软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2016,27(Suppl.(2)):197−206

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

基于多视频的虚实融合系统*

潘成伟[1,2], 张建国[1,2], 王少荣[2], 汪国平[1,2]

[1](北京大学 信息科学技术学院,北京 100871)

[2](北京市虚拟仿真与可视化工程研究中心,北京 100871)

通讯作者: 潘成伟, E-mail: pancw@pku.edu.cn; 汪国平, E-mail: wgp@pku.edu.cn

摘 要: 提出了一种基于多视频的虚实融合可视化系统的构建方法,旨在将真实世界中的图像和视频融合到虚拟场景中,用视频图像中的纹理和动态信息去丰富虚拟场景,提高虚拟环境的真实性,得到一种增强的虚拟环境.利用无人机采集图像来重建虚拟场景,并借助图像特征点的匹配来实现视频图像的注册.然后利用投影纹理映射技术,将图像投影到虚拟场景中.视频中的动态物体由于在虚拟环境中缺失对应的三维模型,直接投影,当视点发生变化时会产生畸变.首先检测和追踪这些物体,然后尝试使用多种显示方式来解决畸变问题.此外,系统还考虑有重叠区域的多视频之间的融合.实验结果表明,所构造的虚实融合环境是十分有益的.

关键词: 三维重建;增强虚拟;投影纹理映射;背景建模;运动追踪

中文引用格式: 潘成伟,张建国,王少荣,汪国平.基于多视频的虚实融合系统.软件学报,2016,27(Suppl.(2)):197−206. http://www.jos.org.cn/1000-9825/16034.htm

英文引用格式: Pan CW, Zhang JG, Wang SR, Wang GP. Virtual-Real fusion system integrated with multiple videos. Ruan Jian Xue Bao/Journal of Software, 2016,27(Suppl.(2)):197−206 (in Chinese). http://www.jos.org.cn/1000-9825/16034.htm

Virtual-Real Fusion System Integrated with Multiple Videos

PAN Cheng-Wei[1,2], ZHANG Jian-Guo[1,2], WANG Shao-Rong[2], WANG Guo-Ping[1,2]

[1](School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China)

[2](Beijing Engineering Research Center for Virtual Simulation and Visualization, Beijing 100871, China)

Abstract: This paper proposes a method for constructing a virtual-real fusion system integrated with multiple videos aiming to create an augmented virtual environment, where images and videos captured from real world are fused to virtual scene. With the help of textures from images and motion from videos, the virtual environment is more realistic. Unmanned Aerial Vehicles are used to take photos and reconstruct the 3D virtual scene. By matching features, video frames can be registered to the virtual environment. Then images are projected to virtual scene with the method of projective texture mapping. Due to lack of the corresponding 3D models in the virtual environment, distortions will occur when images are directly projected and the viewpoint changes. This paper first detects and tracks those moving objects, then it gives multiple ways of displaying moving objects to solve the distortion problem. Fusion of multiple videos with overlapping areas in the virtual environment is also considered in this system. The experimental results show that the virtual-real fusion environment that is build based in this paper has lots of benefits and advantages.

Key words: 3D reconstruction; augmented virtual reality; projective texture mapping; background modeling; motion tracking

虚拟现实技术将现实环境的要素和属性进行抽象并加以表现,通过逼真的绘制方法进行显示,例如目前应

* 基金项目: 国家自然科学基金(61232014, 61421062, 61472010, 61121002); 国家科技支撑计划(2015BAK01B06); 国家重点基础研究发展计划(973)(2015CB3518806)

Foundation item: National Natural Science Foundation of China (61232014, 61421062, 61472010, 61121002); National Key Technology Research and Development Program of the Ministry of Science and Technology of China (2015BAK01B06); National Basic Research Program of China (973) (2015CB3518806)

收稿时间: 2016-05-10; 采用时间: 2016-09-07

用十分广泛的虚拟地球软件(Google earth 和 Microsoft virtual earth 等),可以让用户查看世界各地的卫星图像、地形、三维建筑物等信息,这些可视化元素构建的虚拟环境并不能很好地反映真实世界,主要是因为虚拟环境是静态的.这种静态性主要体现在:(1) 三维模型的纹理是事先采集的,不能反映真实环境的变化情况.例如时间、季节的变化.(2) 三维模型都是静态的,不能反映真实环境的动态情况.例如行人、车辆的运动.

