中国科学: 信息科学 2016 年 第 46 卷 第 12 期: 1737–1747
SCIENTIA SINICA Informationis
《中国科学》杂志社
SCIENCE CHINA PRESS
虚拟现实纵览专刊 · 评述

混合现实中的虚实融合与人机智能交融

陈宝权*, 秦学英

山东大学计算机科学与技术学院, 济南 250101
* 通信作者. E-mail: baoquan@sdu.edu.cn

收稿日期: 2016–10–20; 接受日期: 2016–11–27; 网络出版日期: 2016–12–22
国家重点基础研究发展计划 (973 计划) (批准号: 2015CB352500) 和国家自然科学基金 (批准号: 61672326, 61173070) 资助项目

摘要 混合现实技术与虚拟现实技术的进展, 受到全球各界的广泛关注, 并迅速引发了在教育、医疗、游戏等各个领域对未来的展望. 混合现实旨在将虚拟世界与现实世界融为一体, 由于其视觉景象是视点依赖的, 因此穿戴式头盔显示器提供了可沉浸式观察混合现实场景的基础工具, 并由计算机提供与用户观察适配的虚实融合景象. 然而, 虚实融合景象的生成, 本质上是相异时空场景的相互嵌入, 依赖于空间几何与光照环境的共享与相互作用, 即几何一致性与光照一致性; 在社会学意义上是行为规范的融合, 需要符合人类社会学与心理学的规律, 即行为一致性. 混合现实提供了用户与虚拟世界交互的自然界面, 通过对人类动作和行为的理解, 在现实世界的时空中搭建用户与虚拟世界的桥梁, 在客观上具备对虚拟世界与现实世界联结的直观性. 在此虚实混合的世界中, 人类智能可以通过对虚实混合场景的观察来理解世界, 通过自然交互按需驱动机器对虚拟世界施加影响, 并临场获得计算机对场景或者数据的反馈, 从而实现人类与计算机的沉浸式深度交互, 进而实现机器智能与人类智能的深度融合. 从非穿戴式到穿戴式设备实现了混合现实的沉浸感, 从静态场景到有人类活动的场景实现了混合现实的社会化属性, 而虚拟世界从具体场景扩展到大数据时代的数据可视化分析语义, 则使混合现实成为人类智能与机器智能交互融合的平台.

关键词 混合现实 虚实融合 社会化融合 智能融合

1 引言

虚拟现实试图将人类与计算机营造的世界直接联结, 使得人类可以在虚拟世界中遨游, 获得与现实世界相同的知觉感受. 而混合现实 (mixed reality) 技术 [1], 则试图将人与虚拟世界及现实世界三者同时联结起来, 使得虚拟世界与现实世界发生直接联系, 从而在更大程度上影响世界. 由于混合现实也建立在人类自然知觉感知的基础之上, 因此本质上为所见即所得的人机交互界面 [2]. 该界面将人类从复杂深奥的计算机用户界面中解放出来, 越过繁琐的菜单和参数选择, 回到人类的原始感官通道, 使人可以直观地理解世界. 因此, 混合现实建立了用户与现实世界中的虚拟世界之间的直接通道, 计算机营造的虚拟世界与现实世界在这里自然交汇.

引用格式: 陈宝权, 秦学英. 混合现实中的虚实融合与人机智能交融. 中国科学: 信息科学, 2016, 46: 1737–1747, doi: 10.1360/N112016-00249

近两年, 世界顶级公司发布的产品引发了对虚拟现实技术和混合现实技术的广泛期待, 并在产业界引发了一轮巨大的创业浪潮. Oculus Rift、HTC Vive 的虚拟现实头盔、微软的 Hololens 增强现实头盔及其各种炫酷的演示系统的发布, 引发了无数对于未来技术发展的憧憬. 新一代头盔显示器由于较好地解决了定标的精度和稳定性, 使用户不再感到明显的眩晕, 从而奠定了广泛应用的基础, 但是, 这并非意味着混合现实技术的成熟. 事实上, 混合现实的发展仍然处于初级阶段. 头盔显示器是可穿戴设备, 可以提供混合现实所需的沉浸感, 为混合现实的基础性设备, 但目前还有很多限制, 包括移动范围过小、视野过窄等等. 更重要的是, 头盔显示器中的虚实融合内容的制作与交互, 还潜伏着更多更大的挑战. 由于人类感官已经发展成为高度精密的系统, 模拟人类知觉的人造系统的精度和广度远未超越. 从技术上说, 尽管混合现实技术前景灿烂, 其广泛的实用化尚需假以时日, 有待技术的全面突破.

混合现实涵盖了增强现实 (augmented reality) 与增强虚拟 (augmented virtuality) [1], 并可在此基础上扩展到遥在 (tele-presence) 和现实景物消去 (diminished reality) 等技术. 混合现实式的遥在是将远程的现实直接虚拟化, 通过网络传输与现实场景融合, 成为多时空现实的融合, 例如微软目前发布的 Holoportation; 景物消去是将现实场景中的物体隐匿. 在这样的环境中, 现实通过实时重建实现虚拟化, 通过网络传输及与现实的融合, 完成与现实的混合; 或者通过对现实的虚拟化, 实现景物消去. 因此, 虚实融合可以涵盖丰富的内容.

