

增强虚拟现实交互的具身感

文 / 史元春

1 具身感

具身感 (sense of embodiment) 指人认为自己拥有、控制和处于一个身体的感觉, 这个身体可能是自己, 也可以是另外一个表示自己的虚拟对象。具身感起源于心理学中关于人的意识和身体关系的问题, 随着虚拟现实的普及和元宇宙的兴起, 使之成为一个需要研究的内容。

一个经典实验称之为“橡胶手错觉实验”, 1998 年发表在 Nature 的一篇论文中。实验很简单, 一块挡板的两边有两只手, 一只是被试者的真手, 另外一只是橡胶手, 实验员同时用毛刷刷两只手, 当刷橡胶手, 不刷真手时, 被试者也会有被刷的感觉; 甚至还做了比如拿东西去扎假橡胶手的实验时, 真手也会有被扎的感觉, 马上躲开。

这个心理学实验的一个有趣的结论是, 虽然橡胶手是物理的假手不能动, 也不逼真, 但是用户对这只假橡胶手产生与他自身相关的感觉, 现在称之为具身感的感觉。相比橡胶手, 今天虚拟现实的交互中, 提供用户的是一个可以同步其姿态、动作, 甚至是表情等信息丰富的虚拟化身, 因而自然而然产生一个问题, 虚拟化身对用户具身感会产生什么影响?

2 虚拟现实中的具身感

量变会引起质变, 以及今天技术的进步, 使虚拟现实三维显示能力大幅度提高, 已经可以给人一种逼真的具身感 (外在形象)。2020 年, 一位医

学博士 Spiegel B. 出版了一本专著 VRx:How Virtual Therapeutics Will Revolutionize Medicine, 描述了一种关于 VR 中具身感的体验。在一个完美 VR (也就是今天我们预期的元宇宙) 中, 可以看到自己的腿在数字桌面上跳舞, 因为运动的高度一致, 一瞬间可能分不清这条腿是不是我的, 这时就产生了一个疑问——这是我的腿吗? 无疑 VR 有效地证明了可以让用户产生具身感, 也就是你分不清可能打造某种具身感。目前的技术仍然在不断进步中, 从一开始只有守护的追踪、上肢运动, 到今天可以逼真的服饰和表情的渲染, VR 提供了越来越强的, 更像现实中的人和人运动控制的能力。

所以, 有人畅想在未来社会中一定会有一个数字世界, 用户可以完全沉浸其中, 对于整个虚拟世界和自己产生这种接近现实世界的具身感。在一段视频中只能看到人物的上半身, 这是因为我们的捕捉和感知技术还不能获得相关数据, 是否技术发展了就可以? 为什么我们还要讨论如何增强虚拟现实交互的具身感? 原因在于, 从人机交互的角度来考虑具身感时, 到底具身感在 VR 或者在元宇宙中应该如何存在? 交互对具身感的需求带来一些新问题的同时, 也会产生新的研究机遇, 也就是新的作用方式。

3 交互和增强

3.1 存在的问题

问题 1 虚拟现实人机交互对具身感需求如何?



从人机交互的角度似乎有一个被忽略的问题，我们真的需要完全物理真实的具身感吗？在具身感中，反馈是一个很重要的方面，身体姿态的控制，运动的反馈和显示，比如触觉的反馈，完美的再现是需要的，不同的交互任务对于具身感的要求是不同的。比如比较流行的 Beat Saber 虚拟现实游戏中，用户只需要看到双手持握的武器，并不需要看到双手就可以沉浸其中；相反，在社交应用中想和某人握手，需要虚拟化身具有尽可能细节化的外表。

问题 2 如何维持具身感同时增强交互？

如果能够提供最最高、最强的具身感，为什么还要考虑交互对具身感的需求？原因在于，在一个虚拟环境中，无论是元宇宙还是虚拟现实都是一个虚拟的环境，它有一个重构的过程。同样，具身感也需要重构，并一定程度上还会打破原来与生俱来的具身感，反而可以用来提高交互的效能。比如，通过重定向，在 VR 中可以重新定向用户的行进路径；即重构他对环境的具身感，而偏差的积累用户感觉不到，就可以在小的物理空间中，给用户在虚拟空间一个非常大的环境，甚至通过现实中圆形的路径，可以让用户感觉在始终走直线。在虚拟现实中可以解决有限空间的限制，提供物理动作，可以跑动、追逐、打闹，可以在一个大的场地中开展游戏。利用具身感的重构还可以优化用户的交互姿态和体验。比如，人长时间举着手臂很不舒服，利用姿态的变化和补偿，就可以在一个很舒适的体位上操控。这样问题变成了如何维持具身感，同时增加交互的效能，当然效能和具体的任务也有关系。

