Miyaoku 等提出了 C-Blink 技术^[68], 他通过安装在大屏幕上的摄像头采集手机屏幕的色彩 (手机屏幕闪现不同的颜色作为初始化) 来计算手机相对于大屏幕的 2D 位置, 进而实现 Click, Grab 和 Pitch 等交互操作. C-Blink 技术利用了手机屏幕不同的颜色作为定位的基准, 但存在一个问题是, 需要外置摄像头并且手机屏幕需要时刻面向屏幕, 带来交互上的不便. Hachet 等^[69]提出使用手持 PDA 上的相机和标志卡实现大屏幕虚拟现实交互. 他们采用双手交互的方式把卡片相对相机运动映射为交互动作, 实现了一个三自由度交互界面, 来控制大屏幕虚拟现实漫游. Hachet 的方法需要额外的实物标志板, 并且双手都被占用, 这限制了一些交互操作. Ballagas 和 Rohs 等^[70,71]采用手机摄像头捕捉在大屏幕上闪现的视觉编码标志, 计算手持设备相对大屏幕的坐标, 进而实现了包括 Point & Shoot, Sweep 等一系列的交互技术. 这个技术由于视觉编码标志在屏幕上的时隐时现, 会导致屏幕的混乱引起交互的不便. 此外, Pears 等^[72]也提出了类似的方法实现手机与大屏幕的注册. Jiang 等^[73]组合手持相机和鼠标实现了针对大屏幕的空间交互方法, 通过相机图像与大屏幕之间的仿射变换, 计算光标在大屏幕的位置来实现远距离的大屏幕操作. Wang 等^[74]的 TinyMotion 系统使用手机摄像头, 采用简化的运动差分算法实现基于手机的二自由度控制, 并实现了滚卷、缩放、菜单选择、光标移动以及手势/手写等输入任务. TinyMotion 技术也可以用来操作大屏幕, 但是需要在跟踪精度上进行改进. 完全使用内置摄像头来检测手机本身的运动是一种完美的方法, 其他一些学者则是采用各种简化的光流法实现手机的姿态估算.

与将手持设备作为直接指点设备不同, Boring 等^[75]提出了直接在手机屏幕上点取手机摄像头拍摄的视频图像来实现远程交互的 Touch Projector 系统. 该系统不同于把手持设备直接映射为屏幕光标运动的方法, 系统虽然也使用手机内置相机与远程屏幕定位, 但其交互是通过直接在手机屏幕上操作捕获的屏幕视频控制远程屏幕的. 基于这一思想, 他们开发了包括冻结、缩放等任务在内的跨屏选择/拖动交互技术. Yang 等^[76]把具有触摸功能的手机嵌入到桌面鼠标上, 开发了 LensMouse. 在 LensMouse 提供的交互技术中, 把手机触摸屏作为桌面显示器的辅助窗口, 用于显示操作主屏幕的工具条/弹出式菜单/对话框等图形界面元素, 使用手机屏幕的触摸功能与这些界面控制元素交互, 这样节省了主显示屏幕的空间, 并进行了一系列的实验, 对设备和交互技术的有效性进行了评估.

此外, 随着头盔显示技术的进步, 人们也把手持智能设备作为用户穿戴头盔时的输入设备, 如 Gugenheimer 等^[77] 把手机放置在头盔前, 利用其触控能力操作头盔内的虚拟场景, 这种模式具有较好的手眼协调性; Henrikson 等^[78] 利用电影叙事隐喻设计了多设备故事板, 解决了佩戴头盔时不能看到手持 pad 问题, 并可在多用户间共享。

总之, 手持移动设备作为一类智能型多通道输入输出设备, 为解决包括增强现实、虚拟现实、多媒体展示以及普适计算的交互问题提供了一条有效的途径, 已经引起越来越多的国内外学者的关注。在现有的硬件技术水平条件下, 使用基于手持移动设备进行三维交互亟待解决以下几个方面的问题:

- 手持移动设备具有受限的显示能力, 这是区别传统三维交互设备的特点之一, 如何充分发挥这一特点, 建立面向多显示输出的交互模型是设计手持移动三维用户界面的基础问题;
- 采用怎样的融合策略, 在受限的计算能力范围内有效地感知设备相对于环境或用户的相对运动, 以及采用多通道输出指引用户的注意, 对提升手持移动三维交互的鲁棒性和可用性是至关重要的;
- 开发新的交互隐喻, 构建更为高可用的交互技术, 是一个具有挑战意义的研究问题, 对扩大手持移动三维交互适用范围具有重要意义。

4.4 语音交互技术

语音输入是一种很自然的输入方式, 它能将不同种类的输入技术 (即多通道交互) 结合起来形成一种更有连贯性和自然性的界面。如果功能适当, 尤其是用户的两只手都被占用的时候, 语音输入将成为虚拟现实用户界面中很有价值的工具。语音有许多理想的特点: 它解放了用户的手; 采用一个未被利用的输入通道; 允许高效、精确地输入大量文本; 是完全自然和熟悉的方式。在虚拟现实用户界面中, 语音输入尤其适合非图形的命令交互和系统控制, 即用户通过发布语音命令来请求系统执行特定的功能、更改交互模式或者系统状态。此外, 语音输入也为虚拟现实中的符号输入提供了一种完整的手段^[36]。这主要有三种方式: 单字符语音识别^[79]、非识别语音输入和完整单词语音识别。虚拟注解 (virtual annotation) 系统^[80] 是使用非识别语音输入的一个虚拟现实用户界面范例。

由于语音界面对于用户来说是“不可见的”, 用户通常不需要对语音界面可执行的功能有一个总的视图, 因此为了捕捉用户的真实意图, 就需要通过语义和句法过滤实现纠错 (使用语义或者句法的预测方法来限制可能的解释), 或者是使用形式化的对话模式 (问答机制)。由于语音技术初始化、选择和发布命令都在一次完成, 有时可以用其他的输入流 (如按键) 或者一个特殊的声音命令初始化语音系统。这消除了一个语音开始的歧义, 称为“即按即说” (push-to-talk) 系统^[36]。

语音交互主要有两种方式: 语音识别和语音对话系统^[81]。在使用语音交互开发虚拟现实人机交互时, 首先要考虑语音界面要执行的交互任务, 交互任务将决定语音识别引擎的词汇量大小, 即任务和它所运行的范围越复杂越需要更多的词汇量。对于仅有少量功能的应用, 采用简单的语音识别系统可能就足够了, 高度复杂的应用则需要包含语音对话系统来保证理解语音输入的全部功能。

语音交互的关键是语音识别引擎, 一些现有的语音识别软件包括微软 Speech API, IBM ViaVoice, Nuance 和国内的科大讯飞等, 它们都达到了很好的性能。如今, 随着语音识别技术的逐步开放和开源, 语音技术门槛逐渐降低。主要的语音开源交互平台有 CMU-Sphinx, HTK-Cambridge, Julius 和 RWTH ASR 等。近年来, Google 眼镜、穿戴式设备、智能家居和车载设备的兴起, 将语音识别技术推到应用的前台。自苹果 iPhone 4S 内置 Siri 以来, 几乎所有的手机都开始内置语音助手类的应用。目前具有代表性的语音助手产品有智能音箱 Google Home 和语音助手 Amazon Echo, 这两者在前文都已经有所介绍。

4.5 力/触觉交互技术

相对于传统的视觉交互和听觉交互, 触觉交互能使用户产生更真实的沉浸感, 在交互过程中有着不可替代的作用. 传统人机界面中的力/触觉交互把力/触觉交互看作为界面交互中的一种特殊输入输出方式, 作为输入设备时, 它们用来捕捉用户动作, 作为输出设备时它们为用户提供触觉体验^[82]. 虚拟现实中的力/触觉交互则以自然交互为研究重点, 是未来人机交互的重要发展方向.

触觉反馈相关技术主要包含两个方面. 基于振动的触觉反馈技术是以驱动器所产生的机械振动作为触觉刺激源的触觉反馈技术^[83], 主要用于可穿戴设备的相关研究中, 作为警示信号^[84]或空间导航^[85,86], 以及分担过于复杂的视听通道信息负荷^[87,88]. Spelmezan 等^[89]在可穿戴设备的研究中展示了一组全身触觉反馈模式的示范设计和相关实验, 成功地用触觉信号代替了身体运动动作指导场景中的声音提示, 并呈现出良好的指导效果. 随着触摸屏的快速普及, 面向触屏的触觉反馈技术也在近年出现了比较大的进展. Hoggan 等^[90]围绕键盘触感问题的研究表明, 利用触觉反馈增强手指触摸屏上的键盘触感能够提升相应的文字输入效能. Altinsoy 等^[91]研究了当屏幕上的图形按钮被触摸时, 系统提供听觉、触觉以及二者的组合反馈时的用户表现. 关于触觉反馈的相关技术, 还有 Jansen 等^[92]提出的一种能在触觉桌面上提供局部或全局反馈的触觉反馈技术, Bau 等^[93]提出的一种基于电磁振动原理的触觉反馈技术等. 更早的关于触觉反馈技术可以参照 Stone 的综述^[94].

