

支持大规模视频融合的混合现实技术

Massive Video Integrated Mixed Reality Technology

周忠/ZHOU Zhong
孟明/MENG Ming
周颐/ZHOU Yi

(北京航空航天大学, 北京 100191)
(Beihang University, Beijing 100191, China)

虚拟现实(VR)是一个新兴的科学技术领域,该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,极大地扩展了人类认识、模拟和适应世界的能力。

虚实融合(MR)技术将虚拟环境与真实环境进行匹配合成,降低了三维建模的工作量,并借助真实场景及实物提高用户的体验感和可信度。随着当前视频图像的普及,MR技术的探讨与研究更是受到关注。

视频融合技术利用已有的视频图像,将它们融合到三维虚拟环境中,可以实现具有统一性的、深度的视频集成。该技术最早可追溯到1996年Paul Debevec^[1]提出的一种视点相关的纹理混合方法,即实现了不在相机视点的真实感漫游效果,但仍在很多方面存在难点。

1 MR的技术特点

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VR增强技术^[2],

收稿日期: 2017-09-23

网络出版日期: 2017-11-06

基金项目: 国家高技术研究发展(“863”)计划(2015AA016403); 国家自然科学基金(61572061、61472020)

中图分类号: TN929.5 文献标志码: A 文章编号: 1009-6868 (2017) 06-0006-004

摘要: 视频融合是视频监控控制相关应用领域关注的虚拟现实(VR)热点问题之一。根据实现虚实融合的维度不同,将相关研究方法分为4类:视频标签地图、视频图像拼接、视频叠加到三维场景、视频融合到三维场景。介绍了VR视频融合技术方面的相关工作,认为虚实融合技术正在快速发展中,虚实信息的可视关联对于人类认知和人工智能都已表现出显著的提升作用,未来将可能作为一种基础的地理信息资源来提供,有着重要的应用价值。

关键词: VR; 混合现实; 虚实融合; 视频融合; 视频监控控制

Abstract: Video fusion is a hot topic in some virtual reality (VR) application areas such as video surveillance. According to dimension number, related works in virtual-reality integration are surveyed in this paper in four categories: map with video tag, video image stitching, video overlapping on 3D scenes, and the video fusion in 3D scenes. Then our work in VR video fusion is presented in this paper, especially the new idea of video model and mixed reality rendering. The virtual reality technology is developing rapidly, and the visual correlation of the virtual and real environments has shown significant improvement on both human cognition and artificial intelligence. In the future, the video model may be provided as a fundamental geographic layer resource, which has significant values for various applications.

Keywords: VR; mixed reality; VR integration; video fusion; video surveillance

混合现实则是包含这两类技术及其融合。

增强现实技术与增强虚拟环境技术,可分别形象地描述为“实中有虚”和“虚中有实”。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。增强虚拟环境技术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。两项技术的MR方式如图1和图2所示,图中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。

随着VR技术的发展,其与现实世界正趋向于深度融合,一些技术开始兼具“虚中有实”和“实中有虚”这两种模式。20世纪70年代电影《星球大战》中展示的全息甲板是科幻中的终极理想状态,但U.C.Berkeley提出的Tele-immersion远程沉浸系统、微软的Holoportation、Magic Leap所设计的光场头盔显示原型等已经具备了这种深度MR特点。

2 视频融合技术分类

早期的视频融合技术只能做到将图片向地面或简单立面映射,现在的新技术则已经能够快速将视频实时地映射到复杂的三维模型上。根

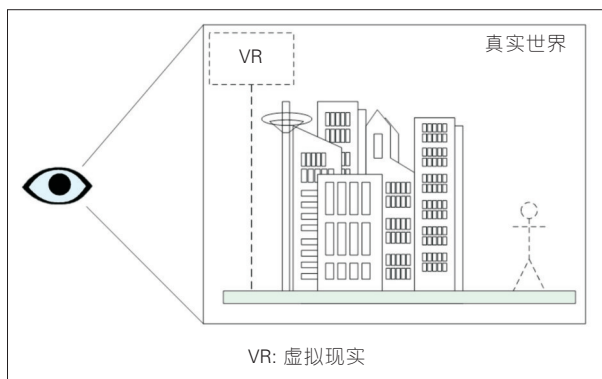


图1
增强现实技术“实中有虚”