混合现实由现实与虚拟两部分构成, 其中虚拟部分关心用户与虚拟世界的联结, 因此涉及两方面的内容: 虚拟世界的构建与呈现, 以及人与虚拟世界的交互. 由于呈现的虚拟世界是与人类感官直接联结的, 因此, 完美的虚拟世界的营造是通过建立与人类感官匹配的自然通道来实现的, 通过真实感渲染呈现虚拟世界, 营造音响效果, 提供触觉、力觉等各种知觉感知和反馈. 因此, 用户与虚拟世界的交互必须要建立相同的知觉通道, 通过对用户的自然行为分析, 形成感知、理解、响应、呈现的环路. 这是虚拟现实技术的核心内容. 混合现实则省却了对复杂多变的现实世界进行实时模拟, 因为对现实世界的模拟本身是非常困难的, 取而代之的是需要建立虚拟世界与现实世界的联结并模拟二者的相互影响. 然而, 要使虚拟世界与现实世界融为一体, 在技术上形成了诸多挑战, 不仅要感知用户的主体行为, 还需要感知一切现实世界中有关联的人、环境甚至事件语义, 才能提供恰当的交互和反馈. 因此, 混合现实涉及到广泛的学科, 从计算机视觉、计算机图形学、模式识别到光学、电子、材料等多个学科领域. 然而, 正是由于混合现实与现实世界的紧密联系, 才使其具备强大且广泛的实用价值.

混合现实技术是虚拟世界与现实世界无缝融合的技术, 虚拟现实代表了计算机营造的世界, 使人类的知觉感知延展到计算机中; 而混合现实技术则是在保持对现实世界正常感知的基础之上, 通过建立虚拟世界与现实世界之间的联系, 再将人类感官延伸到虚拟世界. 混合现实技术中所关注的虚拟世界可以有丰富的内容. 从早期的虚拟现实世界的局部场景, 与现实世界无缝融合, 使得我们可以看到匪夷所思的场景, 典型的是电影《阿凡达》呈现的世界. 然而, 计算机的强大能力不仅在于对场景的营造能力, 还在于对信息搜集、数据整理和分析呈现的能力. 在信息爆炸的时代, 信息容量和复杂度远远超过人类所能够掌控的范围, 在宏观上把握信息的内涵, 提供对数据蕴涵的语义分析, 才有可能使人类理解数据. 混合现实技术可以在数据分析的基础上建立用户与数据的联结, 从而使得用户可以直接感知数据分析的结果, 将人类感知延展到数据语义层面.

混合现实技术由于涵盖了虚拟世界与现实世界, 既需要虚拟现实技术的支持, 也需要增强现实技术的支持. 虚拟现实技术的第一个核心问题是对虚拟世界的建模, 一般包括模拟现实世界的模型或者人工设计的模型. 对现实世界模型的模拟, 即场景重建技术. 虚拟现实的第二个问题是将观察者知觉与虚拟世界的空间注册, 满足视觉沉浸感的呈现技术; 第三个问题是提供与人类感知通道一致的交互技术, 即感知和反馈技术. 增强现实技术在虚拟现实技术的基础上, 还需要将现实世界与虚拟世界进

行注册, 并且感知真实世界发生的状况、动态, 搜集真实世界的数据, 进行数据分析和语义分析, 并对其进行响应. 因此, 我们将混合现实的虚实融合分为 3 个层面: (1) 虚实空间产生视觉上的交互影响, 例如遮挡、光照、运动等; (2) 虚实世界产生社会学意义上的交互融合, 例如行人互相避让的行为; (3) 虚实世界产生智能上的交互融合.

2 可视空间的虚实融合

混合现实技术是虚实空间的相互嵌入和交互. 由于现实空间在物理上不仅包含了三维欧氏空间, 还包含了这个空间中的几何、光照、材质、物体运动等各种可视信息. 当虚实空间融合的时候, 需要满足该物理空间的规律. 空间注册技术将不同的空间包括观察者坐标系映射到统一的坐标标架下, 在此基础上, 通过有限重建现实场景, 生成虚实空间的相互作用效果, 包括遮挡处理、光照处理、实时渲染、动态物体运动及交互效果等.

2.1 场景构建

场景构建是虚拟现实与增强现实的共性基础技术, 一般包括三维建模和场景重建两种形式. 构建现实世界不存在的物体一般采用建模软件, 需要借助设计者的专业技能, 尽管有 Maya, 3D Max 等专业软件仍然是费时费力的. 场景重建是将现实世界中的场景在虚拟世界中复制其几何模型和材质, 通过渲染技术呈现其视觉效果, 因此是现实场景的再现. 作为关注主体对象的物体, 需要高度精确重构现实场景, 这对于大规模自然场景尤其是动态场景具有极大的挑战性. 增强现实技术则试图回避对现实场景的建模问题, 现实世界可以直接观察到, 因此省却了对现实世界的建模和呈现的过程. 然而, 由于虚拟景物会与现实场景产生空间上的交互, 为了避免空间上的冲突, 使虚拟景物具有合理的行为, 例如: 轿车行使在道路上、行人避开障碍物等, 往往仅需要有限重建局部场景, 以提供虚拟世界与真实世界交互所需的几何结构, 在精度需求上往往有较大的差异.