目前, 具有代表性的力/触觉交互技术研究有微软公司的 3D 触觉反馈触摸屏, 它包含一个 LCD 平板屏幕、多个力传感器和一个可来回移动的机械臂, 当用户用手触摸屏幕时, 力传感器会捕获用户的施力信息, 并结合其他参数由机械臂平稳地移动屏幕以生成力反馈和物体形态触觉^[95]. 类似的研究还有微软公司的感光形状记忆聚合物显示屏研究^[96], 以及 Biet 等^[97]在超声波触觉板上的压膜效应设计等. 力/触觉交互技术也出现了一些代表性的商业产品, 例如安卓手机的通用触觉反馈效果, Immersion 公司的 TouchSense®触觉反馈技术, SensAble 科技公司的 PHANTOM®系列触觉交互设备利用机械臂提供力反馈^[98], Senseg 公司的 Tixel™触觉像素技术利用静电摩擦力提供接触表面不同的阻力触觉¹³⁾, Tactus 公司的微流体屏幕技术利用流体的移动构造动态的表面凸起与平复^[99]等.

4.6 虚拟现实多通道交互

多通道交互是指在一个系统中组合了 2 个或以上输入通道 (如语音、视频、触觉和手势等) 的协作方式, 由于它充分利用了人类不同的感觉通道, 因而使得交互更为自然有效. 在多通道用户界面中, 用户可以使用自然的交互方式如语音、手势、眼神、表情、唇动等与计算机系统进行协同工作, 人和机器都被看作是信息交流的主动参与者, 输入通道之间有串行/并行、互补/独立等多种方式, 人机交互向人与人交互的形式靠拢, 大幅度提高了交互的自然性和高效性, 将是未来虚拟现实人机交互的主流形式.

Cohen 等^[100]通过实验证明: 如果能用语音方式替代键盘方式, 那么不仅能避免键盘和鼠标之间频繁切换带来的不自然, 而且通过两者的结合能有效地弥补单个通道的不足, 提高交互的效率. Hauptmann^[101]在直接操纵界面里设计了一个将语音和手势结合起来的模拟实验, 发现更多的人倾向于将两者结合起来使用. Oviatt 发现^[102], 在对地图进行操作时, 如果单纯使用语音会产生很多不便, 但如果同时使用其他通道, 如指点, 那么用户将会感觉更方便. Bowman 等^[7]指出, 在虚拟现实中使用多通道交互具有如下优点:

13) Senseg. The Senseg Tixel. <http://senseg.com/technology/senseg-technology>, Accessed April 1st, 2014.

- 降低耦合度. 使用在主输入通道不同的输入通道进行虚拟交互, 可以减少用户的认知负荷. 如果用户不需要在操作和系统控制之间进行切换, 他们可以始终专注在主要的操作上.
- 减少错误和修正. 在输入模糊或者有干扰的时候使用多输入通道是非常有效的, 特别是当应用了基于识别的输入 (如语音或手势). 多通道的输入联合可以显著提高识别率.
- 灵活性和行为补充. 完成同样的任务, 如果用户使用多通道, 输入会更为灵活. 此外, 基于任务的感知结构用互补的方式使用不同的通道.
- 智力资源的控制. 使用多通道交互可以减少认知负荷, 但同时访问多个智力资源也可能导致交互的低效率.

下面, 我们将结合前文介绍的多种交互技术, 从虚拟现实中的多通道交互和多通道信息融合两个方面介绍研究现状和发展趋势.

4.6.1 典型的多通道交互技术

多通道交互用于图形显示的第一个探索性研究是 Bolt 的 Put-That-There 系统^[34], 该系统集成了语音输入和基于六自由度 (DOF) 跟踪器的指点输入, 在背投屏幕前进行 2D 基本图形元素的创建和编辑. 以 Bolt 为基准, 众多的研究者开始了对多通道交互的进一步探索. 在虚拟现实中, 由于能够跟踪和捕捉人体运动的 3D 数据, 这为其多通道界面设计提供了更大的设计空间, 同时也带来了设计上的复杂性. 由于手势与语音结合是最为直观的多通道交互方式, 因而在虚拟现实中讨论的也是最多的.

Lucente 等^[103] 使用语音识别和基于视频的手、体跟踪输入, 使用户能够在 2D 显示器前操作大的对象, 如选取、拖动和比例操作. LaViola 的 MSVT 系统^[104] 使用全手操作和语音输入两个输入通道用于科学数据可视化, 他的目标是不仅为科学可视化应用构造自然和直观的用户界面, 而且进一步探讨了多通道输入的组合风格: (a) 互补性; (b) 并发性; (c) 专用化; (d) 通道间的传递. Cohen 等的 QuickSet 系统^[105] 是采用无线手持设备的基于 Agent 的协同多通道系统, 系统实时分析连续的语音和笔手势多通道输入, 通过合一整合用于地图导航, 他们与美国海军研究实验室 (NRL) 合作实现了 QuickSet 的三维版本, 并集成到 NRL 的虚拟战场规划平台 Dragon2 中, 他们使用一个定制的 6 DOF flight stick 设备, 用于导航、视点控制、选择和数字墨水的勾画, 语音和手势的输入由 QuickSet 并行识别和融合. Latoschik^[106] 对虚拟现实环境下的多通道交互进行了深入的研究, 提出了多通道虚拟现实交互的用户界面框架及基于传输网 (tANT) 的多通道融合、姿势处理框架, 并分析了多通道交互中的指称手势问题.

另外一些研究者以手部手势 + 语音为主, 考虑视线、体手势等其他输入通道在虚拟现实中的应用问题. 例如 Koons 等^[107] 使用 3D 指点手势、声音和眼动跟踪等交互技术, 实现了一个地图交互系统, 他们还讨论了并发多通道出现的问题. Oviatt 等^[108] 描述了一个虚拟现实多通道 3D 虚拟航行器辅助维护系统, 包括基于肢体跟踪的化身驱动、手势识别 (7 个基于 CyberGlove 的手势) 和语音输入, 实现了时间并发输入通道的融合.

多通道虚拟现实交互环境的体系结构是另外一个重要问题. Althoff 等^[109] 描述了多通道虚拟现实交互的体系结构和一些技术问题. 他们的系统集成了多种输入设备 (包括传统的鼠标、键盘、游戏杆、触摸屏以及空间鼠标、跟踪器和数据手套等), 也采用语音输入, 使用动态头部和手势输入. Touraine 等^[110] 的多通道虚拟现实交互框架支持分布式设备服务, 使用专门的服务器用于管理跟踪器、数据手套、语音或手势识别, 以及触觉设备和多通道融合. 多通道融合的结果事件通过网络传递到图形服务器进行绘制更新. Ilmonen 的 FLUID 也提出了类似的思想^[111]. 输入设备的管理是保障多通道系统实

现的一个重要问题,尤其在虚拟现实环境下,由于使用的设备种类更多,也因此出现了一些设备管理工具,如 VRPN^[14] (virtual reality peripheral network) 和 OpenTracker^[112] 等。

事实上,近年来在混合现实和增强现实研究领域出现的混合用户界面也可以把它视为一类多通道界面,系统在这类界面中集成了不同尺寸的显示设备,如将大屏幕显示、头盔 (HMD)、或手持、移动设备作为虚拟现实系统的输入设备。混合用户界面在交互上使用 2D 和 3D 交互混合的模式,这类系统在输出显示上与多通道交互一样,需要把输出信息分解到不同的设备上或其他知觉通道;在输入上,并不像语音 + 姿势组合模式那样强调融合,而是强调不同交互技术之间的无缝转换。美国哥伦比亚大学的 Feiner 等^[113] 在 1991 年就提出了混合用户界面的概念。最近,他所领导的研究小组致力于把这一概念推广到虚拟现实和混合现实环境中,如小组成员 Benko 等^[114] 提出了交叉维手势交互技术:在一个混合沉浸式环境中用触摸平板显示器和 HMD 作为显示设备构成多显示器混合现实环境,输入设备是带有跟踪器的数据手套,使用集成的 2D 和 3D 手势来支持数据在本地 2D 显示和 3D 显示上下文之间的无缝转换。在 Benko 等^[115] 的另一个工作中,使用同样的技术并引入 Tablet PC 作为手持 Widget,通过 Magic Lenses 隐喻实现场景动态缩放,更多关于混合现实的研究见综述^[116]。