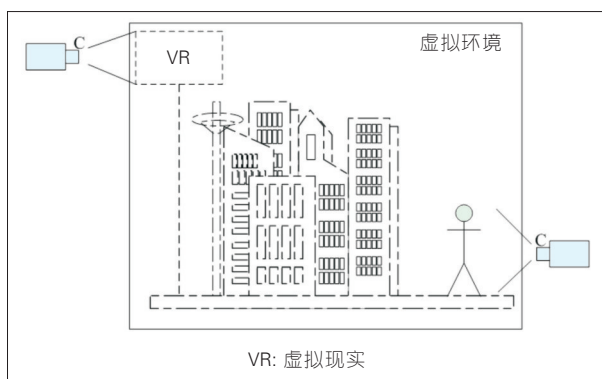


图2
增强虚拟环境技术“虚中有实”

据实现MR的维度不同,可将相关方法分为4类:视频标签地图、视频图像拼接、视频叠加到三维场景,视频融合到三维场景。需要说明的是:这些技术针对的是易于获取的普通摄像头视频,FreeD等基于多视图几何的三维重建方法不在讨论中。

2.1 视频标签地图

基于视频监测控制对于多地点视频有效组织的需求,采用视频标签与地图的索引集成,实现在地图上放置和观看视频。美国FX Palo Alto实验室、美国三菱电机研究所、韩国电子通信研究院和法国原子能署CEA LIST等机构在该方面开展了工作。

美国FX Palo Alto实验室提出了用于室内的多相机实时监测控制系统,动态物体跟踪系统(DOTS)^[3],该系统通过对平面设计图的特征分割,获得系统中视频与位置信息的关联,直接向用户提供一种视频分析结果的展示手段。美国三菱电机研究所MERL的Ivanov等也实现了用于室内

居住环境可视化的类似系统,通过加入运动传感数据等信息,为建筑系统的设计人员和管理人员提供整栋楼的上下文信息^[4]。视频标签索引的融合方法除了用于监测控制系统,还广泛应用于地理信息系统(GIS),它们通过建立提取的视频内容与数据库中GIS数据的对应关系,进行视频与GIS的融合。韩国电子通信研究院的Tae-Hyun Hwang等基于这个思想,将视频和虚拟场景中的meta元素提出,建立了基于meta元素的视频与GIS系统的关联^[5],通过简单的点击地图查询即可在手机上直接访问视频。此外,法国原子能署CEA LIST的Gay-Bellile等通过增强现实AR和相机追踪的方式建立了实时视频与2D GIS的关联^[6]。

2.2 视频图像拼接

视频图像拼接是指将空间上可配准、相互之间具有足够重叠区域的图像序列经过特征对齐、空间变换、重采样和拼接合成之后形成宽视角

甚至全景图像的方法。经过数十年的发展,图像拼接算法目前已经比较成熟,全景相机出现了低成本、消费级的特点,利用鱼眼相机来降低对相机数量的要求,达到小型化成为新的“爆点”。图像拼接主要针对的是窄基线相机图像序列,宽基线相机不具备统一的单应性,特别是遮挡大大影响重叠区域的匹配,其图像拼接是目前研究的难点。

2.3 视频叠加到三维场景

视频叠加到三维场景的方法以2D和3D特征注册为基础进行虚实融合,允许用户在相机视点的转移路径上观看叠加的结果,其效果优于视频标签地图方法。

塞尔维亚利兹大学CG&GIS实验室中Milosavljević等提出基于GIS增强的视频监测控制系统,将视频窗口叠加到3D模型视图窗口的上方显示,在3D GIS环境中建立视频与空间信息的位置关联^[7-8]。美国微软公司的Snively等于2006年提出了照片旅游系统^[9],该系统利用对互联网上照片集的匹配,重构一个稀疏的三维点云场景,通过图像变换的渲染方法进行照片浏览。荷兰代尔夫特理工大学的Haan等人受到Snively等共平面视点转移方法的启发,于2009年面向监测控制系统提出了第一人称式的场景导航方法^[10],通过交互地在场景中放置画布的方法完成视频的注册,进而通过动态视频嵌入实现导航,缓解了视点移动时视频间重叠区域的视觉差异现象。