为了增强虚拟环境的真实性,我们利用相机捕获真实对象的图像、视频或三维模型,并将它们实时注册到虚拟环境中.图像中的纹理信息能够反映出真实世界中三维对象的样貌,视频中的动态信息可以如实地再现出真实环境中发生的动态事件,这样产生的增强虚拟环境(augmented virtual environment),我们称其为虚实融合,即“虚中有实”.虚实融合需要考虑如下几个问题:

(1) 图像与虚拟场景的配准问题,即求解相机的姿态信息,这样才能确定图像中的像素与三维虚拟环境中三维点之间的映射关系,是融合必须解决的基本问题.

(2) 视频是对虚拟环境的动态补充,因此视频中可能存在一些虚拟场景中不存在的三维物体,对于这些动态物体的识别和呈现是需要解决的另一个关键问题.

(3) 对于多个视频,视频间可能存在重叠,需要考虑在重叠区域内的融合,使得视频之间的缝隙过渡平缓.

本文第 1 节给出所参考的相关工作.第 2 节对系统的总体结构进行阐述.第 3 节阐述模型重建与图像注册问题.第 4 节描述视频与虚拟环境的融合过程.第 5 节给出实际系统效果图.最后总结全文,并展望未来的工作.

1 相关工作

基于图像的建模和纹理映射,主要是将相机拍摄的图像映射到三维模型上,产生复杂三维场景逼真的几何和外观效果.Debevec[1]利用多张照片来重建和绘制建筑场景,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的效果.将视频图像在三维场景中进行映射生成动态纹理的想法,最早由 Video Flashlights 系统[2]实现,并将多个视频注册融合到同一个三维环境中,使得用户能够以一个全局的视角观察三维模型和视频,增强了视频的空间表现力.Neumann 等人构建了一个类似的系统并提出了增强虚拟环境[3](augmented virtual environment,简称 AVE)的概念.Sebe[4]等人在 AVE 的基础上,利用背景差分的方法去检测动态物体并使用带纹理的移动的矩形框来显示动态目标.Wang[5]等人提出了 Contextualized Videos 的概念并强调了几种视频呈现的方式,旨在提供一个全局可靠的监控环境.Kim[6]等人提出了使用视频中的动态信息增强 Google Earth 等航拍地球地图的方法,对视频进行分类处理和增强现实.增强虚拟环境在室内监控系统[7,8]、网络摄像头[9]、城市意识[10]等领域也有广泛的应用.

在有关投影机位置矫正的研究中,文献[2]利用边缘检测的方法对摄像头的位姿进行校准,但这种方法不是实时的,且需要一定的人工交互.在 AVE 系统[3]中,相机的初始位姿由 GPS 等传感器获得,然后根据特征的自动校准进行优化.Abrams[9]等人提出了一种基于多边形匹配的方法,需要用户在二维图像和三维环境中,标记出多组相互匹配的多边形区域.

在多重纹理融合方面,Haan[11]在一些特定的环境下进行了尝试,例如球场,机场航站楼等.他融合了很多前人的工作,首先需要对摄像头的位置作矫正,其次对视频内容进行匹配,选择融合策略(融合或者替换).2012 年,加拿大 York University 提出了一个 3DTown 的概念[10],其本质是将 Haan 的想法扩展至城市级别,重点在于如何实施展示大量视频信息的问题.

2 系统总体结构图

虚实融合系统最主要的两个元素是“虚”与“实”.“虚”指的是三维虚拟环境,是对真实环境的一种描述,因此虚拟环境的构建需要根据从真实世界采集的数据进行建模.“实”指的是摄像机采集的图像和视频数据.基于视频图像的虚实融合,本质上是对三维虚拟环境的纹理重建,更确切地说是动态纹理的重建.纹理重建的前提是需要计算三维空间点与图像中像素的映射关系,即视频与虚拟场景的配准.视频成功注册到虚拟环境中,就可以利

用投影纹理的方式进行纹理映射,将视频叠加到虚拟场景中.多个视频融合的时候,融合的边界需要单独处理.对于视频中的动态物体,由于在虚拟环境中缺少对应的三维模型,直接投影的话,可能存在物体变形的问题,因此需要单独处理.综上所述,虚实融合系统主要包含以下模块:(1) 数据采集与模型重建.(2) 图像、视频在虚拟环境下的注册.(3) 图像与虚拟环境的融合.(4) 多视频在虚拟场景中的融合.(5) 视频中动态场景在虚拟环境中的呈现.系统的总体结构框架如图 1 所示.