Schmalstieg 等^[117] 为解决协同增强现实的交互问题,开发了著名的 Studierstube 系统。最近他及其合作者从多用户、多主机平台、多显示类型、多并发应用和多上下文的角度出发,提出了多维度用户界面的概念^[118],并基于 Studierstube 软件框架开发了几个原型系统进行验证。在他们的多维度用户界面中,所有用户使用的是同一类型的交互设备——带有跟踪器的笔板设备 (personal interaction panel, PIP),显示设备包括 HMD 和视频投影。此外,Butz 等的 EMMIE 系统^[119] 也是一个混合界面的范例,该系统集成了被跟踪的手持设备、固定显示器和 HMD,使用 3D 鼠标作为输入设备,实现跨显示器的拖动操作。Darken^[120] 则从维度一致性的角度研究了虚拟环境下的混合维交互问题。

从现有的研究看,多通道交互技术在虚拟现实中的应用主要面临以下几个问题:

- 交互输入的协调工作。这里既有不同知觉通道输入的融合问题,也有同一通道在不同维度交互技术之间的无缝转换。此外,交互通道的选择也直接决定着系统工作性能;
- 输出显示问题。这是与通常的 2D 多通道技术的又一个不同之处,使用同一虚拟环境配置不同尺寸的显示设备已经越来越普遍,但如何使信息的输出与显示设备匹配是一个值得研究的问题;
- 合理的系统构架。如何协调输入、输出和绘制引擎的协调工作,是另一个值得深入探讨的问题。

4.6.2 多通道交互信息融合

多通道系统使用户同时使用不同的通道进行交互,这些交互通常基于语音、姿势或触觉输入。此外,像面部表情识别或唇读也用于多通道输入。多通道界面能组合单独通道的优势,或根据环境上下文转换通道。

由于多通道技术融合了多个通道的输入流,因而在虚拟现实中使用多通道交互技术可以极大地提高系统控制的性能。但在融合过程中需要考虑几个输入通道之间的组合风格^[104]: (1) 互补性; (2) 并行性; (3) 专用化; (4) 通道间的传递。多通道融合主要有两种方式:

- 早期融合,也称为特征融合。融合是基于原始输入数据在信号级进行融合,这种方式适用于被融合的通道是紧密耦合的,如语音 + 唇读。
- 后期融合,也称为语义融合。其过程是把输入数据映射为语义解释的过程:首先从输入通道获取输入信息流,通过初步的预处理,构建统一的数据表示。这些不同通道输入数据流的统一表示作为

14) VRPN. www.cs.unc.edu/Research/vrpn, 2006.9.

token (每一个 token 具有时间戳标记), 用于构成交互语义; 然后, 基于统一表示的 token 进行数据融合; 最后从词法、语法和句法三个层次上进行融合. Token 按时间排列, 并送入语义分析模块中. 基于框架、神经网络和 Agent 等的推理机制被用于融合过程. 后期融合适用于具有不同时间刻度特性的通道间融合.

语音和基于姿势的融合通常采用后期融合的策略. 在后期融合中, 尽管融合算法的细节有所偏差, 但从策略上看是一致的, 即采用属性/值结构合并和递归匹配的策略^[121]. Nigay 等^[122]在他们的 PAC-Amodeus 系统中给出了一个成熟的方法. 他们介绍了一个可以嵌入到多代理多通道结构中的通用融合引擎, 这个引擎采用融合槽 (melting-pot) 机制和规则驱动策略, 试图融合三类数据: (1) 微时序 (micro temporal: 同一时刻产生的输入信息的融合); (2) 宏时序 (macro temporal: 连续输入信息的融合); (3) 上下文融合 (输入的信息涉及到上下文). 但 Johnston 等^[123]指出, 没有普遍的和定义明确的多通道融合机制, 他们采用基于合一 (unification-based) 的融合方法, 多通道的输入被转化成类型特征结构, 这些特征结构代表不同通道对语义的相应贡献. 合一化操作通过判断这些残缺不全的特征结构是否具有语义上的一致性, 将有一致性的特征结构合并, 最后形成系统能够理解的完整形式. 一些基于统计的方法也用于多通道融合中, 如 Kaise 等^[124]为解决沉浸式增强和虚拟现实的 3D 多通道交互中各个信息源的不可靠问题, 提出了使用来自于语音、3D 手势和参考代理的符号和统计信息进行多通道融合, 以在不同通道间互消歧义 (mutual disambiguation).

Billinghurst 在文献 [80] 中从界面设计的角度给出了多通道交互设计的几点建议:

- 语音输入适用于非图形操作命令或控制任务;
- 包括直证和图标手势在内的手势输入适用于视觉/空间输入;
- 上下文知识适用于分解模糊的输入;
- 多重知识表示适用于理解用户输入;
- 基于语法或类型约束的合一 (Unification) 方法适用于多通道整合.

在国内, 中国科学院软件研究所^[4,125~127]、北京大学^[4,128,129]、中国科学院计算技术研究所^[130,131]、中国科学院自动化技术所^[132]、清华大学¹⁵⁾、浙江大学^[4]等研究机构 and 高校在多通道用户界面领域展开了广泛研究, 在不同的方面取得了大量进展. 这些成果为进一步研究虚拟现实下的多通道交互技术提供了良好的基础.

4.6.3 多通道交互存在的问题

总之, 多通道交互为解决虚拟现实交互问题提供了一条有效的途径, 研究虚拟环境下多通道交互信息融合和显示技术已逐渐成为虚拟交互技术研究领域的重点之一. 在现有的硬件技术水平条件下, 虚拟现实环境下的多通道交互技术亟待解决以下几个方面的问题:

- 多通道的任务建模. 虚拟现实环境下的一般交互任务包括导航/漫游、选择/操作和系统控制三类, 如何针对不同的交互通道, 建立交互语义模型, 在交互任务、输入通道以及交互技术之间建立映射关系, 是虚拟现实多通道交互设计的基础.
- 多通道交互支撑技术. 虚拟现实环境中, 使用的交互设备众多, 这就需要对这些设备进行适当的抽象, 进而建立统一的管理机制. 多通道输入需要识别 (语音识别、姿势识别等), 而识别结果的精度和识别算法的效率直接影响着交互的效果. 此外, 单一通道的虚拟交互技术也是一个重要的内容. 因此, 有必要对这些支撑技术进行研究, 从而为虚拟现实多通道交互提供必要的保障.

15) 清华大学人机交互与媒体集成研究所. <http://media.cs.tsinghua.edu.cn/>. 2006.9.

- 多通道交互信息融合方法. 多通道交互主要借助不同通道的互补性以及通道任务的适应性来提高交互的效率, 因此不同交互通道交互信息的融合是多通道虚拟现实的核心问题. 适当的融合策略和健壮的融合算法, 将保障多通道交互的可用性.

- 混合交互技术. 目前的多通道虚拟现实交互技术主要是语音辅助和三维直接操作, 而一些基于高级隐喻的交互技术也有极高的交互效率和可用性, 如何把这些已有的基于高级隐喻的交互技术用于多通道交互中, 构建更为高可用的交互技术, 是一个具有挑战意义的研究领域.

- 交互技术的测试评估是交互技术研究开发过程中的重要一环. 采用合理的测试方法对交互技术进行评估, 评估结果将为交互技术的选择提供有力的依据. 集成经过测试的交互技术, 构成的交互工具箱, 为广泛应用提供工具, 是交互技术研究的最终目的.

5 面向增强现实的交互

增强现实 (augmented reality, AR) 是交互式计算机图形学中一个相对较新的研究领域. 增强现实强调 3D 虚拟对象与真实世界物体的合成和交互, 这为解决一些领域的应用问题提供了新的手段和方法, 详细的情况请参见 Azuma 在 1997 年和 2001 年对 AR 技术所做的两次综述^[133,134].

Azuma 认为 AR 具有三个关键属性: (1) 在真实环境中组合真实物体和虚拟对象; (2) 实时交互; (3) 真实物体和虚拟对象互相注册. 当前 AR 技术需要解决的基本问题是注册跟踪和交互问题. AR 强调实时交互, 因而从交互的观点来看, 它是一类新出现的 3D 用户界面 (3D UI). 总体来说, 与传统虚拟现实交互和桌面 2D 交互相比, AR 环境下的界面设计需要考虑: 与虚拟对象和真实物体无缝交互、面临多个不同类型的显示器和输入设备以及多用户协同工作.