2.4 视频融合到三维场景

视频与三维场景的融合方法,实质是将相机捕捉的视频图像,以纹理的方式实时注册到虚拟环境中,达到增强虚拟环境的效果,能够允许用户从非相机虚拟视点观察融合结果。这种方法比前几种融合方法在视点可选范围上进一步扩大,实际上是从另一个角度解决了宽基线相机以及

无重叠视域相机的图像拼接问题。但这类技术仍会存在一些难以克服的问题。

在 IEEE VR 03 上,南加州大学的 Ulrich Neumann 等人^[11]系统阐述了增强虚拟环境的概念,实现了随着图像数据变化的动态三维模型效果,解决了非相机视点下贴图扭曲现象^[12]。在 ACM MM 10 上,麻省理工学院的 DeCamp 等人^[13]设计了一套用于智能家居的沉浸式系统 HouseFly,通过鱼眼相机的三维融合,让用户可以漫游于掀顶式楼宇。在 ISMAR 09 上,佐治亚理工学院 Kihwan Kim 等^[14]提出基于动态信息增强 Google Earth 等航拍地球地图的方法,提出对视频进行分类处理和增强显示的方法。2012 年国立台湾大学的 Chen 等人^[15]建立了 GIS 辅助的可视化框架,融入了多分辨率监测控制策略,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。

3 MR 技术进展

近几年,我们在这方面开展了一系列工作,主要特色是将图片建模技术用到 MR 中,以得到准确的虚实对齐效果。其中图片建模技术是利用二维图片恢复场景三维结构的数学过程和计算技术,这一技术能够很容易地达到虚实融合过程中对三维模型精度的高要求,克服了视频投影本身带来的二三维深度不匹配问题。

3.1 基于图片建模的视频模型

该方法的核心部分是一种快速建模视频背景的交互式方法,使用体元和场景树来描述图像中各点之间的建模关系,首先针对单幅图像,在图像几何分析的预处理基础上,进行图像与场景模型的三维注册,然后提出了一种体元的定义,支持交互式的方式进行基本几何结构的恢复,实现单幅图像场景的视频模型生成。场景树结构示意如图 3 所示。

监测控制场景中存在大量相机

视频区域重叠度很小的情况,现有基于多视图的建模方法不能适用。进一步针对低重叠度图像序列,使用点线联合的匹配方法进行新图像与现有视频模型的注册,用户可以进一步进行新图像场景的结构建模,最终链式地匹配和注册更多的图像场景结构,如图 4 所示。

在此基础上,我们定义了一种基于单幅照片建模生成的视频模型,它描述了该照片对应的三维几何结构,

可以供二次开发使用。

3.2 基于视频模型的 MR 方法

我们提出了基于视频模型的 MR 方法,针对每个视频创建对应的视频模型^[16],然后通过纹理投影^[17]和阴影投影方法将视频与其模型进行融合。方法整体流程如图 5 所示,分为两个阶段:(1)预处理阶段。提取视频的背景帧进行交互式建模,得到简单的视频模型,然后与三维场景模型

