



评 述

虚拟现实增强技术综述

周忠^{①*}, 周颐^①, 肖江剑^②

① 北京航空航天大学虚拟现实技术与系统国家重点实验室, 北京 100191

② 中国科学院宁波工业技术研究院, 宁波 315201

* 通信作者. E-mail: zz@vrlab.buaa.edu.cn

收稿日期: 2014-04-08; 接受日期: 2014-07-07; 网络出版日期: 2014-12-16

国家自然科学基金 (批准号: 61170188, 61273276) 和国家高技术研究发展计划 (“863” 计划)(批准号: 2012AA011801, 2012AA011803) 资助项目

摘要 随着近年来计算机三维处理能力的增长和低成本传感显示元件的出现, 虚拟现实得到了快速发展, 特别是与现实世界产生了越来越多的结合技术, 从虚拟和现实的两个角度对虚拟现实进行增强. 论文重点围绕近几年的发展趋势, 论述了增强现实与增强虚拟环境的技术特点, 介绍了虚拟现实增强技术的相关硬件设备发展; 然后分别介绍了增强现实和增强虚拟环境技术的发展现状, 讨论了移动互联网上的虚实增强技术与应用, 并结合作者参与 ISO/IEC 的工作, 介绍了相关国际标准制定最新情况; 最后进行总结并提出需要解决的问题.

关键词 增强虚拟环境 增强现实 虚实增强 混合现实

1 引言

虚拟现实技术建立人工构造的三维虚拟环境, 用户以自然的方式与虚拟环境中的物体进行交互作用、相互影响, 极大扩展了人类认识世界, 模拟和适应世界的能力. 虚拟现实技术从 20 世纪 60~70 年代开始兴起, 90 年代开始形成和发展, 在仿真训练、工业设计、交互体验等多个应用领域解决了一些重大或普遍性需求, 目前在理论技术与应用开展等方面都取得了很大的进展. 虚拟现实的主要科学问题包括建模方法、表现技术、人机交互及设备这三大类, 但目前普遍存在建模工作量大, 模拟成本高, 与现实世界匹配程度不够以及可信度等方面的问题^[1].

针对这些问题, 已经出现了多种虚拟现实增强技术, 将虚拟环境与现实环境进行匹配合成以实现增强, 其中将三维虚拟对象叠加到真实世界显示的技术称为增强现实, 将真实对象的信息叠加到虚拟环境绘制的技术称为增强虚拟环境. 这两类技术可以形象化地分别描述为“实中有虚”和“虚中有实”. 虚拟现实增强技术通过真实世界和虚拟环境的合成降低了三维建模的工作量, 借助真实场景及实物提高了用户体验感和可信度, 促进了虚拟现实技术的进一步发展.

搜索热度代表了大众对于该词的关注程度, 一般来说, 新技术会引起搜索高潮, 然后慢慢下降, 在技术取得突破或出现某热点事件时激增, 最终趋于稳定. 我们使用 Google trends 对比了虚拟现实, 增强现实, 增强虚拟环境和混合现实等词的全球搜索热度, 为了有所参照, 以人机交互 (HCI) 作为参考, 搜索结果对比如图 1 所示. 可以看出, 和人机交互一样, 虚拟现实的搜索热度逐渐下降并趋于稳定, 这说

引用格式: 周忠, 周颐, 肖江剑. 虚拟现实增强技术综述. 中国科学: 信息科学, 2015, 45: 157-180, doi: 10.1360/N112014-00076

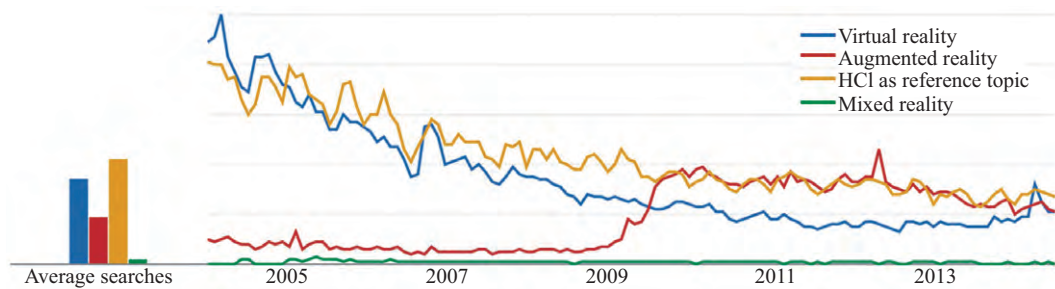


图 1 虚拟现实、增强现实和混合现实搜索量统计对比 (来源: Google trends, 2004.01~2014.06)

Figure 1 The search comparison of virtual reality, augmented reality, and mixed reality (from Google trends, 2004.01~2014.06)

明虚拟现实技术正在成熟中, 逐渐被大众所接受, 其中最近的峰值是 Facebook 收购虚拟现实眼镜企业 Oculus. 与之相反, 增强现实的关注度在 2008 年开始迅速上升, 并连续出现多个热度高峰, 通过查看其关联新闻可以发现, 其热度主要是 2007 年 Apple 公司发布 iPhone, 移动互联网兴起所推动, 特别是在摄像头加上陀螺仪成为智能手机的标配, 为增强现实技术提供了充分的发展空间. 这其中 Google 眼镜和网上虚拟试衣吸引了很多关注度. 而增强虚拟现实, 增强虚拟环境, 增强虚拟 (augmented virtuality) 等词的搜索量远小于以上热词的量级, 反映了增强虚拟环境技术还有待于突破, 尚未出现重要事件.

可以看出, 对虚拟现实技术进行增强正发展成为重要的方向, 具有很大的发展潜力. 本文重点围绕近几年虚拟现实增强技术的发展趋势, 首先论述了技术特点, 介绍了相关的硬件设备发展, 然后分别介绍了增强现实和增强虚拟环境技术的发展现状, 考虑到移动互联网是信息技术发展的又一次革命, 简介并讨论了移动互联网上的虚实增强技术与应用, 接下来结合作者参与 ISO/IEC JTC1 SC 24 分技术委员会的工作, 介绍了相关国际标准制定的最新情况, 最后进行总结并给出了几个需要解决的问题.

2 技术特点

现在业内普遍认可: 从真实世界到虚拟环境中间经过了增强现实与增强虚拟环境这两类虚拟现实增强技术^[2]. 国际上一般把真实世界 (计算机视觉)、增强现实、增强虚拟环境、虚拟现实这 4 类相关技术统称为虚拟现实连续统一体 (VR continuum). 与早期相比, 增强现实或增强虚拟环境的概念已经发生了很大的变化, 技术领域大为拓宽, 但它们的技术特征都离不开如下 3 点: (1) 将虚拟和现实环境进行混合; (2) 实时交互; (3) 三维注册.

下面分别对增强现实与增强虚拟环境进行技术分析. 需要说明的是, 此处的“相机”是指广义上的视觉采集设备, 不单包括摄像头, 也包括红外深度相机、激光扫描仪等.

增强现实技术通过运动相机或可穿戴显示装置的实时连续标定, 将三维虚拟对象稳定一致地投影到用户视口中, 达到“实中有虚”的表现效果. 如图 2 所示, 真实世界是我们所处的物理空间或其图像空间, 其中的人和竖立的 VR 牌是虚拟对象, 随着视点的变化, 虚拟对象也进行对应的投影变换, 使得虚拟对象看起来像是位于真实世界的三维空间中. 图 2 和 3 中的虚线对象代表虚拟环境对象, 实线对象代表真实对象或其图像.

增强现实还有一个特殊的分支, 称为空间增强现实 (spatially augmented reality)^[3], 或投影增强模型 (projection augmented model)^[4], 将计算机生成的图像信息直接投影到预先标定好的物理环境表面, 如曲面、穹顶、建筑物、精细控制运动的一组真实物体等. 本质上来说, 空间增强现实是将标定生成的

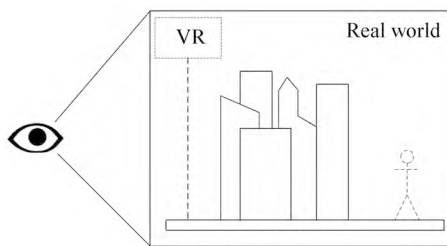


图 2 增强现实技术“实中有虚”

Figure 2 Augmented reality mixed the virtual object into real world

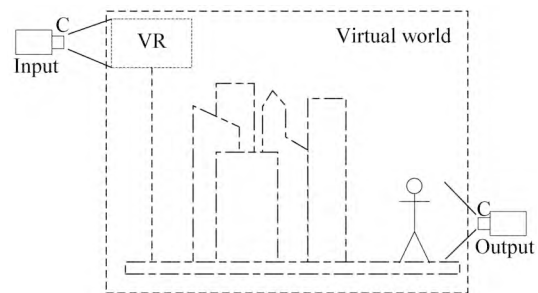


图 3 增强虚拟环境技术“虚中有实”

Figure 3 Augmented virtual environment mixed the real object into virtual environment



图 4 TED 2010 上叠加真实影像的地图

Figure 4 Augmented-reality maps in TED 2010

虚拟对象投影到预设真实世界的完整区域, 作为真实环境对象的表面纹理. 与传统的增强现实由用户佩戴相机或显示装置不同, 这种方式不需要用户携带硬件设备, 而且可以支持多人同时参与, 但其表现受限于给定的物体表面, 而且由于投影纹理是视点无关的, 在交互性上稍显不足. 实际上, 我国现在已经很流行的柱面、球面、各种操控模拟器显示以及多屏拼接也可以归为这一类. 最著名的投影增强模型的是早期的“shader lamps”^[4].

增强虚拟环境技术预先建立了虚拟环境的三维模型, 通过相机或投影装置的事先或实时标定, 提取真实对象的二维动态图像或三维表面信息, 实时将对象图像区域或三维表面融合到虚拟环境中, 达到“虚中有实”的表现效果. 如图 3 所示, 在虚拟环境中出现了来自于真实世界的实时图像, 其中 VR 牌上的纹理和人体都来自于相机采集的图像, 人体甚至可以是实时的三维对象及其表面纹理图像.

与增强现实中存在的投影增强模型技术正好相反, 增强虚拟环境技术中也有一类对应的技术, 用相机采集的图像覆盖整个虚拟环境, 即作为虚拟环境模型的纹理, 用户可以进行高真实感的交互式三维浏览. 当这种三维模型是球面、柱面、立方体等通用形状的内表面时, 这种技术也就是现在已经很普及的全景 (panorama) 图片或视频. 全景视频将真实世界的一幅鱼眼或多幅常规图像投影到三维模型上, 构造出单点的全方位融合效果, 多幅图像之间的拼接可以是图像特征点匹配或相机预先标定等方式. 微软 Bing 地图架构师 Arcas 在 TED 2010 的演讲中演示了一种新颖地图应用研究, 在全景图片增加实时视频内容的叠加¹⁾, 如图 4 所示. 这种增强方式可以反映同一地点各种影像的空间几何关系, 用户可以自由浏览全景, 就像在现场一样, 产生了更加真实的虚拟环境效果.

1) Arcas B A y. Augmented-reality maps. TED Talks, 2010.