在混合现实中作为直接呈现对象的虚拟场景, 其准确性是非常重要的. 场景重建的主要手段有两种, 一种是主动式如激光扫描仪, 一种是被动式如多视角几何重建. 多视角几何重建 [3] 是基于视觉的方法, 通过特征匹配形成形状约束, 确定几何表面的形状. 因此, 其精度是由像素精度决定的, 并受制于特征的显著性. 激光扫描仪是工业标准的重建工具, 比基于视觉的重建方式具有更高的精度, 并且鲁棒稳定, 但是也受制于物体的材质, 例如难以采集高光反射物体的形状. 因此, 在大规模数据采集的过程中, 一般采用激光扫描仪. 车载扫描仪由于其可移动性, 成为大规模景观重建的主要方法. 然而, 激光扫描仪作为一种主动式深度测量仪器, 尽管每一个测量点的精度很高, 但其点云的密度却在某些局部非常稀疏, 并且由于光线的直线传播特性, 无法测量被遮挡的区域, 往往存在数据缺失. 因此, 由激光扫描仪获取的点云数据到重建场景的几何形体, 尚有不小的距离, 几何处理就变得非常重要. 要补足缺失的数据, 重新采集数据是不现实的, 因此需要另辟蹊径. 重构场景对象通常有自身的特点, 例如树木、建筑等等. 使用景物对象本身具有的几何规则或者先验信息作为缺失数据重建的约束条件, 可以在不予补充数据的条件下高精度地重建场景 [4,5], 其中参考文献 [5] 的图 1 展示了具有重复性结构高楼的激光点云重建实例, 尽管原始采样数据在很大程度上存在数据的缺失, 但是因为楼层间三维模型的相似性, 通过层间相似性的迁移产生的约束, 完整重构所有的楼层.

激光扫描重建场景或者物体仍然是一项非常繁琐的工作. 一些景物仅仅援用对称性、重复性等仍然无法很好地重建, 因此需要从不同的角度反复扫描, 以获得足够精度的物体三维模型. 为了规模化

地采集三维模型, 自动地实现模型的高度精确的三维扫描才是解决方案. 由于机器人技术的进步, 借助机器人的行动能力和计算机对几何的感知分析能力, 在机器人手臂操作手持式扫描仪采集三维物体点云数据的同时, 判断点云数据的密度是否充分, 并自动计算下一步的最佳扫描角度, 并反馈给机器人改换扫描姿态. 通过这个过程, 实现物体的自动扫描, 并能够使扫描仪以最优化的方式完成扫描, 文献 [6] 的图 1 是由机器人 Pro 扫描获取的复杂三维模型的示例.

现实场景往往高度复杂, 由各种各样的物体共同构成. 对场景的分析和解析, 也是准确扫描场景三维物体的必要条件. 例如: 当场景中有物体与其他物体接触时, 智能机器人通过分析和试探, 将相连的物体分开, 从而完成对物体的完整扫描. 这对于物体建模的完整性是非常重要的. 由于扫描仪不能选择扫描对象, 而是采集视域内的所有数据, 因此通过几何分析的方法, 产生几何的分割从而生成具有独立性的物体, 是几何重建的重要一步 [6], 例如: 文献 [6] 的图 1 呈现了用 Kinect 采集的室内场景的点云数据重构的场景, 不同的颜色标记了不同的物体. 由于数据采集时采用了精度较低的 Kinect, 使得重建景物的精度有所下降, 但是算法实现了对场景物体的分割与重建. 自动生成的物体的重建 [7] 对于场景的分析以及场景重组具有重要意义, 如文献 [7] 的图 1 所示, 由外形各不相同但风格相似的建筑物, 组成一个巨大的建筑群落. 由于场景的构成存在规律性, 当建立了场景规则之后, 便可以根据规则重新组成新的三维模型, 使得大规模的非重复三维场景重构成为可能.

场景重建至今仍然是重要的课题. 由于场景重建的日益成熟, 虚拟场景的构建已经颇为便捷. 混合现实中对于场景的处理需求, 需要不断地与重建场景发生不同程度的密切关系, 因此是混合现实的重要基础. 由于机器智能的引入, 通过对场景几何的分析, 使得场景的拆分和重组成为可能, 从而引导了新的研究方向.

2.2 场景注册

场景注册是混合现实中非常重要的一个环节, 旨在确定虚拟空间与现实空间之间坐标标架的映射关系. 在沉浸式虚拟现实环境中, 用户处于现实世界中, 需要感知用户的观察方位的变化, 才能为观察者呈现相应视域的景象, 因此往往需要标定观察者头部姿态. 在混合现实环境中, 观察者的头部姿态跟踪更为重要. 为了获得完美的沉浸感, 头盔显示器的跟踪定位高精度、低时延是虚实空间一致性的保障, 能使用户感觉到虚拟物体如同嵌入在现实空间一样. HTC Vive 与 Hololens 的相继发布, 标志着虚拟现实和增强现实头盔在定位部分已经有比较成熟的技术, 能在特殊环境中获得良好的沉浸感体验. 目前的主要限制在于对移动性和环境的制约.

场景注册是非常具有挑战性的技术, 目前比较成功的商品一般采用激光、陀螺仪与加速度计等硬件或主动光学模式. 一般增强现实头盔分为视频式和光学穿透式, 对视频式头盔显示器而言, 其头盔的定位方式一般采用视觉方式定位, 通过对画面的分析来感知头盔的姿态运动参数. 由于视频还提供了场景中丰富的信息, 因此, 使增强现实的虚实融合可以通过视频分析来实现, 这样摄像头就成为一个非常好的传感器, 并且具备廉价便捷的特点. 基于视频的场景注册技术从通过预先计算特征点三维空间位置 (SfM) 的摄像机定标技术 [3,8,9], 逐渐发展到在线重建空间并定位的技术, 例如同时定位与地图构建 (SLAM) 技术 [10], 但基于视觉的技术受困于场景画面特征消失时的算法失效. 由于深度视频传感器 (例如 Kinect) 的出现, 极大地提高了视觉传感器定位技术的鲁棒性, 使得基于视觉的场景注册技术正在走向成熟.