图 3 场景树结构示意图

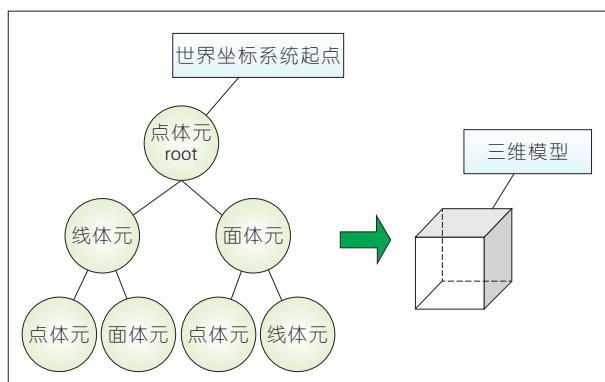


图 4 某火车站 6 幅相机图片的链式建模结果

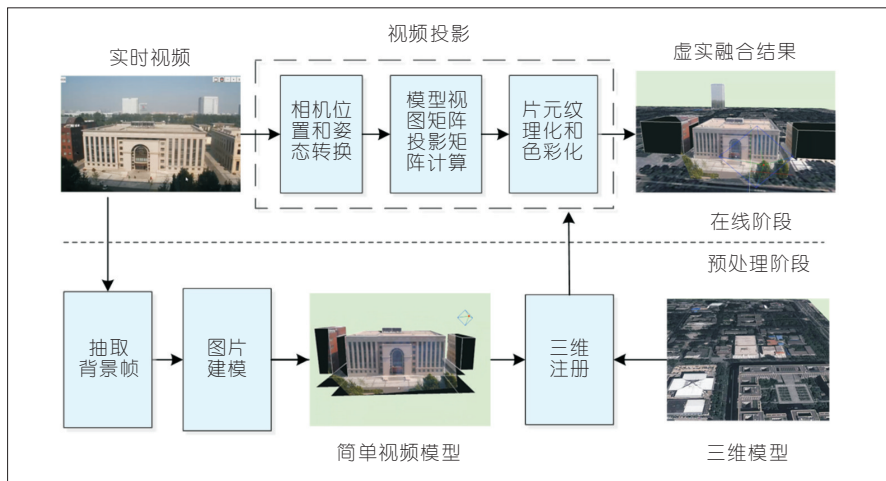
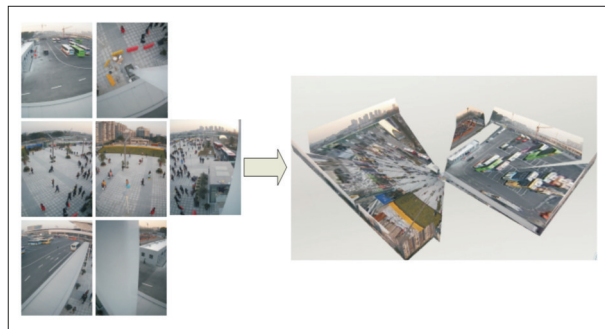


图 5 基于视频模型的 MR 流程

进行注册;(2)在线阶段。该阶段与直接投影的融合方法类似,但不再执行遮挡测试。

3.3 MR 场景中的自动路径规划方法

由于MR中视频中的动态物体并未三维化,这类技术不可避免存在偏离原视点会出现画面畸变的现象。研究视点变化与画面畸变之间的关系,我们给出了一种基于视频投影中的画面畸变的视点质量评价方法,进一步提出了一种MR场景中的自动路径规划方法,来尽量减少畸变现象。

3.4 支持大规模视频融合的视频监测控制

以上技术被应用于支持大规模视频融合的视频监测控制,如图6所

示,各建筑模型是手工建模得到的精细模型,视锥区域是视频模型。通过空间划分和瓦片关联,可以很容易地扩展到大规模的视频模型场景。

4 展望

随着视频监控控制在公共安全、交通等领域的广泛应用,多相机监测控制网络中相机数量日益增多,MR技术将现实世界中大量的相机视频进行整合,提供与真实世界具有几何结构一致性的统一视图,解决的是人类“认知”的问题。另一方面,实际上不同位置的相机在内容上也很难关联分析,对于计算机的智能分析能力也带来了巨大的挑战,这同样需要MR的信息支持。

在MR技术的基础上,我们正在

研究多相机拓扑中的几何-语义联合理解与关联问题,研究并实现了一种基于几何-语义结构分析的多相机场景拓扑连通图的构建方法。以多相机拍摄的监测控制视频为输入,如图7所示,解析相机图像的语义信息与基本几何结构,通过分析目标在不同相机之间的转移状态,计算各相机区域之间的连通概率,从而建立多相机间的拓扑连通图。