AR 增强了用户对真实世界的感知以及与真实世界的交互, 从这个意义上讲, AR 系统的设计就等同于界面设计. AR 技术在医学可视化、设备保养与维修、机器人路径规划、娱乐教育、环境注释 (如古迹漫游) 和军事等领域有着广阔的应用前景, 而 AR 在这些领域的应用就是设计针对领域的 AR 界面.

5.1 基于 3D 虚拟交互的 AR 界面

目前已有的 AR 界面主要使用了基于 3D 虚拟交互和基于 TUI 两种交互隐喻. 这类界面使用六自由度输入设备跟踪用户的手部运动. 跟踪器通常用于虚拟对象的操作或视点控制, 虚拟对象由光学或视频的透视头盔展现. 这些系统中的交互任务与虚拟现实中描述的 3D 交互任务相似, 如 3D 虚拟对象的选择/操作、系统控制等, 这使许多为虚拟环境开发的交互技术可以移植到 AR 环境中. 这方面工作主要有用于简单浏览叠加信息的, 如 Fitzmauric 介绍的 Chameleon 系统^[52] 以及 Rekimoto 等的 NaviCam 系统^[55]. 为了超越简单的信息叠加, Bell 等^[135] 把 WIM (一种 3D 虚拟交互技术: 缩微世界) 技术引入到这类界面中从而为真实环境提供三维地图. 其他的一些研究者也对此进行了研究, 如 Kiyokawa 等的 InfoPoint^[136].

把 3D 虚拟交互应用到 AR 界面中是一个重要的研究和开发方向^[6], 已经在娱乐、设计和训练等领域中得到成功应用. 它们提供了“无缝空间交互”, 用户能够与叠加到物理环境中任何位置的虚拟对象交互; 支持自然、熟悉的交互隐喻, 如简单虚拟手. 然而, 这种交互风格也存在以下几个问题: (1) 3D 虚拟交互没有足够的触觉反馈; (2) 通常需要一个头盔显示器; (3) 用户需要为物理对象和虚拟对象使用不同的输入通道 (真实的手用于物理对象, 专门目的的输入设备用于虚拟对象, 用户需要在两者之

间切换), 这就导致了在自然交互流中的“空间缝隙”。

5.2 基于 TUI 的 AR 界面

1991 年, Mark Weiser 提出了“无处不在的计算”的概念^[137], 作为未来人机交互的范式。实物用户界面 (TUI) 就是这种全新交互范型的一种尝试。TUI 是由 Ishii 和 Ullmer 在 CHI 1997 的会议上提出的^[138], 他们把 TUI 定义为: 把数字信息连接到日常物理对象或环境中, 以增强物理世界。TUI 有以下几个特点: (1) 在实物中嵌入信息和功能; (2) 以实物的形式表示信息; (3) 通过操纵实物操纵数字信息; (4) 空间可重配置。Ishii 在 MIT 领导的 TUI 研究小组做了大量的工作^[16], 并试图通过 MCRpd 提供统一的交互模型^[139]。最近, 其他一些研究者开始着手建立 TUI 的分类体系和交互范式, 如 Kenneth^[140] 基于体现 (Embodiment) 和隐喻两个角度建立了 TUI 技术的分类体系; Orit 等^[141] 试图通过 TAC (Token and Constraints) 范式来抽象和统一 TUI 交互, 并为 TUI 软件工具的开发提供描述方法。

TUI 一经出现, 就迅速应用到 AR 中。目前, 在 AR 界面中使用 TUI 技术已成为一个新的研究热点。基于 TUI 的 AR 界面主要三类: 实物 Widget 界面、实物 AR 界面和转换 AR 界面。

实物 Widget 界面。这类 AR 界面把 GUI 界面中的交互元素实物化, 设计出专门用于控制虚拟对象的物理工具 Phicons。它不使用 HMD 显示器, 而是使用顶部或背部投影把数字信息注册到工作表面上, 同时, Phicons 也置于同一工作表面上。Ullmer 等^[142] 的 MetaDesk 系统是这种方式的一个重要范例。

实物 Widget 界面中, 用户与物理实物和虚拟对象交互, 使用人手直接操作这一相同的输入通道, 这使交互更为容易。但该方法仅在工作表面上叠加虚拟 3D 对象, 而不是在 3D 空间中的任何位置上叠加虚拟对象, 因此, 交互被限制在 2D 表面上, 进而不能实现空间交互, 也就是说, 在交互流中存在“空间缝隙”问题。

实物 AR 界面。Billinghurst 等在总结了几种原型系统的基础上, 提出了实物 AR (TAR) 新界面设计隐喻。实物 AR 界面把 AR 显示器和 TUI 控制结合起来, 2D 或 3D 虚拟对象被注册到被标记的物理对象上, 使用视频透视的 AR 注册技术显示给用户 (即用户通过装在 HMD 上的相机观察真实世界的环境)。用户与虚拟对象的交互是直接通过操作相应的物理对象来实现, 而不是通过实物 Widget 间接控制。由于多个用户能够同时与虚拟对象交互, 因此, 它自然实现了协作式界面。Billinghurst 等^[143] 从空间时间复用的角度讨论了以下几个原型:

- **空间复用界面类型。**空间复用界面是指多个设备, 每个设备只具有一个功能。Poupyrev 等^[144] 的 Tiles 系统是一个 AR 创作界面。Tiles 在物理外观上是一系列具有不同黑白图案的卡片, 图案形状表示了特定的语义。Tiles 有三种类型: 数据 Tiles、操作 Tiles 和菜单 Tiles。系统基于视频标记跟踪技术, 通过 Tiles 之间的空间位置关系以及 Tiles 所表示的语义 (数据、操作或菜单命令) 执行交互任务 (为防止误操作, 需要一定的延时)。通过 Tiles 实现的界面模型定义了一个系统性和通用性的操作集, 这是一个简单、通用的 TAR 用户界面范式, 它可以用来设计广泛的应用。

- **时间复用界面类型。**时间复用是指一个设备在不同的时间点上具有不同的功能。Kato 等^[145] 的 VOMAR 系统提供给用户类似鼠标的单一通用工具——Paddle 来操作所有对象。Paddle 是一个带有跟踪标记的纸片, 两个手都能使用它, 并能让用户使用静态手势 (如靠近关系 + 延时) 和动态手势 (如摇动) 与虚拟对象交互。Paddle 用虚拟家具布置任务来评估这个系统。

TAR 界面方法所提倡的界面风格在 3D 增强现实和增强表面之间架起了桥梁。首先, 作为透明界

16) MIT Media Lab Tangible Media Group. <http://tangible.media.mit.edu>.

面, 它提供了与虚拟对象和物理对象无缝的双手 3D 交互, 并且不需要专门的输入设备; 其次, TAR 消除了数字空间和物理空间之间的空间交互缝隙, 对虚拟对象的操作可以在空间的任意位置进行. TAR 界面的主要限制有: 一是需要使用 HMD; 二是为了进行可靠的跟踪, 标记必须是可见的.

转换 AR 界面. Billinghurst 等^[146] 的 Magic Book 使用手持增强现实显示器 (HHD) 和视频跟踪技术叠加虚拟模型到真实的书页上, 实现了可以在真实和虚拟之间无缝变换的可转换混合现实界面. Magic Book 界面提供了一种新的界面隐喻, 采用了 TAR 界面元素并且支持多用户协同工作, 但相对于虚拟现实限制了身体输入 (没有手部跟踪).

AR 用户界面还有许多领域值得探索. 对于未来在实际应用中开发和利用 AR 来说, 开发新的输入/输出设备、跟踪技术和注册技术是至关重要的. 此外, 设计新的交互技术以便用户与物理对象和虚拟对象无缝交互, 也是一个令人感兴趣的未来研究方向. 另一个可能产生丰硕成果的领域是已有的 3D 交互技术和界面设计原理如何应用和调整到 AR 界面中. 最后, 在 AR 交互的人类因素方面和界面的可用性评估方面也有待开展更多的研究工作.