场景注册技术还包括采用场景中的三维刚体来进行姿态估计, 通过匹配算法计算三维物体与摄像机之间的相对姿态参数, 间接实现对场景的注册. 这种场景注册技术通常在已知三维物体的几何模型时采用 [11], 特别当该物体是混合现实技术中关注的主体对象时尤其如此. 但是, 基于视觉的技术需要

分步实现, 在计算机视觉的初始状态, 需要确认兴趣物体是否在画面中出现以及在画面中的位置, 才能对其进行姿态的估计. 这个过程如果采用计算机自动实现, 需要采用目标检测和识别技术, 来确定目标的初始位置. 一般来说, 目标检测和识别也将提供兴趣物体的初始姿态. 然后, 需要通过对物体的特征匹配, 搜索到兴趣物体与相机间的相对姿态参数, 实现场景注册的目标. 三维注册的实现, 有两个方面的作用: 如果三维物体是静止的, 那么其相对姿态参数就是现实场景与摄像机之间的空间变化参数; 如果三维物体是运动的, 那么也定位了二者之间的相对位置关系, 可以使用该物体的运动驱动虚拟物体的运动, 产生现实与虚拟场景间的交互.

场景注册的时效性非常重要. 基于视觉的场景注册, 需要解决定位精度、定位速度和时间延迟三大障碍. 混合现实作为虚拟世界与现实世界交汇的环境, 由于混合现实的实时性要求, 即使精确定位并且系统的每一个在线环节都达到实时, 仍然不能保证虚实空间的配准. 由于混合现实系统计算性能的局限, 一般经过定标、渲染、融合等环节呈现结果的时候, 已经与事件发生的时刻产生了时间差异, 形成时间延迟. 当时间延迟达到一定程度, 不仅导致混合现实环境中的虚拟物体与现实世界失配, 并且常常引发用户的眩晕感. 当头部姿态急剧变化时, 空间失配尤其明显. 因此, 混合现实系统需要保持非常低的时间延迟, 这往往成为混合现实系统的技术瓶颈, 因为这需要将定标等算法时间控制在数毫秒之内.

场景注册作为混合现实的核心技术而备受关注. 相对而言, 场景注册技术是混合现实中比较成熟的技术. 尽管如此, 这也仅仅是混合现实技术的起步阶段. 混合现实天然地将现实世界与虚拟世界联系在一起, 由于客观世界的复杂性, 使得计算机生成的虚拟世界与现实世界的融合, 需要符合现实世界的规范, 因而极具挑战性.

2.3 高度真实感的虚实融合

混合现实中虚拟世界与现实世界的交互影响, 本质上是虚拟空间与现实空间融为一体时产生的. 场景注册解决了虚拟空间的坐标标架与观察者空间标架的转换关系, 但是并不能保证虚拟景物在现实场景中的合理性. 当虚拟空间的景物要嵌入现实空间中时, 虚拟物体需要避免与现实空间冲突, 共享现实空间的光照环境, 产生相互遮挡, 产生作用力与反作用力, 等等. 因此, 在混合现实中, 需要不断地协调虚拟世界与现实世界之间的关系, 保持虚实场景之间的和谐共存. 本质上, 可以将现实世界看做由几何、材质、光照等静态或者动态复杂物体构成的可视空间; 虚实场景的融合技术, 则是虚拟物体在该可视空间的嵌入.

基于视频的混合现实技术是采用摄像头将真实场景记录下来, 同时将虚拟景物融入视频画面. 在混合现实技术中, 由于其实时性要求, 虚拟景物还远不能达到照片品质, 而可见性与光照一致性直接关系到虚拟场景融入视频场景的真实感. 为了实现视觉上的真实感, 混合现实需要在空间保持虚实物体之间的几何一致性、光照一致性与合成一致性 [12]. 几何一致性的保持指虚拟物体的位置、大小、透视关系与视频图像序列保持一致; 光照一致是指虚拟物体的光源及光照模型与视频图像序列保持一致; 合成一致是指虚拟物体对视频图像序列的影响要与实际情况保持一致, 比如, 虚拟物体对视频图像序列中的景物投射阴影、在水面上形成倒影等. 几何一致要求对视频图像序列精确定标或者空间注册; 光照一致要求对虚拟物体按视频图像中的光照明条件进行图形绘制; 合成一致要求融入虚拟物体的情况下, 对视频图像序列进行再绘制. 为了获得具有真实感的混合现实环境, 需要虚拟物体融入视频场景中的可见性计算、视频场景的自动光照环境重建、局部几何重建等技术.

几何一致性的保持, 除了通过场景注册和场景重建等技术, 将虚实世界对齐以外, 还需要处理现实景物对虚拟物体的遮挡, 否则, 虚实世界的融合感将被破坏. 空间中场景动态引发的对于感知的需求

是非常具有挑战性的, 特别是对遮挡处理的挑战性. 目前遮挡处理一般仅限于静止的规则前景产生的遮挡, 例如: 桌面产生的遮挡. 动态场景, 例如行人产生的遮挡, 则是非常困难的, 场景重建技术很难发挥作用, 因为动态行人的重建非常困难. 仔细思考遮挡处理的挑战可以发现, 遮挡处理其实往往不需要前景的完整几何, 而仅仅需要判断现实景物是否对虚拟物体构成遮挡; 如果形成遮挡, 那么也只需要前景物体的侧影轮廓线. 因此, 从精度要求来说, 遮挡处理需要较为粗糙的深度估计, 但是要求非常精致的侧影轮廓线处理. 由于混合现实是一项在线实时的技术, 因此, 采用手工交互的方式无法满足要求, 需要自动实时地从背景视频中分离出前景物体. 由于前景物体与背景缺乏普适的特征区分, 因此需要复杂的智能技术来分离前背景 [13]. 增强现实中的遮挡问题一直是非常困难的, 即使深度视频的出现使得遮挡问题的处理似乎变得更加容易了, 但是, 由于在侧影轮廓线附近, 深度视频常常存在数据缺失, 使得无法产生精确的轮廓线. 当然, 深度视频的出现使得遮挡关系的确立变得直接, 并且可以更好地确定前景与背景之间的颜色模型. 彻底解决混合现实中的遮挡问题尚待时日.