从我们开展的上述研究和相关的研究经验可以看到，可以通过可计算的方式，对交互需求和具身感进行联合优化。因为考虑具体的因素有很多，可以对两个维度计算求解，即具身感的呈现和交互任务之间有一个最优的备选集合，并且根据具体应用的交互需求来获得最优的交互方式。

3.2 研究进展

如上所述，我们关注的虚拟现实交互中具身感的建模与优化，就是如下两个基本问题。

（1）如何建立虚拟现实中具身感可计算的模型

$$\text{SoE} = f(I, S, \dots | C)$$

包括对虚拟化身、对虚拟的环境，以及先验认知的研究；也就是具身感与哪些因素有关，是否可计算。

（2）如何针对交互任务的复杂需求和具身感进行联合优化

$$\max_{I, S, C} g(\text{SoE}, \text{Per})$$

包括具身感和交互效能，综合考虑的最大化。

3.2.1 虚拟现实具身感的可计算模型的研究

首先是在影响具身感的因素方面相关工作的研究进展。虚拟化身和用户本身的外观是否一致对具身感的影响，这是最直接的、较直观的。在虚拟现实可以用动物甚至某个物体代表一个用户的化身。例如，有研究改变用户在虚拟现实中的肤色、长相等，对用户具身感的影响；还有研究与用户自身相对的环境，研究环境的变化对具身感的影响，比如研究虚拟环境与物理环境是否一致对具身感的影响；以及研究用户做出的行为在虚拟环境中的反馈，其中主要的是运动反馈。

运动反馈。目前在虚拟环境中完成精细和准确的交互任务比较难。我们从交互动作的重定向来研究具身感（见图 1），这也是近几年在 VR 和人机交互中的研究热点。相比于其他应用重构具身感带来的好处，我们希望建模运动偏差对具身感定量影响。这项研究不但可以理解和验证先前的工作，也为未来技术实践的界面设计提供数据和原理的指导。也就是，运动偏差重构出来的与实际运动偏差之间存在什么关系，利用这个关系在新的交互界面和交互任务的设计上做相应的优化。

（1）研究方法和结果

在虚拟现实环境中，通过高精度的动作捕捉



图 1 从交互动作的重定向来研究具身感

设备获取用户的物理动作，比如我们在实验室用 OpticTrack 动作捕捉系统就可以非常精细地采集一个 / 多个关节，甚至手指的微小运作，在此基础上添加转动偏移量，通过用户实验，研究偏移量大小、方向和相互作用对于具身感的影响，建立运动一致性与具身感的可计算模型。

通过多个实验，收集了多个用户的观测数据、统计偏差、运动偏差和具身感的分布关系，从而得到如下三个结论：① 偏差量的方向对于具身感没有显著的影响；② 父关节与子关节的偏差被发现的概率呈负线性的关系；③ 当运动幅度偏差大于 13° 时，用户在多数情况下会发现此偏差，进而显著降低具身感。

这些数据的分布可以用来指导不同的交互运动，根据具身感需求不同，添加运动偏差的域值；所采集的数据建立的计算模型可以给用户一段手臂的运动，满足交互应用对于沉浸感的要求。实验的结果最终要反馈到用户实验，通过运动偏差与具身感的计算模型，调整偏差的大小方向和分配，实现交互动作相应放大的技术。

实验结果显示，物理与虚拟动作完全一致的情况下可以提升 14% 的速度，减少 16% 的运动负载，利用这个偏差，使我们能够更快地达到运动交互的目的。这个计算模型对于高具身感、中具身感和低具身感三个要求的应用，试验用户可以容忍运动不一致的程度，同时用这种不一致完成相应的功能。

(2) 模型的应用

应用一 虚拟现实，高具身感的复健。现在有大量利用此技术做身体功能的建立或恢复。虚拟现实的应用对于用户沉浸感的要求非常高，因此我们人为或者有机地加入对身体运动的偏差。由于复健患者身体机能不如常人，做不好复健任务心里会很受打击，不愿意去做，而我们算法会让他在不注意的情况下，其虚拟身体呈现出一种更健康、移动也更快的合理“假象”，给他一个最大化复健的动力，使其康复锻炼可以坚持下去。

应用二 中具身感的 VR 游戏。用户在游戏中虚拟化身的具身感是中等强度要求的，通过让其虚拟身体变慢或变快可以增强或减少游戏的难度，甚至可以模拟游戏中受伤等更丰富的体验，使用户这种代入感和游戏效果都更好。

应用三 低具身感的面向界面排布。在 VR 一些任务中也会有界面排布，此任务中交互效率最优先。用户可以接受如实际物理运动中手臂移动，所以必须给他一个可以快速获取目标，而减少运动带来的疲劳等交互中常见的问题。