表 1 ISMAR (2008 ~ 2012) 发表 S&T 论文分方向统计
Table 1 AR research categories in the literature set of ISMAR (2008 ~ 2012)

Category	2008	2009	2010	2011	2012	Total	Percentage (%)
Tracking	4	4	8	8	6	30	23.4
Interaction	1	6	7	0	3	17	13.3
AR App.	3	3	3	5	2	16	12.5
Mobile AR	4	4	0	3	4	15	11.7
Evaluation	4	2	2	2	0	10	7.8
Display	4	2	2	0	1	9	7.0
Rendering	1	1	0	0	4	6	4.7
Calibration	2	0	0	1	1	4	3.1
Visualization	1	1	0	1	1	4	3.1
Modeling	0	1	0	1	1	3	2.3
Multimodal AR	1	0	0	0	0	1	0.8
Others	1	0	2	5	5	13	10.2
Summary	26	24	24	26	28	128	100.0

IEEE ISMAR (International Symposium on Mixed and Augmented Reality) 是专门讨论虚拟现实增强技术的国际会议, 发表的论文包括两类: 科技 (Science & Technology, S&T) 论文和艺术人文 (Arts, Media & Humanities, AMH) 论文. 其微软学术搜索计算的 H 因子为 41, 从 1999 年到 2011 年间共发表论文 753 篇, 引用总计 7514 次 (其中自引 992 次), 特别是从 2004 年开始引用频次明显提高²⁾. Zhou 等^[5] 在 2008 年综述了 ISMAR 及其前身会议的 10 年间发表 S&T 论文情况, 相机跟踪、交互、标定、应用、显示和移动 AR 是最主要的 6 个方向, 而高引用论文分布的前几个方向主要是相机跟踪、交互、标定、应用、编著和移动 AR. 可以看出, 当前相机跟踪、交互和标定是虚拟现实增强技术研究关注的热点, 另外, 应用模式也是业内正在积极探索的内容, 应用 (含移动应用) 及其引用比例均能够占整个会议的 1/5.

本文还对 ISMAR 2008 年至 2012 的论文 (regular paper) 进行了统计, 同样按照前述的几个方向进行分类. 从表 1 和表 2 可以看出, 在近几年的发展中, 热门方向大致分布不变, 相机跟踪和交互依然占主导地位, 尤其是近 5 年来的最佳论文奖有 3 篇研究相机跟踪. 而 AR 应用和移动 AR 的探索明显增多, 验证了移动互联网的影响与推动.

IEEE VR (IEEE Virtual Reality Conference) 是 IEEE 关于虚拟现实技术的专业国际会议, 虚拟现实增强技术也是其接受的论文主题之一. 该会议源于 1993 年的会议 Virtual Reality Annual International Symposium. 根据微软学术搜索, 20 多年间共发表论文 461 篇, 引用总计 6511 次 (其中自引 144 次). 该会议一般包括长文 (Long Paper) 和短文 (Short Paper), 近年来部分长文会被推荐至 *IEEE Trans. Vis. Comput. Graph* 发表. 本文统计了 IEEE VR 近 5 年发表论文情况, 如表 3 所示, 相机跟踪与标定、交互、AR 应用依然是最主要的研究方向, 但相比于 ISMAR, 多出了重点研究方向 —— 感知. IEEE VR 的文章以系统级应用为主, 覆盖面广于 ISMAR, AR 应用、交互、感知、跟踪和图形是前 5 个研究方向, 其论文数量约占论文总数的 78.5%.

2) <http://www.ismar.net/>.

表 2 ISMAR (2008 ~ 2012) 不同方向的高引用论文比例
Table 2 Proportion of highly cited papers in ISMAR (2008 ~ 2012)

Category	Papers (%)	Citations (%)
Tracking	23.4	22.6
Interaction	13.3	8.8
AR App.	12.5	10.3
Mobile AR	11.7	32.9
Evaluation	7.8	9.9
Display	7.0	3.3
Rendering	4.7	2.3
Calibration	3.1	1.0
Visualization	3.1	0.8
Modeling	2.3	1.1
MultimodalAR	0.8	0.8
Others	10.2	6.1

表 3 IEEE VR (2009 ~ 2013) 发表论文分方向统计
Table 3 Research categories in IEEE VR (2009 ~ 2013)

Category	2009	2010	2011	2012	2013	Total	Percentage (%)
AR App.	8	12	7	4	5	36	20.3
Interaction	3	11	10	4	6	34	19.2
Perception	9	8	3	5	4	29	16.4
Tracking calibration	2	3	9	6	2	22	12.4
Graphics	3	3	3	5	4	18	10.2
Collaboration & Avatars	4	4	0	1	5	14	7.9
Display	6	0	0	0	5	11	6.2
Evaluations	0	0	4	1	1	6	3.4
Others	0	0	0	3	4	7	4.0
Summary	35	41	36	29	36	177	100.0

3 相关硬件设备发展

虚拟现实增强技术需要与真实世界相结合, 不可避免需要相机等传感装置及显示设备, 因此相关硬件设备的发展对该技术有着重要的影响. 近年来, 摄像头质量的提高和普及, 红外安全激光技术的成熟和消费级产品出现都大大促进了虚拟现实增强技术的进步和应用. 本节简介增强现实与增强虚拟环境常用硬件设备的最新发展情况.

3.1 相机

摄像头是增强现实技术最重要的硬件设备, 大量的相机跟踪和标定技术都是以简单摄像头为基本配置. 摄像头作为一种廉价、标准、易于获取和集成的采集设备, 有着巨大的市场需求, 业内竞争极其激烈. 特别是随着智能手机的出现, 一直处于高速发展中. 前后双摄像头已经成为了智能手机的标

准配置. 现代的摄像头成本越来越低, 尺寸越来越小, 分辨率越来越高, 成像质量也越来越好. 据报道, iPhone 元件供应商之一, Lead-mall 出售的 800 万像素摄像头模块价格仅为 9 美元. 2013 年 7 月 30 日, Apple 公司的一项面向移动设备的摄像头专利获得授权引起大量关注, 这项技术具有 3 传感器, 3 镜头的设计, 据称能大幅度提高成像质量^[6]. Nokia 公司在 2012 年 2 月 27 日发布了第一款高达 4100 万像素的手机, 基于塞班系统的 Nokia 808, 接着在 2013 年 7 月 11 日发布了基于 Windows Phone 平台的 Lumia 1020 手机. Lumia 1020 也配置了 4100 万像素摄像头, 可以捕捉分辨率为 3800 万像素、采样 500 万像素的图像. 上述硬件设备的发展不但使智能手机拍照功能竞争加剧, 而且对数码相机产业甚至高端的单反相机前景都形成了新的挑战.

全景包含了全方位的图像信息, 在可视角度和交互性上具有优势. 目前, 全景图像的合成主要来源于 3 种: 普通相机拍摄、软件后期合成和全景相机拍摄. 其中普通相机拍摄是让同一相机进行连续微小运动, 拼接得到的多帧图像, 从而合成出全景图像. 软件后期合成则是完全依靠图像特征点匹配等方法对重叠的图像进行拼接. 上述两种方法缺点都是不能做到连续实时地合成全景视频, 如果要得到高质量的全景图像还需要手工图像处理, 工作量较大. 全景摄像机通过事先标定方式, 可在采集或回放时实时拼接合成, 得到连续实时的全景视频, 后续工作量小, 适合实现大范围无死角的全景监控. 全景摄像机大致可以分为单成像传感器型和多成像传感器型两大类.

在单成像传感器全景摄像机中又可以分为球面折反式和鱼眼式. 球面折反式是在常规单镜头摄像机前面安装一个球面反射镜, 通过对入射光线的折反来实现对周边区域的 360° 环视. 由于在摄像机的正前方安装了一个反射镜, 会挡住摄像机的正前方入射光, 在图像的中心部位形成盲区, 因而只能用于水平方向的 360° 环视. 典型厂商有美国的 Bellissimo 公司、中国的普维光电等. Kogeto Dot 是一个小巧的摄影附件, 以三点式固定在 iPhone4 上, 在镜头前套上一个全景拍摄用反射镜, 下载免费配套 APP 即可实现全景视频拍摄. 鱼眼摄像机设计相对简单, 通常是在普通的成像传感器前加装一个超短焦距的鱼镜头, 实现大视场的观测. 典型厂商有德国的 Mobotix、中国的海康威视等. 然而, 这种鱼眼相机由于镜头需要接受大角度的光线, 会产生较强的图像畸变, 特别是从镜头边缘获取的图像模糊不清. 此外, 单成像传感器相机整体曝光不能适应所有区域, 高像素密度芯片价格更贵, 虽然存在上述缺点, 但单成像传感器尺寸可以做到很小, 除镜头外的其他部分容易开发和加工.

多成像传感器拼接型全景摄像机使用多个相机拼接的方式扩大了图像分辨率, 缓解成像过程动态响应不足, 解决图像严重扭曲分辨率不均等问题. 最为有名的是 Google 公司街景采集车上的全景相机, 典型产品有美国 Immersive Media 的 Dodeca 2360、加拿大 PointGray 的 LadyBug、美国 Elphel 公司的 360 全景相机等. 其中, PointGray 公司的 LadyBug 5 全景系统采用 6 个 5 百万像素的 CCD 传感器, 可以获得近 3 千万像素的全景合成图像³⁾. Elphel 公司的 Eysis4Pi 全景系统采用 26 个 5 百万像素 CMOS 传感器, 可以获得近亿像素的全景合成图像⁴⁾, 其价格高达 10 万美元一台. 这些产品目前尺寸较大, 价格昂贵. 国内中科院宁波所研发了 8 镜头拼接的 360° 全景摄像机, 并已进入产品化⁵⁾. 腾讯等公司的街景由 5 个单反相机的照片后期拼接而成, 随着技术和工艺的成熟, 全景相机的价格有望大幅度降低.

在新的多相机拼接设备的支持下, 大范围监控从“看得更多更清楚”向“找到和表现出最需要的信息”发展. 美国军方支持大范围监控的拼接型摄像机最近开始正式服役, 公布的测试数据在 40 km² 范围可保持 15 cm 清晰度, 掀起了新一轮摄像技术的变革. 从 2007 年开始, 美国国防部先进设计计划

3) <http://ww2.ptgrey.com/products/ladybug5/>.

4) <http://www3.elphel.com/eyesis-4pi/>.

5) <http://www.panoeeye.com.cn/>.

局 (DARPA) 资助 BAE 系统公司研发了自动实时地面持续监控影像系统 (ARGUS-IS), 投资 1850 万美元, 耗时 30 个月制造完成. 2010 年, 在 UH-60 黑鹰直升机上测试通过, 现已在美军新一代无人直升机——波音 A160T “蜂雀” 上装备 [7]. ARGUS-IS 实现了 18 亿像素的宽视场航空影像连续拍摄, 该系统由 4 组相机构成, 每组相机由 92 个 500 万像素镜头组成, 拍摄速度每秒 5 帧 [8]. 从 5 km 的高空进行拍摄可以监控约合 $4 \times 10^5 \text{ km}^2$ 区域, 并可以有效地分辨出地面上最小 15 cm 的物体. 利用 ARGUS-IS 从高空监测整个军事港口的情况, 借助这一先进的监视系统, 指挥官能够对整个战场的一举一动进行实时监控. 在 2012 年 6 月 21 日, *Nature* 杂志网络版报道了 Duke 大学研发出的一种“超级照相机”原型. 超级照相机能将 98 台独立相机 (每台拥有 1400 万像素) 的传感器获得的画面拼接起来, 从而创造出一张 9.6 亿像素的图像 [9]. 该照相机的分辨率非常高, 可以看见 1 km 外 3.8 cm 宽的物体. 当用近乎鱼镜头的 120° 视角来拍摄西雅图时, 这台 93 kg 重的相机可以捕捉到足够多的细节, 甚至能够读到两个街区之外交通标识牌上的“罚款”字样. 其中, 光学器件仅占了这台相机体积的 3%, 而 $75 \text{ cm} \times 75 \text{ cm} \times 50 \text{ cm}$ 的相机尺寸则既能保证容纳下相机的电路板, 又可使其免于过热. 但是, 这台相机不具备连续拍摄能力, 只能对一个特定时刻进行拍摄.

3.2 位置与角度传感器

GPS 和加速度计已经成为了智能手机的标配, 位置和角度传感器可以对相机跟踪起到重要的辅助作用. GPS 主要根据空间卫星和地面 GPS 传感器之间的往返时间进行三角测量得到数据, 除卫星数量和通信因素外, 容易受高楼遮挡、气候等条件影响, 目前我国的普通廉价 GPS 大致只能实现精度 5~10 m 的位置定位. 差分 GPS 可以接入差分网, 以地面基站作为准确“地标”进行高精度位置测量.