光照一致性是增强现实中产生融入感的必要条件, 牵涉到光照环境的一致性和高度真实感的渲染技术. 虚拟物体需要采用真实环境的光照来进行渲染, 因此用 Debevec 球采集光照环境的 HDR 映照图, 是一个非常简便的方法. 但是, 采集光照环境本身是一种破坏性的操作, 在被关注的区域留下明显的人为标志, 而且一旦被关注的物体是运动的, Debevec 球的光照采集便不具操作性. 在晴天的室外场景, Debevec 球难以精确捕捉阳光强度和角度. 虚拟物体本身的运动和光照环境的变化, 使得对光照环境的重建变得很重要. 一般的室内场景光照环境比较固定, 而在室外场景中, 变化的因素虽然仅仅取决阳光的强度和天空光的分布, 但是光照条件会时常发生变化. 这样通过在线实时的传感器, 随时随地获取光照环境的变化就非常重要. 事实上, 可以通过摄像头捕获影像的分析来感知户外场景的光照变化, 因为场景的光影是由场景几何、材质和光照环境共同作用的结果. 从影像中逆向求解光照环境, 通过对长时的场景影像的分析, 自动感知到环境光照的变化, 使得虚拟景物能够自动与背景融洽. 由于视频场景的虚实融合均为基于图像的参数估计方法, 光照环境的变化只需要恢复光照的分布, 而与光照的绝对值没有直接关系, 因此采用基于图像的方式可以感知准确的光照参数变化 [14, 15]. 即使恢复了光照环境, 也依赖于实时真实感渲染技术, 才能逼真呈现融洽的物体的光影变化. 更加精细的光影变化, 体现在虚拟物体与现实场景间的阴影投射和相互反射的变化, 由于光影的产生与几何、材质和光照均有密切关联, 因此, 仅仅依赖影像计算来产生虚实物体间的正确光影修正仍然是非常困难的, 依赖于一定的几何重建和材质估计.

混合现实有一个很特殊的需求, 就是虚拟景物需要跟现实世界共享统一空间, 因此, 作为虚拟物体的景物与现实场景中的景物之间的作用需要符合自然法则, 并避免空间冲突和视觉冲突, 例如: 杯子不能刺入桌面而是置放在平面上, 球掉在地面上会反弹, 等等. 构建真实场景的模型, 能很好地辅助实现虚实空间的协调融合, 但是其建模精度往往需要因地制宜. 一个虚拟皮球从空中掉落, 碰撞到地面后会反弹, 因此需要重建地面方程以及估计地面材质系数, 以产生逼真的反弹音效和反弹的角度和高度. 这是增强现实中场景建模的特点. 虚实融合的一个特例是两个实拍场景的无缝融合 [16]. 一般来说, 总是需要选择一个现实场景作为虚拟场景, 通过将其虚拟化, 融入现实场景. 微软的 Holoportation 则是采用三维重建的策略, 用 Kinect 重建虚拟化远程用户, 再融合到现实场景中, 因此获得与远程用户面对面交流的体验.

3 社会空间的虚实融合与交互

混合现实环境中, 会在两个方面涉及到人类的社会属性, 一方面, 用户是现实的人类; 另一方面,

虚拟物体可能是虚拟人群.

3.1 混合现实环境中的虚拟人群模拟

混合现实中更为高层的冲突, 来源于虚实空间中人类行为之间的行为交互. 当虚拟角色在现实空间运动时, 不仅需要满足可视空间虚实融合的要求, 适应视频场景环境的约束条件, 还需要使其行为符合社会规范, 满足人类行为的心理范式, 其姿态和行为必须适应视频场景环境的约束条件, 例如自动躲避行人、车辆以及障碍物. 在有人类活动的场景中, 虚实物体间还需要保持行为的一致性, 产生虚拟角色间的行为避让, 因此是社会空间的虚实融合.

虚拟人群或者车流模拟近年来有长足进展, 然而一般的模拟均在虚拟现实环境中展开. 混合现实中, 将虚拟人群嵌入现实世界, 其动态行为会与现实场景产生交互行为 [17]. 但在混合现实环境中, 现实场景一般是难以改变的, 因此, 虚拟人群的行为需要考虑到与现实世界的冲突与协调, 适应虚拟环境中的物体、场景或情节, 其运动需要满足基本的物理规律和自洽性, 从而产生对现实世界环境状况的感知需求. 对现实世界的状态感知因状况的不同会有较大的差异, 很难有固定的设置. 例如: 如果虚拟物体是一群虚拟人, 那么虚拟人应该行走在路面上, 并且应该避免障碍物而产生自动的路径规划; 如果一群真实人群与虚拟人群相遇, 那么虚拟人群需要动态地避开真实人群. 在这样的情节中, 由于真实人群一般不能观察到虚拟人群 (除非佩戴增强现实式头盔) , 因此这种避让不是对等的, 而是由虚拟人群单方面执行. 这样的情节设定不仅需要重建现实场景中的三维结构, 实现虚拟人群行走和避障所需的基础数据, 还需要感知到真实人群所在的三维空间位置及其变化. 这种行为的交互大规模的发生时, 需要由计算机自动对大规模虚拟人群在视频场景中的动态约束行为建模. 此外, 自适应地构建视频场景与虚拟化身人群的空间位置关系, 实现虚拟化身人群在真实视频场景中的动态约束下的大规模人群行为建模.