6 问题与展望

以上概述了虚拟现实人机交互技术目前的研究状态, 得益于图形学技术和显示技术的发展, 虚拟现实技术有了长足的进步, 尤其是头盔技术的发展以及虚拟现实与互联网和人工智能的结合, 使得虚拟现实有成为一种新型个人计算平台的可能, 这就促使人们开始关注更高效、更适人的交互技术, 以应对虚拟现实应用的多样化需求. 从已有技术现状看, 虚拟现实环境下人机交互还面临如下挑战问题:

- 用户在虚拟环境中自然交互行为与状态感知问题. 虽然人们已能通过不同类型的传感器采集到包括人体的外观、深度、脑电、生理生化参数以及语音等信息或信号, 但这些信息或信号只是人的自然行为或状态的外显测量, 对人的自然行为或状态的刻画是不精确的, 如何从这些信息流中准确、实时地提取出刻画人的自然行为或状态的特征原语, 是解决虚拟现实环境下交互感知的核心问题.

- 虚拟现实环境对人交互行为与意图的理解问题. 虽然借助机器学习、模式识别等技术计算机已能部分地、简单地理解用户的意图, 但用户状态转换以及复杂动作的意图理解还具有很多不确定性问题. 因此, 如何理解人的认知加工机制, 并抽象出用于理解人的交互意图的计算模型, 是实现虚拟现实交互认知的核心问题.

- 虚拟现实交互输出反馈的自适应问题. 虚拟环境生成的反馈信息要为人所感知, 而人的感知通道具有多态性, 如何在多模态、强时变的输出空间中产生符合认知加工机制的反馈呈现, 是实现虚拟现实交互反馈的核心问题.

因此, 在未来我们期待虚拟现实人机交互在以下几方面能有突破性的进展:

- 虚拟现实人机交互基础研究. 着力研究人类感知机理与基于认知的自然人机交互理论与方法, 包括研究人类多感官 (视、听、触、力、嗅、味) 通道对虚拟现实、增强现实等生成环境的感知机理, 发展新的感知装置, 为虚拟现实、增强现实呈现设备与装置以及人机交互技术的研制与研究提供基础理论支持; 支持虚拟现实环境下人类生理反应机制、认知过程和情绪机制等研究, 研究多通道反馈呈现与适应性调节方法, 构建生理计算理论; 推动虚拟现实环境下人类交互行为研究和多人协同交互研究, 发展符合人类固有运动技能的全身交互技术, 探究多通道交互机制, 构建自然人机交互理论与方法.

- 虚拟现实人机交互共性关键技术. 加快自然人机交互核心技术研发及产品化. 提升轻量级、微小型化、低功耗智能交互硬件设备的设计研发水平, 研发能精确、实时快速感知人类运动行为、语音

以及生理状态的智能感知设备, 为智能化、网络化虚拟现实提供智能感知终端, 实现对人的行为及状态的真实感知; 突破个体手眼分离条件下的语言、手势、姿态以及眼动等多通道交互技术, 实现交互意图理解, 为移动开放空间中的个人虚拟现实类应用提供自然交互技术; 研究面向有限空间范围的多人、多通道交互技术, 为群体类虚拟现实应用提供支撑平台; 研究语言、姿势、动作、表情、眼动等行为与生理变化之间的关联关系, 形成虚拟现实临场感评测工具, 为提升人在复杂环境中的工作绩效、专业技能人员培训效率以及运动、认知或精神方面受损的患者治疗与康复水平提供共性支持, 也为评价虚拟现实系统本身的评测提供工具。

• 标准与规范建设. 推动虚拟现实在保真度、显示技术、交互设备、交互技术与界面等方面的标准与规范建设. 鼓励虚拟现实软硬件与互联网接入的接口标准化。

参考文献

- 1 Wang C W, Gao W, Wang X R. The Theory, Implementation and Application of Virtual Reality Technology. Beijing: Tsinghua University Press, 1996 [汪成为, 高文, 王行仁. 灵境 (虚拟现实) 技术的理论、实现及应用. 北京: 清华大学出版社, 1996]
- 2 Green M, Jacob R. SIGGRAPH'90 workshop report: software architectures and metaphors for non-WIMP user interfaces. Comput Graph, 1991, 25: 229–235
- 3 van Dam A. Post-WIMP user interface. Commun ACM, 1997. 40: 63–67
- 4 Dong S H, Wang J, Dai G Z. Human-Computer Interaction and Multi Model User Interface. Beijing: Science Press, 1999 [董士海, 王坚, 戴国忠. 人机交互和多通道用户界面. 北京: 科学出版社, 1999]
- 5 Jacob R J K, Girouard A, Hirshfield L M, et al. Reality-based interaction: a framework for Post-WIMP interfaces. In: Proceedings of ACM CHI, Florence, 2008. 201–210
- 6 Bowman D, Kruijff E, LaViola J, et al. 3D User Interface: Theory and Practice. Boston: Addison-Wesley Professional, 2004
- 7 LaViola J, Zeleznik R. Flex and pinch: a case study of whole hand input design for virtual environment interaction. In: Proceedings of the 2nd IASTED International Conference on Computer Graphics and Imaging, California: Palm Springs, 1999. 221–225
- 8 Lin C P, Wang C Y, Chen H R, et al. RealSense: directional interaction for proximate mobile sharing using built-in orientation sensors. In: Proceedings of the 21st ACM International Conference on Multimedia, Barcelona, 2013. 777–780
- 9 Mine M. Virtual environment interaction techniques. UNC Chapel Hill CS Dept: Technical Report TR95-018. 1995
- 10 Liang J, Green M. Jdcad: a highly interactive 3D modeling system. In: Proceedings of 3rd International Conference on CAD and Computer Graphics, Beijing, 1994. 499–506
- 11 Poupyrev I, Billingham M, Weghorst S, et al. Go-Go interaction technique: non-linear mapping for direct manipulation in VR. In: Proceedings of UIST, Seattle, 1996. 79–80
- 12 Bowman D, Hodges L. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In: Proceedings of Symposium on Interactive 3D Graphics, Providence, 1997. 35–38
- 13 Wloka M, Greenfield E. The virtual tricorder: a uniform interface for virtual reality. In: Proceedings of UIST, Pittsburgh, 1995. 39–40
- 14 Janzen I, Rajendran V K, Booth K S. Modeling the impact of depth on pointing performance. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, San Jose, 2016. 188–199
- 15 Stoakley R, Conway M, Pausch R. Virtual reality on a Wim: interactive worlds in miniature. In: Proceedings of ACM CHI, Denver, 1995. 265–272
- 16 Pierce J, Forsberg A, Conway M, et al. Image plane interaction techniques in 3D immersive environments. In: Proceedings of Symposium on Interactive 3D Graphics, Providence. 1997. 39–43
- 17 Pierce J, Stearns B, Pausch R. Two handed manipulation of Voodoo Dolls in virtual environments. In: Proceedings of Symposium on Interactive 3D Graphics, Atlanta, 1999. 141–145
- 18 Mine M, Brooks F, Sequin C. Moving objects in space: exploiting proprioception in virtual-environment interaction. In: Proceedings of SIGGRAPH, Los Angeles, 1997. 19–26