加速度计 (accelerometer) 可以测量设备的加速度方向, 以 iPhone 为例, 可以安装“Sensor Data”采集传感数据. 加速度计的测量值从 -1 到 1 变化, 当手机轴为完全水平时, 测量值为 0, 当轴为竖直时, 测量值为 -1 或 1, 利用简单的三角函数可以大致计算出手机的倾斜角. 手机用的廉价加速度计测量精度低, 大多只能测量一个倾斜角, 所以一般只是用来监测设备的竖直状态, 控制图片或电子书阅读方向等. iPhone 3GS 在移动设备上集成加速度计、陀螺仪 (gyroscope) 和磁力计 (magnetometer) 等设备进行结合计算, 率先推出了“电子罗盘”APP, 可以测量设备的三个旋转角度, 可用于人机交互, 并促进了增强现实 APP 的出现, 例如街景图像的全景控制等.

3.3 可穿戴增强显示设备

Bimber 和 Raskar^[10] 按应用场景不同, 对增强现实显示设备进行了分类, 如图 S1 所示, 有头戴式、手持式和空间投影式. 头戴式数字头盔 (head-mounted display, HMD) 一直是增强现实的传统研究内容, 一般被分为光学透射式 (optical see-through, OST) 头盔和视频透视式 (video see-through, VST) 头盔. 光学透射式是指用户透过透明镜片看到真实世界, 并通过反射或投影方式看到虚拟环境或对象; 视频透视式是指将头盔上集成摄像头采集的外部图像与虚拟场景相合成, 然后输出到用户眼前的小屏幕上, 其原理和技术进展可参考 [5,10]. 2012 年, 初创公司 Oculus 推出了大视场低延迟的消费级 HMD Oculus Rift, 取得了重要突破. 但到目前为止, 绝大多数 HMD 仍然价格昂贵、标定复杂、精度和分辨率不够理想, 突破性的进展不大.

精确的三维空间位置跟踪是 HMD 的一个难点, 限制了其应用范围和成本. 但实际上, 近年来, HMD 的其他部件, 如摄像头、陀螺仪、微型投影机、微型显示屏等均在尺寸、成本和技术指标上有了很大突破, 因此类似 Google glass 增强现实眼镜、Oculus 虚拟现实头盔的消费级产品出现也就可以理

解了. 不考虑三维位置定位后, Google glass 在光学设计, 骨传导和器件集成等方面取得了突破, 成功将智能手机的功能集成到超轻薄的可穿戴眼镜上. Google glass 已于 2014 年上市, 受此影响, 多家公司开始加入这一类产品的研发行列中, Vuzix, GlassUp, Sony, Olympus 等公司在 2013 年发布了多款增强现实眼镜. 2014 年 3 月 20 日, 著名的 Facebook 公司宣布以 20 亿美元收购虚拟现实头盔显示器公司 Oculus —— 一家成立仅有 2 年的硅谷创业公司. Oculus 公司率先研发了一种消费级的高性能头盔显示设备 —— Oculus RIFT, 售价仅为 300 美元. 之后, sony、三星等公司纷纷跟进. 随着业内大公司的积极推动, 可穿戴式三维显示似乎进入了一个新的高速发展时期, 也大大增加了增强现实和虚拟现实技术的影响力.

3.4 体感交互设备

近年来国际上三维体感交互设备的突破性产品连续出现. 体感交互设备可以采集人体的三维运动数据, 提供了将真实世界传感数据合成到虚拟环境的重要方式, 是增强虚拟环境技术的重要设备. 体感交互设备的突破主要来自于飞行时间 (time-of-flight, TOF) 技术, 它随着低成本红外半导体传感器的成熟而提出, 测量原理与三维激光扫描仪大致相同, 都是测量光的往返时间. 所不同的是激光扫描仪是逐点扫描, 而 TOF 是对光脉冲进行调制并连续发送和捕获整个场景的深度. 因此与激光扫描相比, TOF 相机的优点是捕获速度非常快, 缺点是分辨率低、测量精度低.

最早出现的是 TOF 深度相机, 知名的有 ZCam, Mesa, PMD, Canesta 等公司的产品, 它们的光发射频率约几千万 Hz, 捕获速度最高可达到每秒 100 帧, 但分辨率最高不超过 320×240 , 价格昂贵, 一般需要几万元人民币. 微软在 2010 年推出了 Kinect, 它的红外 LED 向外投射光斑阵列, 通过一种基于采样深度数据比较估计的光编码算法, 大幅度降低了光的发射频率, 提高了深度图计算的速度和分辨率 (320×240), 成本也大幅度降低到一千多元, 掀起了研究和娱乐等应用热潮. 2012 年微软又推出了 Kinect for Windows, 进一步提高了深度图的分辨率, 达到 640×480 , 并在适用距离和精度上有了较大提高. 其 SDK 提供了更稳定的人体骨骼、面部跟踪以及三维重建 API^[11].

2012 年 5 月, Leap Motion 公司推出了小型运动控制系统 Leap 3D, 可以追踪多个物体并识别手势, 其识别精度为 0.01 mm, 再次掀起了整个互联网领域对体感交互设备的惊奇与研究热潮. 2013 年 7 月 Leap 3D 正式发售, 其定价仅为 79.99 美元. Leap 3D 控制器只有 8 cm 长, 集成了 2 个 130 像素的网络摄像头传感器和 3 个红外 LED, 采用的是将光编码技术和双目立体视觉相结合的算法⁶⁾. 2013 年加拿大 Thalmic Labs 公司研发的手势控制腕带 MYO 则独辟蹊径, 更显神奇. 它通过检测用户运动时胳膊上肌肉产生的生物电变化, 不止实时, 甚至提前在物理运动之前进行手势识别, 其售价仅为 149 美元, 再次成为业内焦点⁷⁾.

这些体感交互设备能够将真实世界的人体运动在虚拟环境中实时精确表示, 增强了虚拟现实的交互能力. 随着相关设备的发售和解密, 我国也有一些公司和研究机构跟踪和研究.

3.5 三维立体显示设备

现在市面上可以购买的立体显示设备已经有很多了, 最常见的是基于偏振片或奇偶快门切换的立体电影电视、显示器、立体投影等, 这类设备需要佩戴对应的 3D 眼镜, 每帧包括对应左右眼的两幅图像. 基于偏振片的立体显示是在空间上进行了偏振过滤, 因此在亮度上有所损失; 基于快门切换的立体显示是在时间上进行了左右眼图像交替显示, 因此在帧速上有所损失. 这些双目立体显示方式只能

6) <http://www.digitalversus.com/motion-sensors/leap-motion-the-leap-p15766/test.html>.

7) <https://www.thalmic.com/myo/>.

表 4 硬件设备对应的常用虚拟现实增强技术
Table 4 Common augmented virtual reality technique corresponding to the hardware

	AR	AVE
Camera	Image capture; tracking	Image capture
Projector	Projection augmented model	N/A
Gyroscope	Rotation tracking	Rotation tracking
GPS	Location tracking	Location tracking
HMD	Display; assistant tracking	N/A
Depth Camera	Tracking based on depth map	Motion capture; 3D reconstruction
3D scanning LiDAR	Offline 3D scanning	Offline 3D scanning; accurate tracking based on point cloud, such as the google self-driving car
Multi-Projector	Multi-screen tiled display	N/A
Multi-Camera	Image capture; accurate tracking	3D reconstruction; virtual-real fusion

观看到固定的视差图像, 并不能获得跟随头部或眼睛运动的真实观看体验. 近年来出现了一些“裸眼 3D”产品, 这些产品通过栅格或光栅来控制光线的投射方向, 每帧绘制 10 幅或更多幅图像, 在不同方向上独立显示, 从而用户可以在不同位置看到不同图像. 这就要求显示器能够支持很高帧速的图像带宽, 因此设备价格更为昂贵, 而且只有在面向显示器的一定正向范围才能看到较好的 3D 效果. 随着移动设备的兴起, 在多点触摸的电容屏上集成各种各样的膜成为工业界的流行做法, 比如 EyeFly3D 等. 也出现了小尺寸用于手机或掌机的 3D 屏幕, 如三星公司在 2010 年发布的 W960 手机, 采用了 3.2 英寸的 3D AMOLED 裸眼 3D 触摸屏, 不过分辨率较低, 只有 WQVGA (240×400). 日本任天堂公司在 2012 年发布的 3DS LL 配备有 4.88 英寸裸眼 3D 液晶屏, 分辨率达到 800×240, 左右眼都可以观测到 400×240 像素. 还有的三维显示器内置摄像头跟踪人的头部或眼睛, 进行视点相关的三维绘制.

另外, 近年来, 有机发光二极管 OLED (organic light-emitting diode) 受到热捧, 我国大量投资建设 OLED 生产线, LG、三星 2014 年先后推出了 55 英寸 OLED 曲面电视. OLED 的优势之一是可以制作大尺寸柔性曲面屏幕, 其技术成熟以及成本下降将为大尺寸投影显示市场带来很大冲击, 也将促进增强现实技术的新发展.

长期以来, 三维空间立体成像 (俗称真三维裸眼 3D) 得到广泛的关注和期待, 但一直缺乏重大突破. 2007 年 Jones 等人 [12] 在 ACM SIGGRAPH 大会上首先展示了一种用高速旋转的镜子进行 360° 反射实现空间三维成像的方法, 很快得到大量关注. 近几年, 他们在该方面开展研究, 每年都在 SIGGRAPH 大会展示基于该技术的远程再现系统进展. 最近全息三维成像研究更是取得了基础性突破, 2010 年 11 月 4 日, Nature 封面文章介绍了全息成像的突破性工作 [13]. 亚利桑那大学研发了一种新型光致折变薄膜, 用纳米脉冲激光向该材料写入全息图像, 可以产生全息立体成像用于远程再现, 迅速成为热点新闻. 在三维空间立体成像的未来应用上, 很多人提出了各种各样的用户远程再现 (tele-presence) 或远程沉浸 (tele-immersion) 的想象, 例如著名的好莱坞大片《星球大战》(Star Wars)、《黑客帝国》(MATRIX) 和《阿凡达》(AVATAR) 等均来自于这个主题. 总的来看, 三维空间成像还有待突破和成熟, 尚未出现能大规模推广的产品.

在以上硬件设备介绍的基础上, 我们将主要硬件设备对应的常用虚拟现实增强技术进行了总结, 如表 4 所示, 其中“N/A”代表无典型技术.

4 增强现实

虚拟现实技术带来了人机交互的新概念, 带给用户强烈的真实感和临场感的体验, 但这要求精细的三维模型和复杂的渲染计算. 增强现实技术“实中有虚”, 摄像机采集的图像真实地反映了现实场景, 可以减少场景的建模和渲染工作量, 提供了一种轻量级并且真实感强的增强技术. 下面主要从相机跟踪注册和虚拟场景实时融合绘制的角度对增强现实技术最新发展情况进行介绍.

4.1 相机跟踪注册

在增强现实中, 相机跟踪注册是在相机连续观测的二维图像与虚拟的三维场景之间建立空间投影关系, 也就是对虚拟眼睛 (即相机) 的位置和姿态进行实时参数估计. 基于特殊标识的相机跟踪需要在场景中放置较为容易识别且一直可见的标识物, 这大大限制了增强现实的应用范围. 相比之下, 基于自然特征的注册技术没有上述约束条件, 通过对场景中的二维或三维特征进行提取和跟踪来实现相机自定位. 该技术首先在机器人视觉导航领域得到了广泛关注和研究, 目前逐步开始应用于增强现实领域.

最近 10 年间, 国际上对基于自然特征的相机自定位技术展开了深入研究. 在单目相机三维注册技术方面, Davison^[14] 于 2003 年提出了基于单目视觉的实时同步定位和地图构建 (SLAM) 研究, 使用全状态扩展 Kalman 滤波器跟踪少量 Harris 角点, 以实现在室内环境下逐帧更新相机的姿态, 在视觉定位领域做出了重要推动. 该算法在室外光线变化较大和快速运动的情况下性能较不稳定.