对人群等智能对象的驱动是非常复杂的过程, 其与现实场景和真实人群交互的过程, 不仅需要解决空间冲突的问题, 还需要考虑人类的心理感受和社会规范. 在这个过程中, 对现实世界的感知需要扩大到更为宽泛的层面, 可能需要对现实世界全方位的感知才能实现, 包括对人类心理特征、行为习惯以及社会规范的建模与感知. 客观地说, 这一点不仅对混合现实是重要的, 对虚拟现实中人群行为的模拟也是非常重要的. 从这个意义上, 混合现实技术提供了一个研究人类行为心理与社会规范的环境, 也是混合现实技术中具有巨大挑战性的任务.

3.2 混合现实环境中的交互

人类与计算机营造的虚拟场景之间存在巨大的鸿沟. 在混合现实这样的环境中, 试图通过虚拟与现实的融合构建人类知觉与计算机间的直接关联, 因此传统的人类与计算机之间的交互方式难以适用, 需要采用自然的人机交互方式.

自然的人机交互是虚拟现实的核心技术, 通过对人类自然语言、肢体语义等的解读, 由计算机计算出虚拟景物的变化, 做出恰当的响应. 这里涉及到两个方面, 既要对现实行为进行感知, 又要根据感知结果做出响应. 交互既体现在与景物或者信息之间的交互, 也体现在用户或者用户群与虚拟场景间的互相理解和反馈. 当用户与虚拟人群交互时, 用户的意图感知与虚拟人群的响应, 具有更为强烈的社会属性.

自然人机交互技术是现实与虚拟的通道. 感知人类的交互意图是人机交互中非常重要的一环. 早期一般通过虚拟手套等各种接触式的交互工具来实现自然人机交互, 如虚拟现实手套等. 采用基于视

频的方法识别人体的肢体语言, 毋须接触用户, 也是非常重要的发展方向, 但是因为对于背景运动非常敏感, 很容易受到干扰. 由于深度视频采集设备逐渐普及, 极大地提高了行为识别的准确率和鲁棒性. 基于深度视频骨架的动作识别 [18], 变得越来越准确, 从而成为动作识别, 以至人机交互的新工具. 例如: 基于 Kinect 的行为识别或者 Leap Motion 的手势识别方法. 由于骨架的准确性, 根据骨架的信息来确定较为精细的动作姿态参数, 已经比较成功, 广泛地用在游戏中. 但是, 这些交互工具还受到深度视频采集环境的要求限制, 难以在较大的环境和室外场景中使用. 而且, 这些动作一般仅仅包含了非常简单基础的语义, 事实上, 行为识别还远远不能达到自然人机交互的境界. 人类的肢体语言含义丰富, 却难以将肢体语言与某一类动作进行严格的划分, 同一种语义的肢体语言, 由不同的人表达, 会产生很大差异. 行为本身一般具有连续性, 由一连串的动作组成. 因此, 即便是获取了准确的骨架信息, 如何理解用户的肢体语言仍然是一个尚待解决的问题, 是目前的研究前沿.

场景感知的不足是虚实融合的重要障碍. 目前混合现实技术的重要成果, 尚主要体现在头盔显示器的定位精度和速度上. 人类通过五官感知现实世界的三维环境、光照变化、物体运动、行为特征等等, 然后根据环境的动态变化, 驱使虚拟物体做出恰当的调整. 混合现实环境中, 无论是现实世界还是虚拟世界的各要素间, 也需要构成合理的情景, 才能使得增强现实技术对现实世界产生切实的影响力. 如果不能理解用户的交互意图, 就无法使得虚拟角色做出恰当的响应. 对人类行为的理解深度, 决定了虚拟角色行为反馈的深度. 例如, 定位双手的位置也许就可以理解拳击的目标对象, 从而产生反馈; 但是, 要理解人类的情感微妙变化并做出响应, 就非常困难了. 计算机对现实世界的感知, 是通过各种传感器来实现的, 而摄像头或深度摄像头, 是目前视觉传感器的主体. 跟人类视觉一样, 视觉传感器能够捕获到高度清晰的细节, 但是, 受制于计算机视觉技术的进步, 从视觉传感器中分析得到的信息仍然是有限的. 近年来, 由于深度学习在计算机视觉领域的应用, 使得视觉技术能够获得越来越高精度的感知信息, 某些能力甚至已经超越人类的识别能力. 展望技术的发展, 视觉技术将极大地促进混合现实技术在应用领域的能力.

迄今为止, 现实世界是人类活动的空间和世界, 而虚拟世界是由计算机生成的世界, 这两个空间是分离的. 如前所述, 混合现实技术是这两个空间的自然接口和界面, 使得现实世界的人类了解和驱动虚拟世界的动态. 从这个意义上说, 机器人可以是具备二者属性的载体: 机器人可以与计算机一体化从而通过数据共享了解虚拟世界的一切, 同时通过其配置的各种传感器获得对周围世界的感知, 这种感知信息也可以通过共享数据与虚拟世界交流, 并且机器人可以具备与人类相同的一个重要特征: 具备可移动的行为能力. 从这个意义上说, 机器人是介于虚拟世界与现实人类的载体, 其与虚拟世界和现实世界的交互既具备人类的自然属性, 又具备虚拟人的属性, 成为第三类交互方式. 采用机器人作为替身, 在遥在、远程手术等领域都已有应用, 只是囿于技术的限制而难以广泛应用. 通过机器人来联结虚拟世界和现实世界, 实现跨越时空的关联, 将成为一种重要的技术手段.