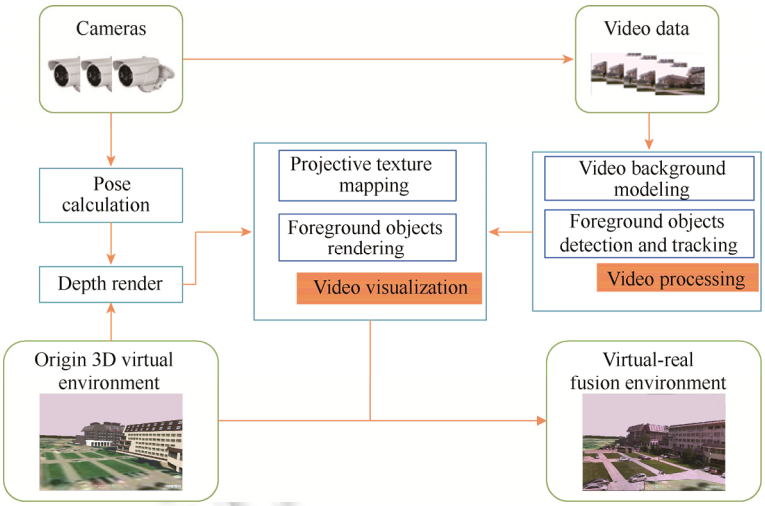


Fig.1 Overview of virtual-real fusion system

图1 系统总体结构图

3 模型重建与图像注册

虚拟环境的构建是虚拟融合系统的关键问题,虚拟环境主要是由许多三维模型组成的,例如地面、建筑等.一个准确的三维模型场景是虚实融合的前提,准确的三维模型有利于图像的成功注册以及图像像素点与三维空间点的正确映射关系的建立.

3.1 模型重建

随着无人机的兴起和广泛应用,我们可以借助无人机采集真实三维空间下不同视角的高清晰度的照片,然后利用这些照片进行三维建模.系统中使用大疆 Phantom 2 型号的无人机进行数据的采集.无人机的重量为 1.2kg,最大飞行时长为 25 分钟(由实际载荷决定),机身上挂载了 GPS、IMU 等传感器,用来控制无人机的飞行轨迹,一个 1 400 万像素 1/2.3 英寸的摄像头用来采集图像.通过控制无人机的飞行高度和相机的朝向,我们能够得到许多不同视点下的三维场景的照片.实际采集照片时,由于遮挡的存在,我们可以使用数码相机拍摄一些额外视角的图像作为补充,用于最终的三维重建.

数据采集完之后,我们使用多视图视觉方法进行三维重建.首先,对每张图像提取 SIFT[12]特征点及其描述符,进而进行特征点的匹配.然后利用 SFM(structure from motion)[13]方法恢复出所有相机的内、外参数以及稀疏点云,进而计算每张图像的深度图,并利用深度图融合[14]的方法获得稠密的三维点云.得到稠密三维点云以后,利用泊松重建[15]的方法得到三维模型的网格.最后,对网格进行纹理重建[16],得到最终的三维模型,如图 2 所示.此外,还可以利用一些建模软件(例如 MeshLab,SketchUp)来完善重建的三维模型.

Fig.2 Some examples of reconstruction

图2 一些三维重建的实例

3.2 图像注册

图像与虚拟场景的配准,本质上是相机标定的问题,需要知道相机的内参和外参.给定图像上的点(x_i,y_i)以及对应的三维点(X_i,Y_i,Z_i),存在一个 3×4 的矩阵 $\boldsymbol{M}$,有如下关系:

$$w\begin{pmatrix}x_i\\y_i\\z_i\end{pmatrix}=\boldsymbol{M}\begin{pmatrix}X_i\\Y_i\\Z_i\\1\end{pmatrix},\ \boldsymbol{M}=\boldsymbol{K}[\boldsymbol{R}\ \boldsymbol{T}],\ \boldsymbol{K}=\begin{bmatrix}f & s & x_0\\0 & f & y_0\\0 & 0 & 1\end{bmatrix} \tag{1}$$

其中,w 是一个放缩因子,$\boldsymbol{M}$ 能够被分解成一个 3×3 的相机内参矩阵 $\boldsymbol{K}$,一个 3×3 的旋转矩阵 $\boldsymbol{R}$ 以及平移向量 $\boldsymbol{T}$.f 是相机的焦距,s 是相机的错切,通常为 0,x_0 和 y_0 表示图片的主点,一般为图片的中心点.理论上 6 组对应点就可以计算出矩阵 $\boldsymbol{M}$,但由于匹配时存在误差,因此需要寻找更多的对应点来进行相机的标定.图像在虚拟场景中的配准,可以看作是在图像与虚拟场景中寻找匹配的点[17]、线[2]、面[9]等.线和面的匹配可以转换成相应点的匹配,因此系统中通过寻找多组匹配点进行相机的标定.