- 19 Szalavari Z, Gervautz M. The personal interaction panel — a two-handed interface for augmented reality. *Comput Graph Forum*, 1997, 16: 335–346
- 20 Schmalstieg D, Encarnacao L, Szalavari Z. Using transparent props for interaction with the virtual table. In: *Proceedings of Symposium on Interactive 3D Graphics*, Atlanta, 1999. 147–154
- 21 Billinghamurst M, Baldis S, Matheson L, et al. 3D palette, a virtual reality content creation tool. In: *Proceedings of VRST*, Lausanne, 1997. 155–156
- 22 Poupyrev I, Tomokazu N, Weghorst S. Virtual notepad: handwriting in immersive VR. In: *Proceedings of VRAIS*, Atlanta, 1998. 126–132
- 23 Nancel M, Chapuis O, Pietriga E, et al. High-precision pointing on large wall displays using small handheld devices. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Paris, 2013. 831–840
- 24 Olivier C, Bezerianos A, Frantzeskakis S. Smarties: an input system for wall display development. In: *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems*, Toronto, 2014. 2763–2772
- 25 Fakourfar O, Ta K, Tang R, et al. Stabilized annotations for mobile remote assistance. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, San Jose, 2016. 1548–1560
- 26 Pick S, Puika A S, Kuhlen T W. SWIFTER: design and evaluation of a speech-based text input metaphor for immersive virtual environments. In: *Proceedings of 2016 IEEE Symposium on 3D User Interfaces*, Greenville, 2016. 109–112
- 27 Mapes D, Moshell J. A two-handed interface for object manipulation in virtual environments. *Presence: Teleop Virt Environ*, 1995, 4: 403–416
- 28 Conner B, Snibbe S, Herndon K, et al. Three-dimensional widgets. In: *Proceedings of Interactive 3D Graphics Symposium*, Cambridge, 1992. 183–188
- 29 Arun K, Joseph J, LaViola J. Exploring 3D user interface technologies for improving the gaming experience. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, Seoul, 2015. 125–134
- 30 Cerney M M, Vance J M. Gesture recognition in virtual environments: a review and framework for future development. *Iowa State University Human Computer Interaction Technical Report ISU-HCI-2005-01*. 2005
- 31 Cadoz C. *Les Realites Virtuelles*. Dominos: Flammarion, 1994
- 32 McNeill D. *Hand and Mind: What Gestures Reveal About Thought*. Chicago: University of Chicago Press, 1992
- 33 Brooks J F P. Grasping reality through illusion: interactive graphics serving science. In: *Proceedings of ACM CHI*, Washington, 1988. 1–11
- 34 Bolt R A. Put-that-there: voice and gesture at the graphics interface. In: *Proceedings of ACM SIGGRAPH Computer Graphics*, New York, 1980. 262–270
- 35 McGuire R M, Hernandez R J, Starner T, et al. Towards a one-way American sign language translator. In: *Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition*, Paris, 2004. 620–625
- 36 Bolt R A, Herranz E. Two-handed gesture in multi-modal dialog. In: *Proceedings of the ACM Symposium on User Interface Software and Technology*, Monterey, 1992. 7–14
- 37 Cavazza M, Pouteau X, Pernel D. Multimodal communication in virtual environments. *Adv Hum Factor*, 1995, 20: 579–604
- 38 Granit L, Simon J, Elisabeth L, et al. A sliding window approach to natural hand gesture recognition using a custom data glove. In: *Proceedings of IEEE Symposium on 3D User Interfaces*, Greenville, 2016. 81–90
- 39 Zhang Z. Microsoft kinect sensor and its effect. In: *Proceedings of IEEE*. Piscataway: IEEE, 2012. 4–10
- 40 Potter L E, Araullo J, Carter L. The leap motion controller: a view on sign language. In: *Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration*. New York: ACM, 2013. 175–178
- 41 Alejandro J, Nicu S. Multimodal human computer interaction: a survey. In: *Proceedings of IEEE International Workshop on Human Computer Interaction in Conjunction with ICCV*, Beijing, 2005. 1–15
- 42 LaViola J J. An introduction to 3D gestural interfaces. In: *Proceeding ACM SIGGRAPH 2014 Courses*, Vancouver, 2014. Article No. 25
- 43 LaViola J J. 3D gestural interaction: the state of the field. *ISRN Artif Intell*, 2013: 514641
- 44 Porta M. Vision-based user interfaces: methods and applications. *Int J Hum Comput Stud*, 2002, 57: 27–73
- 45 Pavlovic V I, Sharma S, Huang T S. Visual interpretation of hand gestures for human computer interaction: a review. *IEEE Trans Pattern Anal Mach Intell*, 1997, 19: 677–695

- 46 Shahidul I, Ionescu B, Gadea C, et al. Full-body tracking using a sensor array system and laser-based sweeps. In: Proceedings of IEEE Symposium on 3D User Interfaces, Greenville, 2016. 71–80
- 47 Zhang Y, Zhou J, Laput G, et al. Skintrack: using the body as an electrical waveguide for continuous finger tracking on the skin. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, San Jose, 2016. 14911503
- 48 Arun K, Joseph J, LaViola J. Exploring the usefulness of finger-based 3D gesture menu selection. In: Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems, Toronto, 2014. 1093–1102
- 49 Hinckley K, Pierce J, Sinclair M, et al. Sensing techniques for mobile interaction. In: Proceedings of UIST, California, 2000. 91–100
- 50 Zhou F, Duh H B, Billinghurst M. Trends in augmented reality tracking, interaction and display: a review of ten years of ISMAR. In: Proceedings of the 7th International Symposium on Mixed and Augmented Reality, Washington, 2008. 193–202
- 51 Abowd G D, Iftode L, Mitchell H. The smart phone: a first platform for pervasive computing. IEEE Pervas Comput, 2005, 4: 18–19
- 52 Fitzmaurice G W. Situated information spaces and spatially aware palmtop computers. Commun ACM, 1993, 36: 39–49
- 53 Michael R. Linking Physical and Virtual Worlds with Visual Markers and Handheld Devices. Dissertation for Ph.D. Degree. Zurich: Swiss Federal Institute of Technology, 2005
- 54 Fitzmaurice G W, Zhai S, Chignell M. Virtual reality for palmtop computers. ACM Trans Inform Syst, 1993, 11: 197–218
- 55 Rekimoto J, Nagao K. The world through the computer: computer augmented interaction with real world environments. In: Proceedings of the 8th Annual ACM Symposium on User Interface and Software Technology, Pittsburgh, 1995. 29–36
- 56 Wagner D, Schmalstieg D. First steps towards handheld augmented reality. In: Proceedings of the 7th IEEE International Symposium on Wearable Computers, Sardinia, 2003. 127–135
- 57 Rohs M. Visual code widgets for marker-based interaction. In: Proceedings of IEEE International Conference on Distributed Computing Systems Workshops, Columbus, 2005. 506–513
- 58 Rohs M, Gfeller B. Using camera-equipped mobile phones for interacting with real-world objects. Adv Pervas Comput, 2004, 176: 265–271
- 59 Yee K P. Peephole displays: pen interaction on spatially aware handheld computers. In: Proceedings of ACM CHI, Florida, 2003. 1–8
- 60 Mehra S, Werkhoven P, Worring M. Navigating on handheld displays: dynamic versus static peephole navigation. ACM Trans Comput Hum Interact, 2006, 4: 448–457
- 61 Rohs M, Antti O. Target acquisition with camera phones when used as magic lenses. In: Proceedings of ACM CHI, Florence, 2008. 1409–1418
- 62 Morrison A, Oulasvirta A, Peltonen P, et al. Like bees around the hive: a comparative study of a mobile augmented reality map. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Boston, 2009. 1889–1898
- 63 Reitmayr G, Drummond T. Going out: robust model-based tracking for outdoor augmented reality. In: Proceedings of 5th IEEE and ACM International Symposium on Mixed and Augmented Reality, Minnesota, 2006. 109–118
- 64 Miyashita T, Meier P, Tachikawa T, et al. An augmented reality museum guide. In: Proceedings of the 7th IEEE and ACM International Symposium on Mixed and Augmented Reality, San Francisco, 2008. 103–106
- 65 Henrysson A, Billinghurst M, Ollila M. Face to face collaborative AR on mobile phones. In: Proceedings of the 4th IEEE and ACM International Symposium on Mixed and Augmented Reality, Vienna, 2005. 80–89
- 66 Wagner D, Barakonyi I. Augmented reality Kanji learning. In: Proceedings of IEEE International Symposium on Mixed and Augmented Reality, Tokyo, 2003. 335–336
- 67 Schmalstieg D, Wagner D. Experiences with handheld augmented reality. In: Proceedings of IEEE International Symposium on Mixed and Augmented Reality, Cambridge, 2007. 3–18
- 68 Miyaoku K, Higashino S, Tonomura Y. C-Blink: a hue-difference-based light signal marker for large screen interaction via any mobile terminal. In: Proceedings of ACM Symposium on User Interface Software and Technology, New Mexico, 2004. 147–156