在 Davison 的工作基础上, Klein 和 Murray^[15] 提出了 PTAM 算法, 将单目视觉 SLAM 系统中的地图构建和帧间特征跟踪进行分离. 首先使用双目立体视觉的方法来建立场景的初始深度图, 然后一方面利用多核 CPU 并行计算估计光流场实现对特征点的实时跟踪, 另一方面利用光束平差法对多个局域关键帧进行相机姿态优化, 因此之后的跟踪性能在很大程度上依赖于初始化过程中摄像机的运动以及每一帧的选择, 初始化过程中摄像机在保持旋转不变的情况下平移要尽可能的小, 而且初始化过程不应持续较长的时间. PTAM 在室内小范围场景内取得了较为精确的相机定位结果, 如图 S2 所示.

Newcombe 和 Davison^[16] 提出一种紧致的相机跟踪算法 DTAM^[17], 不再使用特征点匹配方法, 而是选用大量窄基线图像进行逐点计算逆向深度的方法. 该方法提到相机簇的概念, 即所有满足窄基线关系的图像为一个相机簇. 作者定义了参考帧下每个点的正则代价 (regularised cost), 对相机簇上每个点的梯度进行 Huber 加权, 并加上正则项和辅助项. 在满足正则代价最小的条件下, 计算参考帧每点的深度, 进而进行相机姿态的准确估计和场景的重建. 浙江大学 Tan 等人^[18] 针对动态场景提出了一种单目相机跟踪方法, 该方法改变了关键帧的表示与更新方法, 从而能够适应于场景中部分物体移动或者整个场景渐变的情况, 同时还提出了一种先验自适应的 RANSAC 变种方法, 以适用于他们的方法.

Pollefeys 等人^[19] 在对室外的大场景进行重建时, 同样利用了 GPU 并行流水技术优化的 KLT 算法实现上千个 SIFT 特征点实时跟踪, 同时他们还使用 GPS 和 IMU 进行相机姿态的辅助定位.

相比于单目视觉定位, 基于立体视觉原理的运动估计由于能够获取更为准确的三维点重建, 在相机定位时具备更高的精度和稳定性. 在早期的系统中, 为了追求实时性, 自然特征的跟踪注册通常只考虑在相邻的图像帧间进行姿态估计, 因而在行走期间的误差和系统漂移会不断积累并增加. 如 Nistér 等人^[20] 构造的双目视觉 SLAM 系统为了追求实时性的要求, 仅将视觉估计的运动结果进行简单累加, 其系统误差约 5%; Konolige 等人^[21] 所构建的立体视觉 SLAM 系统的误差约 4%. 为此, 一些学者提出了改进方案减少系统漂移. Klein 和 Murray^[15] 提出通过与 GPS, IMU 以及方向传感器的联动和数据融合, 减少对视觉系统的依赖, 使系统的整体误差漂移率可降低到 1%~2%. Zhu 等^[22,23] 提出

利用全局和局部地标数据改进图像特征的匹配精度, 实现更大范围的相机定位和导航, 该系统使用了两对立体相机前后观察, 并通过动态选择图像帧特征点构造地标数据库, 从而纠正行走期间的全局漂移并降低累积误差。

2010 年微软推出的 Kinect 为视觉跟踪技术提供了一种新的设备, 它的基本原理是主动结构光测距, 可以快速捕获准确的深度图, 在室内小范围场景取得了优秀的结果^[24,25], 如图 S3 所示. Kinectfusion 算法采用 ICP 算法对前后帧的三维点云进行匹配注册, 估计相机姿态. 该方法使用距离符号函数来定义体素的值, 进而可以计算多帧的加权结果. 还有人提出了帧 — 模型的匹配方法, 相比帧 — 帧的方法, 更加鲁棒, 适应于连续帧的重建问题. 然而存在的共同问题是计算量较大而导致实时性不佳. 在大场景定位方面, 目前还没有基于 Kinect 深度相机的较好研究成果. Kinect 相机理想的有效深度在 10 m 以内, 而基于可见光的双目立体相机通过调整基线长度, 可以获得更远的深度估计范围. 因此, 如将 Kinect 深度相机、立体相机和惯性导航单元 (IMU) 等不同模态传感器进行有机结合, 将可能大大提高现有三维注册技术的精度, 实现大范围高精度的增强现实技术。

4.2 虚实场景实时融合绘制

在真实场景中, 物体往往具有不同的深度信息, 物体之间会随着用户视点的位置变化产生不同的遮挡关系. 与此同时, 场景中还会存在其他的动态物体 (如人、车等), 它们的深度还会随着物体自身的运动而发生变化, 从而造成更为复杂的遮挡关系. 在增强现实系统中, 每一个绘制的虚拟物体均需要被准确地放置在场景中, 并应与周边不同深度的景物实现交互, 确保正确的遮挡关系和交互关系. 这就需要场景绘制系统能够在实时估计场景深度的基础上, 实现有效的遮挡处理、碰撞检测以及渲染绘制。

基于深度的虚实遮挡处理方法通常首先计算场景图像上每个像素点的深度信息, 然后根据观察者的视点位置、虚拟物体的插入位置以及求得的深度信息等, 对虚拟物体与真实物体的空间位置关系进行分析. 如果虚拟物体被真实物体遮挡, 则在显示合成场景图像时只绘制虚拟物体中未被遮挡的部分, 而不绘制被遮挡的部分. Yokoya 等人^[26] 提出利用立体视觉设备估计真实场景中的物体深度信息, 而后根据观察视点位置和所估计的深度信息完成虚实物体的遮挡处理. 为了减小运算量, 该方法将立体匹配仅局限在虚拟场景在当前图像中的投影区域内, 其存在的问题是容易导致系统的运算速度随着虚拟场景在图像上投影面积的改变而变得不稳定. 此外虚拟物体与真实场景交界处会产生较为明显的遮挡失真现象, 难以获得令人满意的计算精度. 为了保证计算量的稳定性, Fortin 和 Hebert^[27] 根据场景物体到观测视点的距离将场景划由远而近分成多个区域, 从而处理虚实物体的遮挡, 如图 S4(a) 所示. 而 Hayashi 等人^[28] 在工作场景中布置数量较多的标识块辅助区域定位, 提出了一种基于轮廓的实时立体匹配方法, 能快速而准确地获得真实物体轮廓的深度信息, 不过对标识块的部署要求较高, 如图 S4(b) 所示. 受场景深度捕获算法提取精度和速度的限制, 目前增强现实中的虚实遮挡技术还只能完成简单形状的遮挡关系。

真实物体的深度或模型被获取后, 在增强现实中, 除遮挡关系外, 还需要考虑真实物体对虚拟物体的交互, 主要表现为碰撞检测. 当一个虚拟物体被人为操纵时, 需要能够检测到它与真实世界中物体的碰撞, 产生弹开、力反馈等物理响应. 现有的增强现实研究大多将碰撞检测作为算法验证, 大多精简快速. Salcudean 和 Vlaar^[29] 提出了适用于单点交互的基于高阻尼的接触模型, 根据刚性物体的硬度来模拟物体的冲击反馈. Constantinescu 等人^[30] 利用 Poisson 公式提出硬度可变的接触模型, 并有效模拟了虚拟物体与平面刚体的接触碰撞过程. Moore 和 Wilhelms^[31] 提出了利用单点碰撞的序列组合模拟多点碰撞, 并通过解析算法求虚实物体在碰撞过程中的冲量和接触力. Baraff^[32] 则进一步分析了法向加速度与接触力和摩擦力的关系, 并模拟了二维结构虚实物体的碰撞过程. 由于增强现实的深

度捕获精度还较低, 只能用于实现地形匹配、简单碰撞的效果, 随着深度捕获设备的发展, 将模型和稠密点云相结合的碰撞检测将可能成为重要的研究点. 例如 Leap motion 用稠密点云驱动高质量三维手部建模, 可以实现精确的碰撞检测, 这将提高增强现实在虚拟装配等领域的应用能力.

可以预计, 高精度的深度捕获设备及其相机跟踪算法仍将是增强现实领域研究的重点问题, 它的突破将促进增强现实技术的快速实用化.

5 增强虚拟环境

虚拟环境的建模存在不少限制, 例如建立真实环境的精确模型需要耗费大量人力, 建模形成的庞大数据库难以及时更新或修正, 纹理来自于事先采集, 不能反映真实环境的动态情况等. 增强虚拟环境技术出现的初衷就是为了解决这些问题, 但目前进展来看解决得还不够好.

将增强现实与增强虚拟环境两者进行比较, 增强现实以个人获取的真实世界图像为基础, 让虚拟对象适应用户视点或摄像头的运动变化, 因此在本质上是面向个人的, 适合于支持交互; 而增强虚拟环境以虚拟环境为基础, 通过三维注册让不同地点的 2D/3D 视觉采集实时融合进虚拟环境. 虚拟现实的三维绘制本身就可以是视点相关的, 因此在本质上是面向空间数据的, 适合于建立应用服务. 从这个观点出发, 增强虚拟环境技术更需要和网络结合起来发挥价值. 可以发现, 现有的相关研究确实大多都是以网络系统为基础, 如前述的远程呈现、远程沉浸等. 以下从基于视频图像的增强虚拟环境技术、基于三维角色的增强虚拟环境技术、虚实场景融合以及网络传输等方面对增强虚拟环境技术进行综述.

5.1 基于视频图像的增强虚拟环境技术

增强虚拟环境技术最直接的想法就是利用相机捕捉真实对象的图像或三维模型, 并将图像或三维模型实时注册到虚拟环境中, 使增强后的虚拟环境能够表示真实对象的状态和响应交互. 通过视频图像增强的方法最早是 Katkere 等人^[33]在 1997 年提出, 他们认为视频信息可以用来创建沉浸式的虚拟环境, 进而实现多视频流的有效分析, 进行视频不能提供的操作, 如变换新的任意虚拟视角的视频等. 美国 Sarnoff 公司的 Sawhney 等人^[34]发展了这种想法, 不再用视频创建虚拟模型, 而是用视频去增强已有虚拟的模型. 传统的监控系统采用二维堆叠显示大量视频流, 而他们的 Video Flashlights 系统首次尝试把实时视频的图像作为纹理, 实时映射到静态三维模型, 并在图形硬件的帮助下将多个已标定相机的视频进行统一实时渲染. 这种把多个视频注册到同一个三维环境的尝试, 使得用户能够以一个全局的视角统一观察模型和视频, 扩展了用户的视域, 增强了视频的空间表现力.