4 混合现实中的智能融合

展望未来, 混合现实技术具有巨大的发展潜力. 混合现实能够建立用户、虚拟世界与现实世界之间的紧密联系, 通过自然人机交互实现一体化的知觉感知系统. 因此, 在混合现实环境中, 人类知觉自然延伸到计算机营造的虚拟世界之中. 那么, 在混合现实环境中, 人类知觉与计算机的联结未来可以达到什么样的程度呢?

在混合现实环境中, 用户可能看到对机器零件的标识, 或者一个虚拟的杯子掉落在现实的地面上而摔得粉碎, 这是可视空间的虚实融合; 也可能看到认识的虚拟角色在碰面的时候亲切地问候, 这是社

会空间的虚实融合. 然而, 计算机的世界可以远比对模拟场景的呈现要复杂得多. 由于计算机技术的巨大进步, 计算机存储和处理数据的能力已远远超过人类的能力极限. 然而, 计算机对数据的分析能力尤其不足, 缺乏人类通常具备的洞察力. 当人类知觉可以延伸到计算机营造的世界中, 人类大脑的分析判断能力, 是否可以与计算机的数据分析与处理能力相结合呢?

将现实世界中搜集到的信息, 经过大数据分析后, 通过信息可视化叠加到现实世界中, 通过引入现实世界的空间感, 增加对数据的临场感, 更好地实现对于大数据意义的理解. 由于混合现实环境具有交互功能, 因此, 可以提供给用户驱动数据的机能, 通过人类智能对于信息的分析和判断, 按照用户需求推动计算机对于数据的分析和语义提取, 并进一步呈现分析结果. 这样构成的分析闭环是人类智能与机器智能的交替作用. 混合现实环境中的虚拟世界是计算机营造的世界, 通过自然人机交互技术, 将虚拟世界提供给人类来操纵, 而且这种操纵是以符合人类知觉特征的方式进行, 从而将人类的感知自然延伸到计算机的世界. 从这个角度说, 混合现实能够提供人类智能与机器智能的融合平台.

混合现实是在线实时的反馈系统, 因此人类智能与机器智能的交互是自然融合的, 能够确保这种交流的流畅性. 由于信息及其语义的呈现必须要符合人类的感知特点, 因此信息的可视化成为非常重要的技术组成. 混合现实环境将人类与机器的优势结合起来, 利用人类理解力和洞察力, 又具有机器对数据的处理和分析能力, 由于结合了人类智慧与机器智能, 必将成为大数据时代的新工具.

5 结论

混合现实技术正在崛起, 从大趋势来说, 将成为未来虚拟现实技术中的领头羊. 但是, 混合现实技术在全面实用化之前, 尚有诸多技术困难需要克服. 从屏幕视频式到穿戴式头盔的混合现实技术, 使得用户自如地移动观察混合场景成为可能, 从而赢得了沉浸感. 然而, 如何合成混合场景呈现给观察者, 仍然是一个非常具有挑战性的任务. 由于虚实世界本身存在空间上和行为上冲突的可能, 解决其融合中的几何一致性、光照一致性和合成一致性, 是可视空间融合的核心问题. 当场景中包含虚拟角色时, 混合现实需要考虑在社会空间的虚实融合与交互, 产生虚实人群间的行为一致性, 并符合心理与社会规范. 混合现实中内容, 从场景的融合, 到信息及其分析结果的呈现, 可以囊括计算机生成的各种内容. 由于混合现实是自然人机交互的天然平台, 在混合现实环境中通过交互驱动计算机的分析, 通过可视化呈现反馈给用户, 从而形成人与机器智能之间的深度融合, 提供了人机智能深度融合的平台.

理想的混合现实技术, 是一个视觉上虚实难分的混合环境, 提供从静态场景到动态场景, 从刚体运动到柔性物体运动, 从非智能物体到智能物体, 涵盖虚拟与真实以及亦虚亦实的机器人的混合式环境, 用户既可以是现实的人类, 也可以是机器人; 交互既可以在用户与虚拟对象之间, 也可以在虚拟对象之间, 或者在虚拟对象与现实对象之间. 这样的场景, 既可以融合虚拟与现实的时空, 也可以融合现实中的不同时空, 因此也涵盖了遥在与远程协同等技术. 这目标固然是理想的, 但需要多个领域技术的巨大进步作为支撑. 然而, 令人乐观的是, 由于混合现实涉及广泛的交叉学科和技术, 每一项技术的突破, 都可以在现实中得到应用.

补充材料 文中所提参考文献 [5~7] 的图 1. 本文的补充材料见网络版 info.scichina.com. 补充材料为作者提供的原始数据, 作者对其学术质量和内容负责.

致谢 本专题的特约编辑和匿名审稿人提出了有价值的建设性意见, 作者对此表示衷心的感谢. 感谢作者的合作者们在相关研究中的进展带来的思想. 特别感谢浙江大学 CAD&CG 国家重点实验室

彭群生教授和鲍虎军研究员, 在合作和研讨中不断给予极具启发性的思路, 对本文核心思想的形成产生了重要影响.