- 69 Hachet M, Pouderoux J, Guitton P. A camera-based interface for interaction with mobile handheld computers. In: *Proceedings of Symposium on Interactive 3D Graphics and Games*, New York, 2005. 65–72
- 70 Ballagas R, Borchers J, Rohs M, et al. The smart phone: a ubiquitous input device. *IEEE Pervas Comput*, 2006, 5: 70–77
- 71 Rafael B, Michael R, Jennifer G, et al. Sweep and point & shoot: phonecam based interactions for large public displays. In: *Proceedings of ACM CHI*, Protland, 2005. 1200–1203
- 72 Pears N, Jackson D C. Smart phone interaction with registered display. *IEEE Pervas Comput*, 2009, 8: 14–21
- 73 Jiang H, Ofek E, Moraveji N, et al. Direct pointer: direct manipulation for large-display interaction using handheld cameras. In: *Proceedings of ACM CHI*, Montréal, 2006. 22–27
- 74 Wang J, Zhai S, Canny J. Camera phone based motion sensing: interaction techniques, applications and performance study. In: *Proceedings of ACM Symposium on User Interface Software and Technology*, Montreux, 2006. 101–110
- 75 Sebastian B, Dominikus B, Andreas B, et al. Touch projector: mobile interaction through video. In: *Proceedings of ACM CHI*, Atlanta, 2010. 10–15
- 76 Yang X D, Edward M, David M, et al. LensMouse: augmenting the mouse with an interactive touch display. In: *Proceedings of ACM CHI*, Atlanta, 2010. 10–15
- 77 Gugenheimer J, Dobbelsstein D, Christian W, et al. FaceTouch: enabling touch interaction in display fixed UIs for mobile virtual reality. In: *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, Tokyo, 2016. 49–60
- 78 Henrikson R, Araujo B, Chevalier F, et al. Multi-device storyboards for cinematic narratives in VR. In: *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, Tokyo, 2016. 787–796
- 79 Bowman D, Rhoton C, Pinho M. Text input techniques for immersive virtual environments: an empirical comparison. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Maryland, 2002. 2154–2158
- 80 Billinghamurst M. Put that where? voice and gesture at the graphics interface. In: *Proceedings of ACM SIGGRAPH Computer Graphics*, Orlando, 1998. 60–63
- 81 McTear M F. Spoken dialogue technology: enabling the conversational interface. *ACM Comput Surv*, 2004, 34: 90–169
- 82 Gosselin F, Andriot C, Savall J, et al. Large workspace haptic devices for human-scale interaction: a survey. In: *Proceedings of 6th International Conference on Haptics: Perception, Devices and Scenarios*, Heidelberg, 2008. 523–528
- 83 Schätzle S, Hulin T, Preusche C, et al. Evaluation of vibrotactile feedback to the human arm. In: *Proceedings of EuroHaptics*, Paris, 2006. 577–560
- 84 Cassinelli A, Reynolds C, Ishikawa M. Augmenting spatial awareness with haptic radar. In: *Proceedings of the 10th IEEE International Symposium on Wearable Computers*, Minnesota, 2006. 61–64
- 85 Ross D A, Blasch B B. Wearable interfaces for orientation and wayfinding. In: *Proceedings of the 4th International ACM Conference on Assistive Technologies*, Arlington, 2000. 193–200
- 86 van Erp J B F, van Veen H A H C. Vibro-tactile information presentation in automobiles. In: *Proceedings of Eurohaptics*, Minnesota, 2001. 99–104
- 87 van Veen H A H C, van Erp J B F. Tactile information presentation in the cockpit. In: *Proceedings of Haptic Human-Computer Interaction*. Berlin: Springer Berlin Heidelberg, 2001. 174–181
- 88 Gemperle F, Ota N, Siewiorek D. Design of a wearable tactile display. In: *Proceedings of 5th International Symposium on Wearable Computers*, Argentina, 2001. 5–12
- 89 Spelmezan D, Jacobs M, Hilgers A, et al. Tactile motion instructions for physical activities. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Boston, 2009. 2243–2252
- 90 Hoggan E, Brewster S A, Johnston J. Investigating the effectiveness of tactile feedback for mobile touchscreens. In: *Proceedings of ACM CHI*, Florence. 1573–1582
- 91 Altinsoy M E, Merchel S. Audiotactile feedback design for touch screens. In: *Proceedings of Haptic and Audio Interaction Design*. Berlin: Springer Berlin Heidelberg, 2009. 136–144
- 92 Jansen Y, Karrer T, Borchers J. MudPad: tactile feedback and haptic texture overlay for touch surfaces. In: *Proceedings of ACM International Conference on Interactive Tabletops and Surfaces*, Pittsburgh, 2010. 11–14
- 93 Bau O, Poupyrev I, Israr A, et al. TeslaTouch: electrovibration for touch surfaces. In: *Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology*, California, 2010. 283–292

- 94 Stone R J. Haptic feedback: a brief history from telepresence to virtual reality. In: *Proceedings of Haptic Human-Computer Interaction*. Berlin: Springer Berlin Heidelberg, 2001. 1–16
- 95 Hrvoje B, Christian H, Sinclair M, et al. NormalTouch and texture touch: high-fidelity 3D haptic shape rendering on handheld virtual reality controllers. In: *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, Tokyo, 2016. 717–728
- 96 Kikin G E. Light-induced shape-memory polymer display screen. U.S. Patent 8279200. 2012
- 97 Biet M, Giraud F, Lemairesemail B. Squeeze film effect for the design of an ultrasonic tactile plate. *IEEE Trans Ultrason Ferr Freq Contr*, 2007, 54: 2678–2688
- 98 Massie T H, Salisbury K J. The phantom haptic interface: a device for probing virtual objects. In: *Proceedings of the ASME Winter Annual Meeting, Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*. Amsterdam: IOS Press, 1994. 295–300
- 99 Karlin S. Tactus technology [Resources Start-Ups]. *IEEE Spectrum*, 2013, 50: 23
- 100 Cohen P R, Dalrymple M, Moran D B, et al. Synergistic use of direct manipulation and natural language. In: *Proceedings of ACM SIGCHI Bulletin*, Austin, 1989. 227–233
- 101 Hauptmann A G. Speech and gestures for graphic image manipulation. In: *Proceedings of ACM SIGCHI Bulletin*, Austin, 1989. 241–245
- 102 Oviatt S. Multimodal interactive maps: designing for human performance. *Hum Comput Interact*, 1997, 12: 93–129
- 103 Lucente M, Zwart G J, George A. Visualization space: a testbed for deviceless multimodal user interfaces. In: *Proceedings of AAAI Spring Symposium on Intelligent Envrionments*. Menlo Park: AAAI Press, 1998. 87–92
- 104 Laviola, J. MSVT: a virtual reality-based multimodal scientific visualization tool. In: *Proceedings of IASTED International Conference on Computer Graphics and Imaging*, Innsbruck, 2000. 1–7
- 105 Cohen P R, Johnston M, McGee D R. et al. QuickSet: multimodal interaction for distributed applications. In: *Proceedings of International Multimedia Conference*. New York: ACM Press, 1997. 31–40
- 106 Latoschik M E. A gesture processing framework for multimodal interaction in virtual reality. In: *Proceedings of International Conference on Computer Graphics*, Camps Bay, 2001. 95–100
- 107 Koons B D, Sparrell C J, Thorisson K R. Integrating simultaneous input from speech, gaze, and hand gestures. In: *Proceedings of AAAI Workshop on Intelligent Multimedia Interfaces*, Anaheim, 1993. 257–276
- 108 Oviatt S L, Cohen P R, Wu L, et al. Designing the user interface for multimodal speech and gesture applications: state-of-the-art systems and research directions. *Hum Comput Interact*, 2000, 15: 263–322
- 109 Althoff F, McGlaun G, Schuller B, et al. Using multimodal interaction to navigate in arbitrary virtual vrml worlds. In: *Proceedings of Workshop on Perceptive User Interfaces*, Florida, 2001. 1–8
- 110 Touraine D, Bourdot P, Bellik Y, et al. A framework to manage multimodal fusion of events for advanced interactions within virtual environments. In: *Proceedings of the Workshop on Virtual Environments*, Barcelona, 2002. 159–168
- 111 Ilmonen T, Kontkanen J. Software architecture for multimodal user input-Fluid. In: *Proceedings of ERCIM Workshop on User Interfaces for All*, Heidelberg, 2002. 319–338
- 112 Reitmayr G, Schmalstieg D. An open software architecture for virtual reality interaction. In: *Proceedings of ACM Symposium on Virtual Reality Software and Technology*, Alberta, 2001. 47–54
- 113 Feiner S, Shamash A. Hybrid user interfaces: breeding virtually bigger interfaces for physically smaller computers. In: *Proceedings of ACM Symposium on User Interface Software and Technology*, Hilton Head, 1991. 9–17
- 114 Benko H, Ishak E W, Feiner S. Cross-dimensional gestural interaction techniques for hybrid immersive environments. In: *Proceedings of IEEE Virtual Reality Annual International Symposium*, Bonn, 2005. 209–216
- 115 Benko H, Ishak E W, Feiner S. Collaborative mixed reality visualization of an archaeological excavation. In: *Proceedings of the International Symposium on Mixed and Augmented Reality*, Arlington, 2004. 132–140
- 116 Huang J, Han D Q, Chen Y N, et al. A survey on human-computer interaction in mixed reality. *J Comput Aided Des Comput Graph*, 2016, 28: 869–880 [黄进, 韩冬奇, 陈毅能, 等. 混合现实中的人机交互综述. *计算机辅助设计与图形学学报*, 2016, 28: 869–880]
- 117 Schmalstieg D, Fuhrmann A, Hesina G, et al. The studierstube augmented reality project. *Presence Teleop Virt Environ*, 2002, 11: 33–54
- 118 Schmalstieg D, Fuhrmann A, Hesina G. Bridging multiple user interface dimensions with augmented reality. In: *Proceedings of IEEE and ACM International Symposium on Augmented Reality*, Munich, 2000. 20–29
- 119 Butz A, Hollerer T, Feiner S, et al. Enveloping users and computers in a collaborative 3D augmented reality.