2003 年南加州大学 Neumann 等人^[35]在 IEEE VR 会议上系统阐述了增强虚拟环境的概念. 从表现未来虚拟城市的角度出发, 将实时采集的图像数据投影到地形和建筑数据之上, 实现了随着图像数据变化的动态三维模型效果, 如图 S5(a) 所示. 由于物体的深度不正确, 在非相机视点可能存在贴图扭曲, 他们提出了一种方法, 对运动对象 (行人、车辆等) 进行检测提取, 建立少量三角面片在估计深度上进行 billboard 贴图^[36]. 他们还指出了 AVE 的几个基本用途, 如场景理解、运动物体跟踪、事件检测等. 在 2010 年 ACM Multimedia 会议上, 麻省理工学院的 de Camp 等^[37]设计了一套用于智能家庭的沉浸式系统 HouseFly, 使用了家居的 3D 模型, 把鱼眼相机用于 AVE 系统. 用户在观察时, 看到的是投影到 3D 模型上的沉浸式多路音视频, 可以方便地确定视频中的人物位置和行走路线等, 如图 S5(b) 所示. 在 2009 年 ISMAR 会议上, 乔治亚理工学院 Kim 等^[38]提出了使用动态信息增强 Google

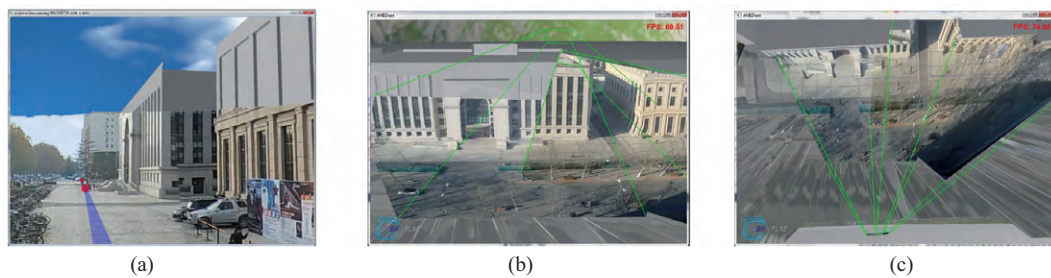


图 5 校园区域的立体监控平台

Figure 5 Campus scope 3D surveillance system. (a) Panorama fusion; (b) 3D/video fusion of mono-camera streaming; (c) Bird's eye view

Earth 等航拍地球地图的方法, 对视频进行分类处理和增强显示, 还在识别自然现象图像并进行图形绘制方面进行了尝试, 如图 S5(c) 所示. 2012 年台湾国立大学的 Chen 等^[39] 建立了 GIS 辅助的可视化框架, 融入了多分辨率监控策略, 以固定视角的相机提供低分辨率图像, 球基相机根据用户交互提供兴趣区的高分辨率图像. 这些工作在面向三维地图的空间数据可视化方面进行了尝试, 说明这类增强虚拟环境技术能够应对超大型地理应用系统的需求, 有着重要的应用潜力.

虚拟空间的相机标定与真实空间的标定不同, 虚拟空间是真实空间的逼真展示, 其场景的模型已经能够替代真实场景作为标定的输入信息的一种, 并配合当前相机采集到的图像, 进行快速的相机精确标定. Abrams 和 Pless^[40] 针对网络摄像头, 尝试了一种使用少量相关点约束的几何标定方法. 基于此方法, 将网络摄像头的图像作为纹理实时地展现在三维几何模型上, 简单实用, 但限于精简模型顶点和固定点数量不多, 渲染效果不够理想.

随着技术的进步, 出现了一些商业的 AVE 产品和方案. 在以上技术的基础上, 现有的产品技术都是先通过三维扫描仪采集城市的 LiDAR 地形数据, 建立三维地形模型, 然后进行相机注册, 结合现有的相机流和存储方案, 输出为视频投影融合显示. 例如美国 Airborne 1⁸⁾, SentinelAVE 等公司在亚特兰大的商业区采用了 42 个相机数据在三维扫描建立的虚拟场景中进行融合绘制. 国内也有一些单位在视频投影融合领域开展研究, 针对视频监控领域进行产品研发.

我们也在这方面开展了工作, 合作设计研发了全景相机软硬件系统, 在校园区域建立上百个相机作为输入的虚实混合实时视频监控平台, 开发支持多类相机的流媒体服务器, 对视频进行同步传输处理, 各类视频采集设备以及离线视频被抽象成虚拟相机, 动态挂载在不同的流媒体服务器之上. 提供快速的相机标定工具, 在街景视图下, 展现以 360° 全景球为主, 普通视频投影为辅的虚实融合效果, 如图 5 所示.

5.2 基于三维角色的增强虚拟环境技术

基于视频图像的增强虚拟环境技术主要解决的是多路视频流的时空理解和可视化问题, 但视频图像本身还是二维的, 所以无法交互, 而且在视觉效果 (特别是浏览视点) 上存在很多限制, 在非相机位置的虚拟视点会上存在图像拉伸、扭曲等变形现象. 一些工作通过实时三维重建技术将真实世界的对象更好地合成到虚拟场景中, 突破三维视觉、交互等方面的限制. 典型的例子是远程呈现或远程沉浸系统里加入的实时重建的虚拟对象, 如虚拟物体、人体姿态、面部表情等.

8) <http://panopticon3d.com/>.

在这类增强虚拟环境中, 目前已经实用的技术是采用体感交互的虚拟角色代替人体, 参与到虚拟环境中, 与虚拟环境或者其他虚拟角色进行交互. 人体姿态的实时三维获取与恢复是此类方法的难点. 微软在 2010 年推出的 3D 体感相机 Kinect 为人体姿态的获取提供了一种无 “marker” 的快速廉价解决方案^[41], 获取的姿态可以通过骨骼绑定等多种方式实时地映射到虚拟角色上. 清华大学的 Ye 等人^[42] 随后又利用多 Kinect 克服了复杂场景下的遮挡问题, 多个虚拟角色可以同时进行捕捉与映射. 但是由于自身的精度不足等问题, Kinect 无法精确地捕捉手势^[43], Leap 3D 的出现弥补了此类问题, 其动作识别精度可达到 0.01 mm⁹⁾. 微软在 2011 年 7 月推出了在线视频聊天程序 Avatar Kinect, 可以使用 Kinect 的传感器来记录面部表情和嘴唇动作, 并将其映射到用户的个性动画头像中, 动画头像跟随用户的动作和口型作出相似的反应, 这些功能已经出现在 Xbox 360 游戏和新版 Kinect for windows 中¹⁰⁾.

向虚拟场景中加入 “三维虚拟化” 的真实物体是一个更难的问题. 虚拟角色增强了人体的交互能力, 但在表现效果上不够真实, 特别是缺乏实时反映真实对象状态的动态纹理. 如果没有实时重建的三维模型, 物体图像像纹理一样直接投影在虚拟模型表面, 出现压扁扭曲和部分重叠的情况, 真实感大为下降. 最新的远程再现/远程沉浸技术是基于三维角色重建的增强虚拟环境技术, 他们采用多个相机从不同角度采集一个区域内的物体并实时三维重建, 将位于不同地点的用户或物体模型通过网络传输共享在同一个三维虚拟环境中, 其主要特点是实现无 “marker” 点的人体自由三维交互, 多个用户可以突破物理空间限制, 在所应用的虚拟环境中协同操作.

早在 20 世纪末, 研究者就开始尝试将真人进行三维重建并加入到虚拟环境中, 以获得更加真实的沉浸效果. 最早的一个成功实例是 1998 年卡内基梅隆大学设计的用于捕获多人交互场景和动态事件建模的 3D Dome 原型系统^[44]. 多个摄像机布置在各个角度, 对处于中心工作区域的物体进行视频采集, 采用立体视觉匹配方法对已录制好的视频序列进行离线建模, 可实现动态事件的多视点观看.

2000 年 5 月, UNC 教堂山分校和宾夕法尼亚大学演示了一种基于三目立体匹配的远程协作系统, 系统采用了 7 个 Sony DFW-V500 的 1394 相机, 互相重叠成三目来使用, 然后进行背景剔除, 在四核的 Dell 服务器上立体计算的并行化, 计算速度与传输速度明显提高^[45]. 在 UNC 的系统里, 用户能够同时观察到两个异地参与者, 来自远程的三目深度视频流在沉浸式显示设备中被实时合成, 两个显示屏幕分别显示来自不同位置的另外参与者, 每个屏幕都带有一组投影机, 能够提供立体显示. 头部跟踪器提供观察者的头部位置信息, 保证正确渲染用户视点所对应的虚拟场景.

法国 INRIA 的 GrImage 课题组于 2007 年研究设计了基于三维重建的无 “marker” 虚实交互系统 GrImage^[46]. GrImage 采用 6 个 PointGrey 工业摄像机在 1 m × 1 m × 1 m 的空间内, 通过可视外壳算法 EPVH 实现了对无标定物体的实时建模, 并基于 INRIA 的 FlowVR 仿真平台进行模型嵌入和虚拟环境仿真, 实现真实物体与刚体、弹性体或流体等虚拟对象的实时交互. 该系统及其后续版本相继在 SIGGRAPH 2007, ECCV 2008, VRST 2008, 法国电视频道 TF1, SIGGRAPH 2009, ACM Multimedia 2010 等会议上展示, 获得了大量关注和好评.

加州大学伯克利分校 (UC Berkeley) 从 2005 年开始一直从事远程沉浸方面的研究. 他们建立了由 48 个相机覆盖的采集空间, 以 4 个相机为一个相机簇, 采用基于立体视觉的三维重建技术对局部采集环境进行实时三维重建. 利用 12 台主机组成一个小型的计算集群, 负责进行分布式的三维重建计算^[47].

在以上三维采集和交互环境研究的基础上, UC Berkeley 与 UIUC, UC Davis 合作建立了异地多

9) <https://www.leapmotion.com/>.

10) <http://www.xbox.com/en-us/kinect/avatar-kinect/>.

点三维远程沉浸 (3DTI) 协同环境 TEEVE. 位于不同地域位置的三个用户通过 TEEVE 系统实现了远程共享, 任一用户均可在自己的终端上看到三人的重建结果及他们与虚拟环境的交互场景. 他们在远程太极拳教学、心理学测试、远程协同舞蹈、远程博物馆探索、地质三维可视化远程研讨分析和远程手术训练等方面开展了很多应用尝试¹¹⁾.

2011 年 10 月, UNC 设计了基于 6 个 Kinect 体感相机的远程再现系统 (Encumbrance-Free Telepresence)^[48], 该系统使用 5 个 Kinect 从不同角度进行人体数据采集, 并利用 GPU 对 5 个 Kinect 所采集的深度数据进行并行处理, 生成真实环境的三维模型和人体三维模型, 另外 1 个 Kinect 用于对远程用户进行跟踪. 该系统能够根据远程用户的位置实现人体三维模型的真 3D 显示, 实时地从 Kinect 所采集的数据生成真实环境和人体的三维模型, 并在真实场景中加入虚拟物体, 实现了用户与虚拟物体的简单交互.

5.3 视频融合

视频融合是实现虚拟环境真实感显示的关键, 但大多数现有虚拟环境系统的图形绘制只是融合静态三维模型和真实光场信息进行真实感渲染. 这方面最早是 Debevec^[49] 所做的开创性工作, 提出手工将三维拓扑和图像内容进行映射, 利用建筑的三维几何结构对照片进行视点相关的绘制, 实现了非相机视点的真实感漫游效果, 但当时并未实现真实的三维空间标定关系.

传统的纹理映射要求纹理坐标与三角形顶点是有先验关联, 必须在渲染之前进行指定. 与这种图像到模型的方法相反, AVE 系统使用的是模型到图像的方法. 视频投影只需要提取获知相机所在的位置和相机的图像, 即可实时地计算模型上每个顶点的坐标. 当相机的位置发生变化, 或者视频中的内容发生变化, 可以直接或间接地计算这种模型到图像的映射关系. 投影纹理映射^[50] 用于映射一个纹理到物体上, 就像将幻灯片投影到墙上一样. 虽然该方法主要用于一些阴影算法以及体绘制算法中, 但是它在计算机视觉, 基于图像的渲染和三维可视化等其他领域也有很大的用途.

同时, 因为图像应该只映射到相机可见的区域, 所以必须判断图像投影时的可见性信息. 阴影贴图方法 (shadow maps) 可以有效地提供可见性信息, 在阴影里则为不可见区域, 不在阴影的区域则为纹理映射的区域. 另一种方法是阴影体方法 (shadow volume), 也可以获得投影可见性, 而且可以避免阴影贴图技术因深度精度不足而造成的锯齿问题.