3.2.2 针对交互任务的复杂需求和具身感进行联合优化

虚拟现实中，除了对自身以外，用户对环境也存在具身感，也就是用户认为自己处于和可以与虚拟环境进行交互的感觉。目前很多工作也希望可以将逼真的环境提供给用户，来提升对环境的具身感。而目前的虚拟现实环境存在一个显著的问题，就是

用户深度感知不准的问题。

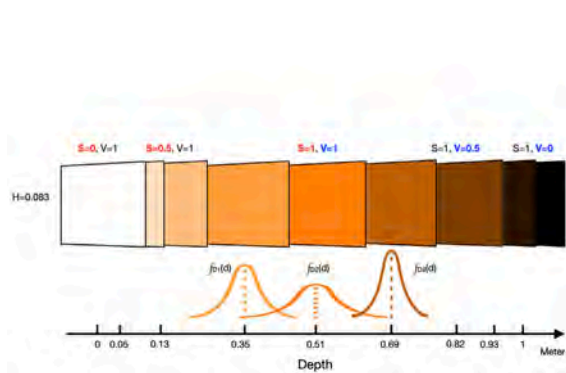
从视觉处理的过程我们得知，用户对环境的感知等都会受很多视觉线索的影响。例如，物体的相对大小（近大远小）；物体之间的遮挡及投影，如手对一个球体的遮挡，遮挡不好就“长”在手上；两只眼睛聚焦物体，观察角度不同，还会产生双目视差；自体感知，眼部聚焦物体产生的反馈，也就是在物理环境中，人获取信息有一个聚焦。现在的摄像机基本上可以通过 AR 技术对所看的图像进行一部分虚化，但是在 VR 环境建模中，如果把人类在物理世界中建立或者与生俱来的深度线索和感知模式完全迁移到虚拟世界中，还是比较困难的。基于视觉感知这些原理，对用户视觉控制和渲染能力创造一个额外线索，来帮助重建人的视觉感知。

这里主要采用深度，用目标颜色进行深度信息编码。目前颜色编码已经有很多成熟的应用，比如用不同的颜色表现温度情绪、用不同色相表现不同的深度等级。我们延续映射的思路。但是深度与颜色之间没有明确的大小关系，而是一种循环关系，如红橙黄绿青蓝紫可以一直循环下去，因此用户很

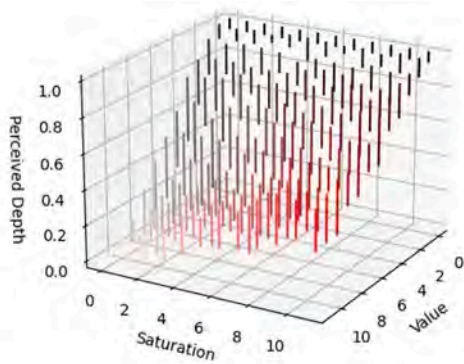
难形成一致的深度到色相映射。

我们在大量的基础实验中可以获得颜色的饱和度和透明度，来估计深度的能力。基于上述数据及规律，我们可以建立设计算法，综合表达这种能力，以获得用户感知的精度。感知精度是指用户会以多大的概率混淆相连的深度，在不混淆区间获得作为最优的颜色深度映射。在这个过程中，我们兼顾相邻颜色级别的混淆概率，以及表达深度的层级数量。为了保持灵活性和支持更多任务，允许具体的交互应用指定颜色范围，用户可以接触到混淆程度和颜色的变化，以这些作为输入，计算最优的颜色到深度的映射方式。

如图 2 所示建立一个数据采集实验。首先探索用户自然颜色深度的匹配关系，在相同色相上共建一个色度和饱和度的二维分布，用于计算用户对于不同颜色组合的混淆概率。所谓最优就是满足对混淆概率——要求用户有 95% 的概率，可以分清不同的颜色和深度。然后搜索一条路径，且路径的节点数最大（即区分的范围），同时以最多的节点数满足混淆概率。在这个选择中计算预期混淆，以获得概率最低这一条路径。其本质上也是一个搜索问题。



(a) 颜色编码在深度维度的混淆概率模型



(b) 深度编码在颜色两维平面的混淆模型

图 2 数据采集

从实验的视频展示中能更直观地看到不同映射和颜色起的作用。测试时以 3D 和 2D 图形作为目标，同时记录用户完整实验时间，包括完成绘制的图形，以及与目标图形之间的差别。经过颜色线索非常有效地提高了用户的输入控制精度，以及有没有颜色

的差异所表示的深度等非常细节的任务设计。最后，我们用可以搜索到的路径和颜色所表达的深度，建立了颜色映射和混淆度关系的路径，同时也做了大量的验证实验。

首先是手势复现。比如，空中手势可以用延伸

映射到深度的方式更加精准地完成 3D 手势的绘制。同时用颜色和深度的关系，用户可以方便观察进行操作，完成 3D 界面排布的任务。例如驾驶辅助中的障碍提示，也是一个很有效的映射关系。