参考文献

1 Azuma R, Baillot Y, Feiner S, et al. Recent advances in augmented reality. IEEE Comput Graph Appl, 2001, 21: 34–47
2 Costanza E, Kunz A, Fjeld M. Mixed reality: a survey. In: Human Machine Interaction. Berlin: Springer, 2009. 47–68
3 Wan G, Snavely N, Cohen-Or D, et al. Sorting unorganized photo sets for urban reconstruction. Graph Model, 2012, 74: 14–28
4 Li Y, Zheng Q, Sharf A, et al. 2D-3D fusion for layer decomposition of urban facades. In: Proceedings of International Conference on Computer Vision, Barcelona, 2011. 882–889
5 Nan L, Sharf A, Zhang H, et al. SmartBoxes for interactive urban reconstruction. ACM Trans Graph, 2010, 29: 93
6 Xu K, Huang H, Shi Y, et al. Autoscanning for coupled scene reconstruction and proactive object analysis. ACM Trans Graph, 2015, 34: 1–14
7 Lin J, Cohen-Or D, Zhang H, et al. Structure-preserving retargeting of irregular 3D architecture. ACM Trans Graph, 2011, 30: 183
8 Zhang G F, Qin X Y, Wei H, et al. Robust metric reconstruction from challenging video sequences. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, 2007
9 Tan W, Liu H M, Dong Z L, et al. Robust monocular SLAM in dynamic environments. In: Proceedings of International Symposium on Mixed and Augmented Reality (ISMAR), Adelaide, 2013
10 Liu H M, Zhang G F, Bao H J. Robust keyframe-based monocular SLAM for augmented reality. In: Proceedings of International Symposium on Mixed and Augmented Reality (ISMAR), Merida, 2016
11 Wang G F, Wang B, Zhong F, et al. Global optimal searching for textureless 3D object tracking. Visual Comput, 2015, 31: 979–988
12 Nakamae E, Qin X, Tadamura K. Rendering of landscapes for environmental assessment. Landsc Urban Plan, 2001, 54: 19–32
13 Zhong F, Qin X Y, Peng Q S. Transductive segmentation of live video with non-stationary background. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, California, 2010
14 Liu Y L, Qin X Y, Xing G Y, et al. A new approach to illumination estimation based on statistical analysis for augmented reality. Comput Animat Virt World, 2010, 21: 321–330
15 Zhang R, Zhong F, Lin L L, et al. Basis image decomposition of outdoor time-lapse videos. Visual Comput, 2013, 29: 1197–1210
16 Zhong F, Yang S, Qin X Y, et al. Slippage-free background replacement for hand-held video. ACM Trans Graph, 2014, 33: 199
17 Zhang Y J, Qin X Y, Julien P, et al. Online inserting virtual characters into dynamic video scenes (in Chinese). J Comput-Aided Design Comput Graph, 2011, 23: 185–191 [张艺江, 秦学英, Julien Pettré, 等. 虚拟群体与动态视频场景的在线实时融合. 计算机辅助设计与图形学学报, 2011, 23: 185–191]
18 Jiang X B, Zhong F, Peng Q S, et al. Action recognition based on global optimal similarity measuring. Multimed Tools Appl, 2016, 75: 11019–11036

Composition of virtual-real worlds and intelligence integration of human-computer in mixed reality

Baoquan CHEN* & Xueying QIN

School of Computer Science and Technology, Shandong University, Jinan 250101*, China*
*E-mail: baoquan@sdu.edu.cn

Abstract Recent rapid developments in mixed reality and virtual reality have attracted worldwide attention. This technology has prospective uses in education, medicine, video games and other fields. Mixed reality is the composite of a virtual world and the real world, typically with virtual objects incorporated into a view-dependent visual scene. A wearable head-mounted display provides a basic tool for immersion in mixed reality scenes, and a computer provides virtual scenes that coordinate with the real world. The composition of virtual and real scenes is essentially the inter-embedding of different temporal and spatial scenes, and depends on interactions with spatial geometry and illumination of the environment, namely geometrical consistency and illumination consistency. In the sociological sense, the composition must also follow the law of human sociology and psychology, that is, behavioral consistency. Mixed reality provides a natural and intuitive interface between users and real and virtual worlds. Mixed reality, through the understanding of human action and behavior, intuitively connects the virtual and real world. Humans can understand the world through the observation of the composition of the virtual and real scenes, and immediately influence the virtual world through natural interaction, receiving intelligent feedback from the computer. Thus, immersive interaction and deep interaction between humans and computers can be realized. Technological improvements from non-wearable to wearable devices have allowed mixed reality to gain immersion; in advancing from static scenes to dynamic scenes, mixed reality gained social attributes; and progressing from specific virtual scenes to semantic analysis and visualization of big data, mixed reality becomes a platform for the integration of human and computer intelligence.

Keywords mixed reality, composition of virtual-real worlds, social composition, intelligence composition

Baoquan CHEN received his Ph.D. degree in Computer Science from SUNY@Stony Brook, and M.S. in Electronic Engineering from Tsinghua. He is endowed Changjiang professor, and dean (Computer Science & Software) of Shandong University. Prior to the current post, he was the founding director of the Visual Computing Research Center, Shenzhen Institutes of Advanced Technology (SIAT), Chinese Academy of Sciences, and a faculty member at CS&E at the University of Minnesota at Twin Cities. His research interests generally lie in computer graphics and visualization.

Xueying QIN received her Ph.D. degree from Hiroshima University of Japan in 2001, and M.S. and B.S. degrees from Zhejiang University and Peking University in 1991 and 1988, respectively. She is currently a professor at the School of Computer Science and Technology, Shandong University. Her main research interests are augmented reality, computer vision and computer graphics, and focus on photo-realistic rendering, camera tracking, illumination estimation, virtual crowd simulation, object detection and tracking, action recognition, and composition of virtual-real scenes.