在 AVE 系统中, 一个区域可能会有多个相机同时可见, 也就是出现了部分纹理重叠 (overlap) 的问题. 麻省理工学院的 de Camp 等^[37] 对虚拟空间划分区域, 每个相机对应一个区域, 避免一个区域有多相机同时可见的情况. 台湾国立大学的 Chen 等^[39] 在视频进行投影之前, 对所有图像先进行拼接, 得到两两相机之间的图像拼接关系; 并定义一个二维的相机地图, 地图里的每个点只对应一个相机. 这些方法通过手工选择避免了纹理混合, 还可以通过视点相关的纹理融合实现自动的混合显示. Harville 等^[51] 提出了一种实现多纹理混合的简单易用方法. 该方法把投影仪在重叠区域的混合因子设置成该点与图像最近边界的距离, 距离边界越远, 混合因子越大, 在边界上实现淡出淡入的效果, 避免图像的不连续现象.

5.4 网络传输

增强虚拟环境技术比增强现实技术更需要和网络结合来发挥价值. 在网络应用中, AVE 系统不仅要为用户提供虚拟环境本身的静态精细模型, 还会传输根据视频等传感器信息新生成的虚拟对象模型,

11) <http://tele-immersion.citris-uc.org/>.

因此涉及到实时修正与变形的动态模型. 这涉及到时间相关的三维动态模型的流式传输, 传统的基于状态参数的分布式虚拟环境技术无法处理类似问题^[52].

动态模型的大数据量源于纹理, 每个虚拟视点的纹理映射都可能涉及到多路实时视频流, 对于视频流来说, 都是常规的视频流选择与流式传输技术, 在数据量上很难做到进一步的实时降低. 另一方面, 就时间序列化的几何模型来说, 基于三维角色重建的增强虚拟环境技术还未能探索动态几何模型的时间相关性, 目前的实时三维重建算法都只能处理孤立帧的模型计算与生成, 这就限制了动态几何模型的压缩与传输技术发展. UC Berkeley 早期研究基于点云模型的压缩技术^[53], 后来与 UIUC 合作研究 RGBD 的压缩方法^[54], 压缩比和质量并不理想, INRIA 并没有进行压缩^[55], 可以看出, 在实时三维重建技术或网络带宽取得突破之前, 动态模型的实时压缩与传输难以提高. 然而, 鉴于近期基于三维角色的增强虚拟环境技术采用体感交互设备发展迅速, 变形模型的实时流式传输可以视为动态模型传输的一个“弱形式”, 特别在移动终端和云计算兴起的时期, 其研究对增强虚拟环境中虚拟角色的物理真实感提高有着重要价值.

自 1996 年 Hoppe 提出渐进网格传输技术后, 出现了大量的几何模型渐进传输算法, 但是渐进压缩技术只能支持对静态模型的流式传输, 因为模型的顶点删除和分裂方式与表面有关, 当模型表面发生形变后, 原始记录不再有效. 近年来, 德州大学达拉斯分校在变形模型的交互共享方面开展了一系列研究. 2010 年 Tang 等^[56] 基于谱变换方法, 将网格在空间域上表示为基矩阵和一系列多分辨率变换矩阵, 可以用于渐进传输和恢复, 只传输变形后的谱系数而非完整模型, 大大降低了网络开销, 初步实现了变形模型的流式传输. Tang 等^[57] 进一步将实时变形模型引入到低计算性能的移动设备上, 移动用户可在触摸屏上操作来改变模型的形状和运动. 为加快运算速度并减少通讯负载, 算法使用谱表示方法计算三维变形, 同时提出了一个渐进的变形流传输技术. 受固有变形算法和多分辨率映射算法的限制. 该方法只能支持小型网格的变形传输, 并且由于模型分辨率是固定的, 无法实现不同终端的帧速保障.

本文作者课题组也在变形模型的渐进传输方面开展了系统的研究, 结合几何频谱和渐进网格技术, 给出一种基于“分解—重建”过程的动态多分辨率变形模型传输方法. 首先提出了一种基于微分坐标的特征保持模型简化算法, 对模型进行精确, 可控的多分辨率细节保持简化^[58], 然后基于谱变换与渐进几何表达技术, 实现对三维模型变形的实时多分辨率共享传输^[59]. 接收方在同步显示模型形变过程中, 可以任意改变模型分辨率而不影响变形传输本身, 因此可在接收方实现动态帧速控制, 使变形质量可以适配终端能力. 图 6 所示为变形模型传输效果. 图 7 所示为变形传输过程中的动态分辨率效果.

在图 6 中, 心脏模型为交互式模型变形在虚拟医疗研讨中的应用, 模型顶点数为 7349 个, 远程用户在共享的虚拟环境中对柔性心脏模型进行交互式变形操作, 并实时共享心脏变形结果; 象模型为离线模型动画序列的实时传输效果, 模型顶点数为 42321 个, 各网络节点在无需本地缓存的情况下, 采用不同的模型分辨率实时同步共享模型动画. 图 7 中模型顶点数目为 8431 个, 在没有采用动态分辨率的情况下 (图 7(a)), 动画帧速只能达到每秒 6 帧左右, 而采用动态分辨率之后, 系统根据当前动画帧速自动调整模型分辨率 (图 7(b)), 尽量简化远离视点位置的模型分辨率, 从而减少计算和渲染开销, 提高动画帧速.

近几年, 云计算成为互联网领域中的研究热点, 它可将存储和复杂的计算从客户端转移到云计算服务环境, 为复杂的交互式物理特效计算提供了可能的分布式架构模式. 从这个特点出发, 本文作者提出了一种基于云计算服务的变形模型远程编辑方法^[60], 在客户端进行低分辨率的编辑控制, 由云服务进行对应操作在完整分辨率上的控制顶点序列的权重计算, 返回给客户端后即可实现实时的编辑变

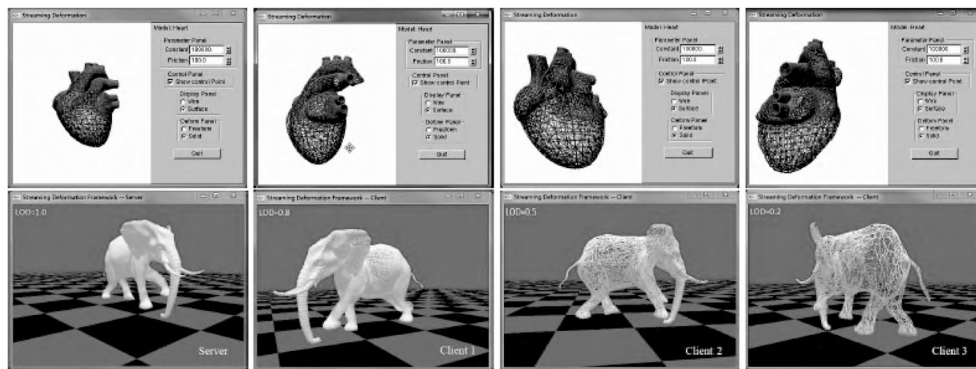


图 6 交互式模型变形 (心脏) 和模型动画 (象) 实时传输效果图

Figure 6 The interactive deform of a human heart and realtime transmission of model animation

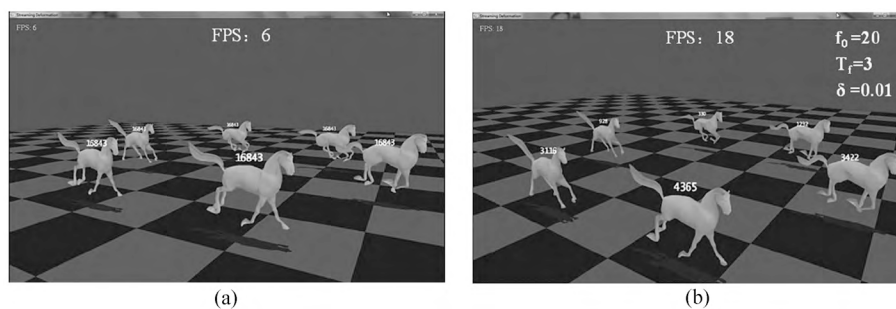


图 7 动态多分辨率渲染与完全渲染效果对比

Figure 7 Dynamic multi-resolution render and the complete render. (a) Without frame rate control (6 fps); (b) with frame rate control (18 fps)

形, 基于云服务的模型编辑流程如图 S6 所示.

在上述基于云服务的变形计算模式下, 用户在对替身模型进行变形操作时, 无需等待目标高精度模型的变形反馈结果, 只需将用户操作句柄通过网络传输到云端, 然后由云服务器在后台通过用户操作队列对目标模型进行变形计算^[60]. 与传统所见即所得的变形编辑方式相比, 这种基于云服务的模型编辑方式实现了用户交互编辑与目标模型变形结果的分离, 既保证了用户客户端编辑操作的交互式呈现, 又实现了对高精度大型模型的变形编辑, 可作为云计算模式下虚拟现实技术应用的一个参考.

6 移动互联网上的相关技术与应用

随着移动互联网的兴起, 一些研究和应用开始将虚实增强技术引入到移动互联网上. 在三维重建的基础上, INRIA GrImage 课题组在 ACM Multimedia 2010 上展示了在法国三个城市之间进行的远程再现原型, 多个位于不同地点的用户 (包括移动电脑用户) 能够共享同一个虚拟环境^[61]. 如图 S7 所示, 两个是以重建出的三维模型的方式加入到共享虚拟环境里, 另一个则是通过二维视频的方式加入. 三个用户通过不同终端设备进行交互, 展示了一个具有移动接入支持的多点远程沉浸系统.

UIUC 的 Klara Nahrstedt 课题组提出基于颜色-深度 LOD 的 DIBR 算法, 并在远程沉浸系统中进行用户心理评估^[62], 另外还提出了移动终端的远程 DIBR 算法^[63], 通过两个参考视图可以有效地

消除三维图像变换所导致的空洞。

随着智能手机的崛起, 移动互联网上增强现实移动 APP 在几年之内飞速增长。这些应用都离不开标配的摄像头, 以及使用陀螺仪、GPS 等设备的辅助定位, 最终以视频图像叠加的方式展示给用户。下面对典型的几类应用 APP 进行介绍。

- 实景识别. Google 推出的图片搜索应用 Goggles, 它的功能是可以利用手机拍照的方式, 使用图像识别技术来识别地标、标识、条形码和二维码等, 也可以用于识别绘画作品、书籍、DVD 和 CD 等, 应用会将获得的识别信息, 通过虚实图像合成的方式有效地放置于图像范围内。

- 户型图绘制. 加拿大蒙特利尔的创业公司 Sensopia 在 2011 年发布了 MagicPlan。该应用通过 iPhone 或者 iPad 对室内进行拍摄利用相机和陀螺仪进行摄像头的跟踪, 在此基础上测量并绘制出房屋户型图。该 APP 综合使用相机跟踪技术与手持设备交互技术, 进而在由特征点建立起来的三维空间中进行测绘。

- 特效制作. Action Movie FX 是可以制作简单的好莱坞电影特效的虚实图像合成工具, 内置几种固定的特效, 如汽车翻毁、飞机坠毁、导弹打击等。该 APP 捕捉物体运动轨迹, 通过图像视频叠加的方法对现实进行特效增强。

- 结合位置服务的 AR 应用. Wikitude World Browser 是一款基于地理位置的增强现实应用, 可以通过指南针、摄像头和 GPS, 将虚拟信息数据标注到现实世界中。在陌生环境下, 用户开启 GPS 定位, 对想了解的地方进行拍摄, 服务器会返回这个地方的有用信息, 如酒店信息、景点名胜的特色图片和视频等。

- 现在已经出现了多款移动增强现实开发引擎或平台, 目前最著名的商业 AR 引擎是高通的 vuforia 平台和 metaio 等。Obvious Engine 是一款 iOS 上的增强现实框架引擎, 它使用物体表面的自然特征来进行相机跟踪, 其开发者介绍, 可以在大多数光照环境下达到每秒 30 帧的跟踪速度, 并且支持 OpenGL 和 Unity 3D 开发集成。

移动互联网上的增强现实 APP 的发展已经进入高速发展阶段, 各种层出不穷的创意吸引着消费者进行下载使用, 但目前的增强现实应用还非常有限, 受制于屏幕、摄像头等设备的进一步更新发展。在未来几年, 各种商业创意、3D 屏幕和深度相机的加入将可能推动增强现实 APP 的大发展。