对于上述两部分的研究，还有一些基础的方法和模型的建立都有助于开展更深度的工作。

4 总结与展望

具身感本来是人与生俱来（first nature）的，我们自然而然地接受身体和自己的控制，这是生物学本性。从历史上看，First nature 是很少或者根本没有控制的东西；而一个重要的观察——对于虚拟现实来说，具身感其实是可以重构的、可以习得的第二本性（second nature），对于很多交互任务来说，具身感不用做到完美，用户可以适应不完美的具身感。

这其实也给我们提供很大的研究空间，以利用重构具身感的机会优化用户与系统的交互过程，实现不同的交互优化。对于很多交互任务，具身感不用做到 first nature 的完美，而是用户在不完美的具身感，甚至可以改变自己。第二天性在人机交互中叫做习惯成自然，也就是习得的，我们拥有一定程度的自主控制，甚至被带动的控制，它给我们提供了很大的研究空间，在虚拟现实中能重构具身感和优化人机交互。

不同的人机交互优化，一直在围绕所谓重构和增强来开展。下面列举一个非常有意义的研究工作。

虚拟现实甚至元宇宙不仅是一个沉浸式的观察，还有交互，一个最基本的交互任务是我们访问空间的对象，也就是用户看到什么就选取什么，即目标点取行为速度要快，否则对用户的交互体验和交互效果有很大的损失。首先人的视觉运动范围有限，现在的 VR 环境中要求用户看清目标才能点取，如果人的头部（头盔）运动就要改变虚拟环境，用户会产生很强的不适感（晕）；而头运动和人的视觉获取和呈现体验之间还不能精准计算匹配，并且人与人之间也有差异，所以这种频繁运动造成的眩晕

是一个很大问题。此外，人的交互过程中，如果看清再去点有一个移动时间，准确率也不高，也会产生眩晕感。

实际上，人眼睛的余光不需要看得很准，肢体感知会带给人一定范围操作和访问的准确性，因此我们研究人无需视觉定睛去看某个对象，可以访问到的一个范围。但是在这个过程中人运动有一定偏差，为此建立用户动作控制偏差的量化模型，首创的无需视觉注意的空中目标选取技术，支持用户不需要凝视某个对象就可以精准访问，从而提高了用户目标选取的准确度和速度，降低了眩晕感。

无需视觉注意的空中目标选取技术，也就是用户控制运动偏差的量化模型可以在 AR/VR、大屏交互进行偏差的补偿和布局的优化。由于它不是在三维空间中准确重构，而是有了偏差的一个重构，所以可以更好改善用户点取的速度和眩晕感。头盔上的实验结果与微软的 Hololens 进行了对比，我们基本的抓取动作速度提高 30% 以上。

未来，具身感将在环境、自身和反馈上进行深入研究，因为它们都有很多的优化空间。在交互中最基础的原子操作就是对象访问，用户需要准确知道访问对象的空间位置、外观形态、对象的语义才能明确表达交互意图和交互行为。用户为了表达交互意图，需要明确了解自身在虚拟现实中的肢体动作、运动能力和交互工具，才能做出合适的交互命令。之后还要迅速观察和反馈效果，认知到其交互行为是否准确完成。交互命令是否达成对后续的交互继续来做这个过程是一个闭环，每一步都有重要的影响，因而需要有良好的具身感支持，否则在对虚拟环境认知交互效率和体验上都不好。

再回到最初的一个基本问题——具身感的测量。具身感怎么感知和测量高、中、低，是一个很交叉的研究。我们看到有的研究方法是用脑电测量，但是它只是相关的大脑区域，不能明确知道这个信号的语义。另外一些研究更接近人的生理表现，比如

（下转第 46 页）

（上接第 40 页）

瞳孔直径反映用户主观感受的变化，但是这些生理信号受干扰的因素，控制实验比较难做。目前最常用的方法就是做主观量表，实验时都是现场访谈和观察，以及靠有经验的专家，这种方法需要更自动化实时观测，即在交互闭环环境中用测量和观察的技术作为依据和改善。

在上述研究的基础上具有生理感知具身感的感

知能力，可以用来实现面向交互任务自适应的虚拟化身；在此基础上，我们还结合了面向多人交互场景具身感的优化和人机物三元融合的具身交互，才能够支撑更广泛的面向未来的虚拟现实，尤其是面向个人体验很好的应用场景，提供至关重要的人机交互能力，不断优化用户的交互体验和交互效率。

（参考文献略）



史元春

青海大学校长，清华大学人工智能研究院智能人机交互研究中心主任、教授，CAAI/CCF Fellow。主持多项国家十三五重点专项等国家研发计划项目。主要从事人机交互、普适计算、多媒体、网络教育技术等相关领域研究。科研成果近年连续获得国际人机交互领域顶级会议最佳论文奖，并两次获得国家科技进步奖。