- In: Proceedings of IEEE International Workshop on Augmented Reality, San Francisco, 1999. 35–44
- 120 Darken R P, Durost R. Mixed-dimension interaction in virtual environments. In: Proceedings of the ACM Symposium on Virtual Reality Software and Technology, New York, 2005. 38–45
 - 121 Vuurpijl L G, Bosch L T, Ruiter J D, et al. Overview of the state of the art in fusion of speech and pen input. Technical Report IST02001-32311, 2002
 - 122 Nigay L, Coutaz J. A generic platform for addressing the multimodal challenge. In: Proceedings of Conference on Human Factors in Computing Systems, Denver, 1995. 98–105
 - 123 Johnston M, Cohen P R, McGee D, et al. Unification-based multimodal integration. In: Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics, Madrid, 1997. 1–8
 - 124 Kaiser E, Olwal A, McGee D, et al. Mutual disambiguation of 3D multimodal interaction in augmented and virtual reality. In: Proceedings of the 5th International Conference on Multimodal Interfaces, Vancouver, 2003. 12–19
 - 125 Li J, Tian F, Wang W X, et al. A multimodal interaction system for children. *J Software*, 2002, 13: 1846–1851 [李杰, 田丰, 王维信, 等. 面向儿童的多通道交互系统. *软件学报*, 2002, 13: 1846–851]
 - 126 Ao X, Li J F, Wang X G, et al. Structuralizing digital ink for efficient selection. In: Proceedings of the ACM International Conference on Intelligent User Interfaces, Sydney, 2006. 148–155
 - 127 Wang L, Fu Y G, Ji L E, et al. A layout system by two handed manipulation based on constraints. *J Comput Aided Des Comput Graph*, 2006, 18: 1243–1249 [王亮, 付永刚, 纪连恩, 等. 基于约束语义的双手交互场景布局系统. *计算机辅助设计与图形学学报*, 2006, 18: 1243–1249]
 - 128 Lin Y M, Dong S H. The realization of algorithm and software platform. *Chin J Comput*, 2000, 23: 90–94 [林应明, 董士海. 多通道融合算法和软件平台的实现. *计算机学报* 2000, 23: 90–94]
 - 129 Pu J T, Wang Y, Chen W G, et al. FreeVoiceCAD — a multi-modal user interface prototype system. *J Comput Res Develop*, 2003, 40: 1382–1388 [普建涛, 王悦, 陈文广, 等. 多通道用户界面原型系统 FreeVoiceCAD. *计算机研究与发展*, 2003, 40: 1382–1388]
 - 130 Chen Y Q, Gao W, Liu J F, et al. Multi-model behavior synchronizing prosody model in sign language synthesis. *Chin J Comput*, 2006, 29: 822–827 [陈益强, 高文, 刘军发, 等. 手语合成中的多模式行为协同韵律模型. *计算机学报*, 2006, 29: 822–827]
 - 131 Wang Z Q, Gao W. A method to synthesize Chinese sign language based on virtual human technologies. *J Software*, 2002, 13: 2051–2056 [王兆其, 高文. 基于虚拟人合成技术的中国手语合成方法. *软件学报*, 2002, 13: 2051–2056]
 - 132 Zeng X Y, Lu P, Zhang M T, et al. A multi-modal game player interface system based on video and speech. *J Comput Aided Des Comput Graph*, 2005, 17: 2353–2358 [曾祥永, 鲁鹏, 张满园, 等. 基于视频与语音的多通道游戏用户界面系统. *计算机辅助设计与图形学学报*, 2005, 17: 2353–2358]
 - 133 Azuma R T. A survey of augmented reality. *J Presence: Teleop Virt Environ*, 1997, 6: 355–385
 - 134 Azuma R Y, Baillot R, Behringer S, et al. Recent advances in augmented reality. *IEEE Comput Graph Appl*, 2001, 21: 34–47
 - 135 Bell B, Höllerer T, Feiner S. An annotated situation-awareness aid for augmented reality. In: Proceedings of ACM Symposium on User Interface Software and Technology, Paris, 2002. 213–216
 - 136 Kiyokawa K, Kurata Y, Ohno H. An optical see-through display for mutual occlusion of real and virtual environments. In: Proceedings of International Symposium on Augmented Reality, Munich, 2000. 60–67
 - 137 Weiser M. The computer for the twenty-first century. *IEEE Pervas Comput*, 2002, 3: 19–25
 - 138 Ishii H, Ullmer B. Tangible bits: towards seamless interfaces between people, bits and atoms. In: Proceedings of ACM CHI, Atlanta, 1997. 234–241
 - 139 Ullmer B, Ishii H. Emerging frameworks for tangible user interfaces. *IBM Syst J*, 2000, 9: 915–931
 - 140 Kenneth P. Fishkin: a taxonomy for and analysis of tangible interfaces. *Pers Ubiquit Comput*, 2004, 8: 347–358
 - 141 Orit S, Nancy L, Eduardo H, et al. The TAC paradigm: specifying tangible user interfaces. *Pers Ubiquit Comput*, 2004, 8: 359–369
 - 142 Ullmer B, Ishii H. The metaDESK: models and prototypes for tangible user interfaces. In: Proceedings of ACM Symposium on User Interface Software and Technology, Alberta, 1997. 223–232
 - 143 Billingham M, Poupyrev I, Kato H, et al. Mixing realities in shared space: an augmented reality interface for collaborative computing. In: Proceedings of the IEEE International Conference on Multimedia and Expo, New York, 2000. 1641–1644
 - 144 Poupyrev I, Tan D S, Billingham M, et al. Tiles: a mixed reality authoring interface. In: Proceedings of 7th

- Conference on Human-Computer Interaction, Netherlands, 2001. 334–341
- 145 Kato H, Billinghurst M, Poupyrev I, et al. Tangible augmented reality for human computer interaction. *J Soc Art Sci*, 2002, 1: 97–104
- 146 Billinghurst M, Kato H, Poupyrev I. The magicbook: a transitional AR interface. *Comput Graph*, 2001, 25: 745–753

A survey on human-computer interaction in virtual reality

Fengjun ZHANG^{1,2*}, Guozhong DAI¹ & Xiaolan PENG¹

1 *Beijing Key Laboratory of Human-Computer Interaction, Institute of Software, Chinese Academy of Sciences, Beijing 100190, China;*

2 *State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, Beijing 100190, China*

*E-mail: fengjun@iscas.ac.cn

Abstract Human computer interaction is one of the core technologies of virtual reality, which has an important significance on promoting the widely use of virtual reality and improving the user experience. Due to the development of various sensors and hardwares, human computer interaction technology in virtual reality has made remarkable progress. This paper firstly introduces the paradigm of human computer interaction in virtual reality, and then makes a summary about the main research contents and development trends of various virtual reality and augmented reality technologies, including 3D interaction, hand gesture interaction, handheld devices interaction, speech interaction, haptic interaction and multimodal interaction. Finally, the paper proposes some existing problems that need to be further studied.

Keywords virtual reality, augmented reality, human-computer interaction, user interfaces, 3D, gesture recognition, haptic interfaces

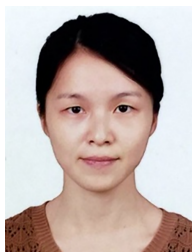


Academy of Sciences, China. His research interests include human-computer interaction and virtual reality.

Fengjun ZHANG was born in 1971. He received the B.S. degree from Yan-Shan University, China and obtained the M.S. degree from Harbin University of Science and Technology, China. He received the Ph.D. degree in mechanical manufacturing and automation from the Harbin Institute of Technology, China. He is currently an associate professor in the Intelligence Engineering Lab, Institute of Software, Chinese



Guozhong DAI was born in 1944, Ph.D. supervisor, professor. His main research interests include human-computer interaction and computer graphics. Email: dgz@iscas.ac.cn.



Xiaolan PENG was born in 1989. She received the B.S. and M.S. degrees in Human Robotic Interaction Lab from University of Science and Technology Beijing, China. She is currently an engineer in the Intelligence Engineering Lab, Institute of Software, Chinese Academy of Sciences, China. Her research interests include human-computer interaction and multimodal learning.