7 相关国际标准制定进程

ISO/IEC JTC 1/SC 24 (国际标准化组织/国际电工委员会第 1 联合技术委员会第 24 分委会) 的全称是计算机图形、图像处理和环境数据表示分技术委员会, 目前拥有 10 个 P 成员国和 23 个 O 成员国, 其中 P 成员国包括澳大利亚、中国、埃及、法国、俄罗斯、日本、韩国、葡萄牙、英国、美国。根据 JTC1 N7752 文件, SC 24 的工作范围主要是与计算机图形、图像处理、虚拟现实、信息交互及信息可视化表示这几个领域有关的信息技术应用接口的标准化, 以及建模与仿真相关参考模型、交互格式、编码等规范。近几年 SC 24 制订的重要标准包括 X3D (ISO/IEC 19776 系列、19777 系列), SEDRIS (ISO/IEC 18023 系列、18024 系列) 等。根据工作规范, SC 24 原下设三个工作组 (Work Group, WG): WG6 多媒体表示/交换, WG7 图像处理/交换, WG8 环境数据表示)。鉴于虚拟现实相关技术的快速发展, 2011 年 8 月, 在美国 Rapid City 举办的 SC 24 全会经过讨论, 考虑增加 WG9 专门进行虚拟现实/增强现实/混合现实领域的标准化工作, 并且把 SC 24 原来工作范围中的“虚拟现实”(VR) 扩展为“虚拟现实/增强现实/混合现实”(VR/AR/MR)。其中韩国代表团提交了正式的新名称, 范围的定义及

详细阐述¹²⁾, 中国及美国代表团提出了对名称和范围的具体建议, 特别是中国代表团建议用“虚拟现实连续统一体”(VR Continuum)代替冗长且有含义交叉的“虚拟现实/增强现实/混合现实”。最终经过审议, SC 24 采用了“增强现实连续统一体”(AR continuum)的名称, 最终 WG6 修改为“增强现实连续统一体表示和交换”(AR continuum presentation and interchange), 新成立 WG9 工作组, 并命名为“增强现实连续统一体概念和参考模型”(AR continuum concepts and reference model)。

2012 年 8 月 19 日至 24 日, SC 24 全会在比利时 Brussels 举行, 各国代表团积极提出多项与 ARC 有关的新提案。新提案的重要评价之一是需要具有足够多产业界人士的兴趣, 最终全会批准以下 3 项新工作提案进入初始阶段投票状态: (1) SC 24 N 3411 增强现实连续统一体参考模型; (2) SC 24 N 3414 用于物理传感器的增强现实连续统一体参考模块; (3) SC 24 N 3415 用于真实特征表示的增强现实连续统一体参考模块。各国在会后组织了答票和提交建议。这三项 NP 提案最初是韩国提出的, 在全会中我国补充提出了 ARC 术语和概念定义及术语制定规则, 也写入了投票文件, 在投票阶段还建议了参考模型完善, 部分传感器接口与格式完善等。在 2012 年底, SC 24 公布了 10 个 P 成员国的投票结果, 通过了增强现实连续统一体的三个新提案, 进入了工作程序。

增强现实连续统一体参考模型 (ARC-RM) 定义了一种结构 (如体系、功能、信息/计算可视化), 用于当前和未来的增强现实和混合现实领域国际标准对比及描述相关关系。该参考模型定义了一系列规则、概念以及内部关系, 应与未来增强现实、混合现实标准的总范围相适应。包含以下内容: 规则, 增强现实连续统一体术语及定义, 用例, 需求, ARC 体系结构 (基于网络或独立), 功能和基础组件, 组件之间的接口及数据流, 抽象层, 与其他标准的关系等。

ARC 系统的通用参考模型如图 S8 所示, 包括主要组件及功能性和组件接口 (数据和控制)。可以看出, 目前的 ARC 模型更重视从相机采集, 然后进行跟踪和识别, 通过消息机制发送给虚实合成模块, 最终完成绘制和显示。这符合目前产业界对增强现实技术应用在 PC 或移动互联网上的强烈需求: 仅配备简单的相机即可实现虚拟现实增强的显示, 需要模型结构简单, 跟踪以实现精确的空间定位, 识别后通过简单的消息机制进行驱动。

ISO/IEC JTC 1/SC 29 (coding of audio, picture, multimedia and hypermedia information, 音频、图像、多媒体和超媒体信息) 技术分委也对增强现实/混合现实技术表现出强烈的兴趣, 正在制定增强现实应用格式标准 Multimedia application format (MPEG-A)-Part 13: Augmented reality application format¹³⁾。2013 年, 在 ARC 新提案的基础上, SC 24 和 SC 29 建立了联合工作机制, 首先由 SC 24 WG9 和 SC 29 WG11 共同推动 ARC 参考模型的标准。由于双方共同的兴趣, 将标准进一步集中在混合和增强现实 MAR (Mixed and Augmented Reality), 也就是虚拟现实增强技术, 从中也可以看出工业界对于虚拟现实增强技术的期待。

2013 年 5 月, 在 SC 29 WG 11 发布的 ISO/IEC JTC 1/SC 29/WG 11 N13613 的基础上, ISO/IEC JTC 1/SC 29-24/WG 11-9 启动了技术草案报告 “WD 1.0 of Mixed and Augmented Reality Reference Model”, 并计划了 2013 年中、年底, 2014 年初的三次讨论会。在目前的草案中, 混合和增强现实 MAR 被定义为一种基于采集的真实世界环境的表示, 其中的元素以计算机驱动的方式被图形、声音等内容所增强, 如图 S9 所示。混合和增强现实系统实现了对最终用户内容同步驱动的虚拟环境内容的实时访问。

ISO 要求有关文件的非成员组织使用需要书面许可, 因此本文不对参考模型等草案的详细内容进

12) ISO/IEC JTC 1/SC 24 N3280. Gerard Kim, Richard Puk, “Reorganization of SC 24 to emphasize VR/AR/MR”.

13) ISO/IEC FDIS 23000-13, 2013 年 8 月 8 日发布进展处于 50.0 阶段。

行介绍, 部分内容可到 ISO 网站¹⁴⁾查询. 在参与标准工作的过程中, 我们有两点认识:

(1) “虚拟现实/增强现实/混合现实”(VR/AR/MR) 得到了工业界的普遍认可, 特别是增强现实已经具有了强大的产业支持, 应该重视跟进标准的制定;

(2) 各国代表团积极在增强现实连续统一体方面提出建议, 韩国代表团的积极主导是跟韩国政府和大企业的重视和支持是离不开的, 中国目前的关注程度还不够, 建议更多的学术界和产业界人士积极参与和建议.

8 总结与展望

近年来, 虚拟现实增强技术已经取得了显著的技术进步, 特别是在产业界的普及型需求和积极推动下, 展示出强劲的发展前景. 本文分别从技术特点、相关硬件设备、增强现实、增强虚拟环境这 4 个方面, 阐述和分析了虚拟现实增强技术的研究进展. 鉴于移动互联网时代的兴起, 本文介绍了移动互连网上已经出现的虚实增强技术与应用, 最后还介绍了相关国际标准的制定进程.

增强现实技术已经得到了学术界和产业界的广泛关注, 在硬件装置、跟踪精度、虚实合成等方面正在快速发展和提高. 例如在 2013 年的增强世博会 (Augmented World Expo) 上, 绝大多数参展商都集中在 SLAM 系统跟踪技术和眼部穿戴式增强设备上, 他们期望出现像 3D 眼镜 Atheer One 这样高度集成的便携式设备, 能够刺激现有增强现实技术的快速应用. 同时, 新型产品的出现也会给研究者带来更多的研究内容, 促进增强现实技术向着更高的方向发展.

增强虚拟环境技术发展缓慢, 还未能像增强现实一样被大众认可. 不同于增强现实技术, 该技术使用的“画布”而非逼真的三维虚拟环境, 难以用二维视频图像等信息达到较好的增强效果. 现有的方法除了存在采集装置、建模精度等多方面的关键技术问题, 还无法达到逼真的合成效果, 例如虚实物体的正确遮挡关系, 虚实光照效果的融合处理, 以及多传感器数据融合的理论问题等. 增强虚拟环境技术还需要深入研究, 有待于以上关键技术的突破和新的硬件设备推动.

总的来说, 虚拟现实增强技术的理论研究、系统开发和应用推广等方面有着重大进展, 尤其是移动互联网时代的到来, 为增强技术带来了广阔的应用群体和飞速的技术进步. 同时, 虚拟现实增强技术的不均衡的发展也为该领域带来了大量有待解决的问题和难题.

今后的几年里, 虚拟现实增强技术的发展将会引入越来越多新的思路. 例如, 本文作者在基于云计算服务的变形编辑方法方面所做的工作, 从一个侧面说明云计算服务可能对虚拟现实相关技术提供新型的分布式计算模式; 目前脑机接口和外骨骼系统的研究可能为虚拟现实相关技术提供新的增强表达方法. 虚拟现实增强技术还有很长的一段路要走, 期待它能够实现现实世界与虚拟世界的无缝融合, 改变普通生活 and 传统行业, 成为人类访问世界的新途径.

致谢 中国电子标准研究所刘棠丽博士、高麟鹏主任等帮助修订了本文关于增强现实国际标准部分的内容, 为本文的完成提供了大量的标准文件、会议总结和新提案文件等, 并对本文提出了宝贵意见, 特此致谢.

补充材料 S1~S9. 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.

14) <http://www.iso.org>.