图8中展示了使用我们的算法在Duke^[18]数据集上恢复出的相机拓扑连通图,各视频图像和地图上对应的语义区域进行对齐,和原始场景中的拓扑连通关系相符。这种细粒度的相机视频融合方式可以很容易地作为一个地图图层推广到大规模GIS系统中,可以从根本上解决现有的海量视频碎片化问题。

MR技术正在快速发展中,这种虚实信息的可视关联对于人类认知和人工智能都已表现出显著的提升作用,未来作为一种基础的地理信息资源来提供,有着重要的发展意义。

参考文献

- [1] DEBEVEC P E, TAYLOR C J, MALIK J. Modeling and Rendering Architecture from Photographs: A Hybrid Geometry-and Image-Based Approach[C]//Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques. USA: ACM, 1996:11-20. DOI: 10.1145/237170.237191
- [2] AZUMA R, BAILLOT Y, BEHRINGER R, et al. Recent Advances in Augmented Reality[J]. Computer Graphics and Applications, 2001, (21): 34-47
- [3] GIRGENSOHN F, SHIPMAN T, TURNER, et al. Wilcox, Effects of Presenting Geographic Context on Tracking Activity Between Cameras[C]//Conference on Human Factors in Computing Systems. USA: California, 2007: 1167-1176
- [4] IVANOV Y, WREN C, SOROKIN A, et al. Visualizing the History of Living Spaces[J]. IEEE Transactions on Visualization and Computer Graphics, 2007, 13(6):1153-1160. DOI: 10.1109/TVCG.2007.70621
- [5] HWANG T H, CHOI K H, JOOL I H, et al. MPEG-7 Metadata for Video-Based GIS Applications[C]//Proceedings of IEEE International Geoscience and Remote Sensing Symposium. USA: IEEE, 2003, (6): 3641-3643. DOI: 10.1109/IGARSS.2003.1294880
- [6] GAY-BELLILE V, LOTHE P, BOURGEOIS S,

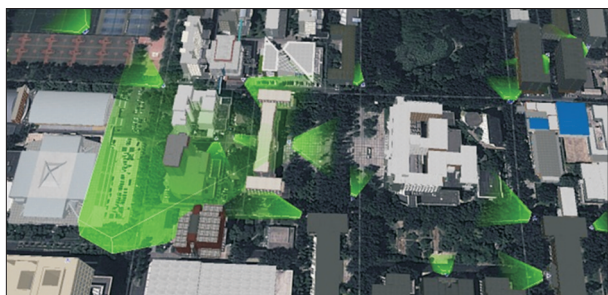


图6
多路相机视频的融合概览

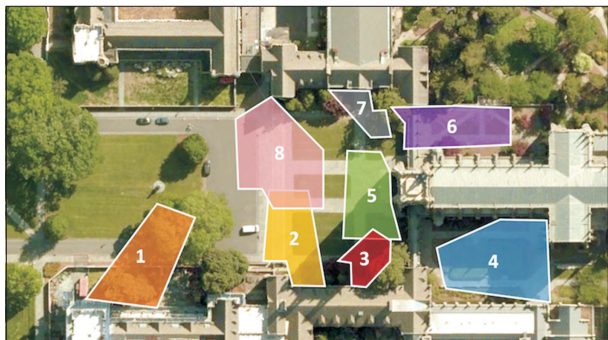


图7
Duke数据集的相机关系

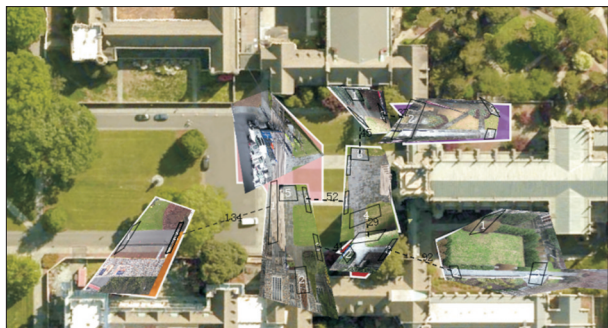


图8
相机图片覆盖区域及拓扑连通

下转第36页