参考文献

- 1 Zhao Q P. A survey on virtual reality. *Sci China Ser-F: Inf Sci*, 2009, 52: 348–400
- 2 Azuma R, Bailiot Y, Behringer R, et al. Recent advances in augmented reality. *Comput Graph Appl*, 2001, 21: 34–47
- 3 Bimber O, Raskar R, Inami M. *Spatial augmented reality*. Wellesley: AK Peters, 2005
- 4 Raskar R, Welch G, Low K L, et al. Shader lamps: animating real objects with image-based illumination. In: *Proceedings of the 12th Eurographics Workshop on Rendering Techniques*, Vienna, 2001. 89–102
- 5 Zhou F, Duh H B L, Billinghurst M. Trends in augmented reality tracking, interaction and display: a review of ten years of ISMAR. In: *Proceedings of the 7th IEEE International Symposium on Mixed and Augmented Reality*. Cambridge, 2008. 193–202
- 6 Gere D S. Image capture using luminance and chrominance sensors. US Patent, 8 497 897, 2013-7-30
- 7 Leininger B. A next-generation system enables persistent surveillance of wide areas. *Defense Secur*, 2008
- 8 Leininger B, Edwards J, Antoniadis J, et al. Autonomous real-time ground ubiquitous surveillance-imaging system (ARGUS-IS). In: *Proceedings of SPIE Defense and Security Symposium*, Orlando, 2008. 69810H
- 9 Brady D J, Gehm M E, Stack R A, et al. Multiscale gigapixel photography. *Nature*, 2012, 486: 386–389
- 10 Bimber O, Raskar R. Modern approaches to augmented reality. In: *Proceedings of ACM SIGGRAPH 2006 Courses*. Boston: ACM, 2006. 1
- 11 Han J, Shao L, Xu D, et al. Enhanced computer vision with microsoft kinect sensor: a review. *IEEE Trans Cybernetics*, 2013, 43: 1318–1334
- 12 Jones A, McDowall I, Yamada H, et al. Rendering for an interactive 360 light field display. *ACM Trans Graph*, 2007, 26: 40
- 13 Blanche P A, Bablumian A, Voorakaranam R, et al. Holographic three-dimensional telepresence using large-area photorefractive polymer. *Nature*, 2010, 468: 80–83
- 14 Davison A J. Real-time simultaneous localisation and mapping with a single camera. In: *Proceedings of IEEE International Conference on Computer Vision*, Nice, 2003. 1403–1410
- 15 Klein G, Murray D. Parallel tracking and mapping for small AR workspaces. In: *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, Nara, 2007. 225–234
- 16 Newcombe R A, Davison A J. Live dense reconstruction with a single moving camera. In: *Proceedings of Computer Vision and Pattern Recognition*, San Francisco, 2010. 1498–1505
- 17 Richard A, Newcombe R A, Steven L, et al. Dense tracking and mapping in real-time. In: *Proceedings of IEEE International Conference on Computer Vision*, Barcelona, 2011. 2320–2327
- 18 Tan W, Liu H, Dong Z, et al. Robust monocular SLAM in dynamic environments. In: *Proceedings of IEEE International Symposium on Mixed and Augmented Reality*, Adelaide, 2013. 209–218
- 19 Pollefeys M, Nistér D, Frahm J M, et al. Detailed real-time urban 3d reconstruction from video. *Int J Comput Vision*, 2008, 78: 143–167
- 20 Nistér D, Naroditsky O, Bergen J. Visual odometry. In: *Proceedings of Computer Vision and Pattern Recognition*, Washington, 2004. 652–659
- 21 Konolige K, Agrawal M, Bolles R C, et al. Outdoor mapping and navigation using stereo vision. In: *Proceedings of Experimental Robotics*. Berlin Heidelberg: Springer, 2008. 179–190
- 22 Zhu Z, Oskiper T, Samarasekera S, et al. Ten-fold improvement in visual odometry using landmark matching. In: *Proceedings of IEEE International Conference on Computer Vision*, Rio de Janeiro, 2007. 1–8
- 23 Zhu Z, Oskiper T, Samarasekera S, et al. Real-time global localization with a pre-built visual landmark database. In: *Proceedings of Computer Vision and Pattern Recognition*, Alaska, 2008. 1–8
- 24 Newcombe R A, Davison A J, Izadi S, et al. KinectFusion: real-time dense surface mapping and tracking. In: *Proceedings of IEEE International Symposium on Mixed and Augmented Reality*, Basel, 2011. 127–136
- 25 Steinbrucker F, Sturm J, Cremers D. Real-time visual odometry from dense RGB-D images. In: *Proceedings of IEEE International Conference on Computer Vision Workshops*, Spain, 2011. 719–722
- 26 Yokoya N, Takemura H, Okuma T, et al. Stereo vision based video see-through mixed reality. In: *Proceedings of IEEE International Symposium on Mixed and Augmented Reality*, California, 1999. 131–141
- 27 Fortin P, Hebert P. Handling occlusions in real-time augmented reality: dealing with movable real and virtual objects. In: *Proceedings of Computer and Robot Vision*, Quebec City, 2006. 54

- 28 Hayashi K, Kato H, Nishida S. Occlusion detection of real objects using contour based stereo matching. In: Proceedings of International Conference on Augmented Tele-existence, Christchurch, 2005. 180–186
- 29 Salcudean S E, Vlaar T D. On the emulation of stiff walls and static friction with a magnetically levitated input/output device. *J Dynamic Syst Measurement Control*, 1997, 119: 127–132
- 30 Constantinescu D, Salcudean S E, Croft E A. Haptic rendering of rigid contacts using impulsive and penalty forces. *IEEE Trans Robot*, 2005, 21: 309–323
- 31 Moore M, Wilhelms J. Collision detection and response for computer animation. *ACM SIGGRAPH Comput Graph*, 1988, 22: 289–298
- 32 Baraff D. Analytical methods for dynamic simulation of non-penetrating rigid bodies. *ACM SIGGRAPH Comput Graph*, 1989, 23: 223–232
- 33 Katkere A, Moezzi S, Kuramura D Y, et al. Towards video-based immersive environments. *Multimedia Syst*, 1997, 5: 69–85
- 34 Sawhney H S, Arpa A, Kumar R, et al. Video flashlights: real time rendering of multiple videos for immersive model visualization. In: Proceedings of the Eurographics Workshop on Rendering, Pisa, 2002. 157–168
- 35 Neumann U, You S, Hu J, et al. Augmented virtual environments (ave): dynamic fusion of imagery and 3d models. In: Proceedings of Virtual Reality Conference, Los Angeles, 2003. 61–67
- 36 Sebe I O, Hu J, You S, et al. 3d video surveillance with augmented virtual environments. In: Proceedings of ACM SIGMM International Workshop on Video Surveillance, California, 2003. 107–112
- 37 de Camp P, Shaw G, Kubat R, et al. An immersive system for browsing and visualizing surveillance video. In: Proceedings of the International Conference on Multimedia, Firenze, 2010. 371–380
- 38 Kim K, Oh S, Lee J, et al. Augmenting aerial earth maps with dynamic information. In: Proceedings of the IEEE International Symposium on Mixed and Augmented Reality, Seoul, 2009. 35–38
- 39 Chen S C, Lee C Y, Lin C W, et al. 2D and 3D visualization with dual-resolution for surveillance. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Providence, 2012. 23–30
- 40 Abrams A, Pless R. Web-accessible geographic integration and calibration of webcams. *ACM Trans Multimedia Comput Commun Appl*, 2013, 9: 8
- 41 Shotton J, Sharp T, Kipman A, et al. Real-time human pose recognition in parts from single depth images. *Commun ACM*, 2013, 56: 116–124
- 42 Ye G, Liu Y, Hasler N, et al. Performance capture of interacting characters with handheld kinects. In: Proceedings of European Conference on Computer Vision, Berlin Heidelberg, 2012. 828–841
- 43 Ren Z, Meng J, Yuan J, et al. Robust hand gesture recognition with kinect sensor. In: Proceedings of the 19th ACM International Conference on Multimedia, Scottsdale, 2011. 759–760
- 44 Narayanan P J, Rander P W, Kanade T. Constructing virtual worlds using dense stereo. In: Proceedings of the IEEE International Conference on Computer Vision, Freiburg, 1998. 3–10
- 45 Mulligan J, Kaniilidis K. Trinocular stereo for non-parallel configurations. In: Proceedings of the International Conference on Pattern Recognition, Barcelona, 2000. 567–570
- 46 Allard J, Menier C, Raffin B, et al. Grimace: markerless 3D interactions. In: Proceedings of ACM SIGGRAPH Emerging Technologies, San Diego, 2007. 9
- 47 Kurillo G, Bajcsy R. 3D teleimmersion for collaboration and interaction of geographically distributed users. In: Proceedings of Virtual Reality, 2013, 17: 29–43
- 48 Maimone A, Fuchs H. Encumbrance-free telepresence system with real-time 3D capture and display using commodity depth cameras. In: Proceedings of IEEE International Symposium on Mixed and Augmented Reality, Seoul, 2011. 137–146
- 49 Debevec P E, Taylor C J, Malik J. Modeling and rendering architecture from photographs: a hybrid geometry-and image-based approach. In: Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM, 1996. 11–20
- 50 Segal M, Korobkin C, van Widenfelt R, et al. Fast shadows and lighting effects using texture mapping. *ACM SIGGRAPH Comput Graph*, 1992, 26: 249–252
- 51 Harville M, Culbertson B, Sobel I, et al. Practical methods for geometric and photometric correction of tiled projector.

- In: Proceedings of Computer Vision and Pattern Recognition Workshop, New York, 2006. 5
- 52 Zhou Z, Wu W. Distributed virtual environment. Beijing: Science Press, 2009 [周忠, 吴威. 分布式虚拟环境. 北京: 科学出版社, 2009]
 - 53 Lien J M, Kurillo G, Bajcsy R. Multi-camera tele-immersion system with real-time model driven data compression. Visual Comput, 2010, 26: 3–15
 - 54 Wu W, Arefin A, Kurillo G, et al. CZLoD: a psychophysical approach for 3D tele-immersive video. ACM Trans Multimedia Comput Commun Appl, 2012, 8: 39
 - 55 Petit B, Lesage J D, Menier C, et al. Multicamera real-time 3d modeling for telepresence and remote collaboration. Int J Digital Multimedia Broadcast, 2010, 2010: 247108
 - 56 Tang Z, Rong G, Guo X, et al. Streaming 3D shape deformations in collaborative virtual environment. In: Proceedings of Virtual Reality Conference, Waltham, 2010. 183–186
 - 57 Tang Z, Ozbek O, Guo X. Real-time 3D interaction with deformable model on mobile devices. In: Proceedings of the ACM International Conference on Multimedia, Scottsdale, 2011. 1009–1012
 - 58 Zhang L, Ma Z, Zhou Z, et al. Laplacian-based feature preserving mesh simplification. Adv Multimedia Inf Process. Berlin Heidelberg: Springer, 2012. 378–389
 - 59 Zhang L, Dou F, Zhou Z, et al. Sharing 3D mesh animation in distributed virtual environment. In: Proceedings of the International Conference on Computer Animation and Social Agents, Istanbul, 2013
 - 60 Zhou Z, Dou F, Li Y, et al. GhostMesh: cloud-based interactive mesh editing. In: Proceedings of the 2013 Pacific-Rim Conference on Multimedia, Nanjing, 2013. 13–16
 - 61 Petit B, Dupeux T, Bossavit B, et al. A 3d data intensive tele-immersive grid. In: Proceedings of the International Conference on Multimedia, Firenze, 2010. 1315–1318
 - 62 Wu W, Arefin A, Kurillo G, et al. Color-plus-depth level-of-detail in 3D tele-immersive video: a psychophysical approach. In: Proceedings of the ACM International Conference on Multimedia, Scottsdale, 2011. 13–22
 - 63 Shi S, Jeon W J, Nahrstedt K, et al. Real-time remote rendering of 3D video for mobile devices. In: Proceedings of the ACM International Conference on Multimedia, Beijing, 2009. 391–400

Survey on augmented virtual environment and augmented reality

ZHOU Zhong^{1*}, ZHOU Yi¹ & XIAO JiangJian²

¹ State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China;

² Ningbo Industrial Technology Research Institute, Chinese Academy of Sciences, Ningbo 315201, China

*E-mail: zz@vrlab.buaa.edu.cn

Abstract With the rapid development of computer 3D processing capacity and the emergence of low-cost sensors, virtual reality (VR) has advanced quickly in recent years, especially in combination with real-world technologies. Such combination, sometimes also noted as mixed reality (MR), mainly includes two directions, augmented virtual environment (AVE) and augmented reality (AR). This paper summarizes AVE and AR trends, illustrates their technical characteristics, and introduces related hardware development. Then, the state-of-the-art technology development for both AR and AVE are surveyed. The technologies and Applications related to AR/AVE in mobile Internet are also introduced. In addition, the paper introduces the current status of relevant international standard drafting from one of the authors' participation in ISO/IEC activities. Finally, the paper concludes by presenting some future problems.

Keywords augmented virtual environment, augmented reality, augmented virtuality, mixed reality



of China Computer Federation and Institute of Electrical and Electronics Engineers.

ZHOU Zhong received his B.S. degree from Nanjing University and Ph.D. degree from Beihang University in 1999 and 2005, respectively. He is an associate professor at State Key Lab of Virtual Reality Technology and Systems, Beihang University, Beijing, China. His main research interests include natural phenomena simulation, augmented virtual reality, and Internet-based virtual reality technologies. He is a member



ZHOU Yi was born in 1988. He received the B.S. degree in Computer Science and Technology from Beihang University, in 2010. Currently, he is a Ph.D. candidate at State Key Lab of Virtual Reality Technology and Systems. His research interests include augmented virtual reality and time-lapsed video understanding. He is a student member of ACM.



and Application Journal.

XIAO JiangJian was born in 1972. He received the Ph.D. degree in Computer Science from University of Central Florida, Orlando, USA in 2004. Currently, he is a professor at Ningbo Industrial Technology Research Institute, CAS. His research interests include Computer Vision, Computer Graphics, and Visualization, etc. He is a member of ACM, a member of IEEE. He also is an associate editor of Machine Vision