在第 3.1 节的三维重建过程中会产生特征点云,每个三维点对应多个图像特征点,因此图像特征点与三维空间点的匹配可以转换成二维图像特征描述符的匹配.考虑到大场景下,三维特征点云的数量巨大,如果直接将图像特征点与特征点云进行匹配,搜索空间大,导致十分耗时且内存开销巨大.我们采用类似文献[18]的方法来均匀采样虚拟视点绘制得到三维场景的代表性视图用于特征的匹配.利用图像检索的方法[19]检索出与待注册图像最相似的合成图像,然后进行图像间的特征点匹配.当获得多组待注册图像特征点与空间三维点的匹配后,利用文献[20]中的方法求解相机的内、外参数.图像注册的方法见算法 1.

算法 1. 图像注册(如图 3 所示).

输入: 特征点云 $\boldsymbol{Q}$ 和图像 $\boldsymbol{I}$;

输出: 图像 $\boldsymbol{I}$ 的 $\boldsymbol{K},\boldsymbol{R},\boldsymbol{T}$.

1.均匀采样虚拟视点位置,绘制 $\boldsymbol{Q}$ 得到合成图像 $V=\{V_i\}_{i=1,\cdots,N}$

2.检索出与图像 $\boldsymbol{I}$ 相似的合成图像 $Q=\{Q_j\}_{j=1,\cdots,k}$

3.SET 匹配点集合 $\boldsymbol{S}=\varnothing$

4.FOR each Q_j do

5.找到图像 $\boldsymbol{I}$ 和 Q_j 之间的匹配 S_j

6.$\boldsymbol{S}=\boldsymbol{S}\cup S_j$

7.End FOR

8.利用 $\boldsymbol{S}$,建立二维像素点和三维点的匹配对,SolvePnP 求解 $\boldsymbol{K},\boldsymbol{R},\boldsymbol{T}$

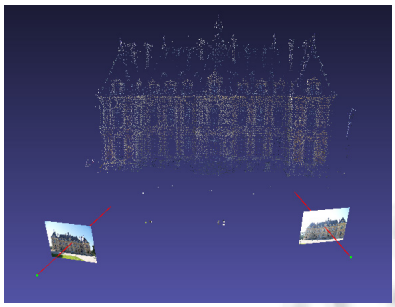


Fig.3 Schematic diagram of image registration. Green points represent positions while red points represent orientations

图3 图像注册的示意图.绿色点表示相机位置,红色直线表示相机的朝向

4 视频与虚拟环境的融合

随着图像成功注册到虚拟场景中,给定空间三维点,便可以计算其对应的像素,利用投影纹理的方法重新生成三维模型表面的纹理,达到“实”增强“虚”的效果.在处理多视频的情况下,空间三维点在多副图像上均产生了纹理,这时就产生了多纹理融合的问题.对于静态对象可以直接以投影纹理的方式融合到虚拟场景,但视频中的动态对象呢?由于三维场景中缺失动态对象对应的三维模型,直接投影,会产生物体变形的问题,因此视频融合时,需要考虑视频中动态场景的特殊处理.

4.1 投影纹理映射

由于标定出了相机的内外参数,可以得到三维场景中三维点与图片中像素的对应关系,这种映射关系本质上由一个 4×4 的矩阵 $\boldsymbol{M}$ 决定.系统中投影纹理映射是基于 OpenGL 实现的,$\boldsymbol{M}$ 可以分解为 4×4 的视图矩阵 $\boldsymbol{V}$ 和 4×4 的投影矩阵 $\boldsymbol{P}$.给定空间三维点,其纹理坐标计算如下:

$$w\begin{bmatrix}s\\t\\q\\t\end{bmatrix}=\boldsymbol{PV}\begin{bmatrix}X_i\\Y_i\\Z_i\\1\end{bmatrix} \tag{2}$$

其中,(s,t)表示纹理坐标,w 用来判断三维点在相机前(>0)还是后(<0),q 表示三维点的深度值,这些值的范围在(–1,1)之间,需要归一化到(0,1)之间.由于考虑深度值,需要将相机内参矩阵 $\boldsymbol{K}$ 变成 4×4 的矩阵,$\boldsymbol{P}$ 和 $\boldsymbol{V}$ 的计算方式如下:

$$\boldsymbol{P}=\overbrace{\begin{pmatrix}2/\mathrm{W}&0&0&-1\\0&2/H&0&-1\\0&0&-2/(F-N)&-(F+N)/(F-N)\\0&0&0&1\end{pmatrix}}^{\text{Normalization}}\overbrace{\begin{pmatrix}f&s&x_0&0\\0&-f&y_0&0\\0&0&-(F+N)&F\times N\\0&0&1&0\end{pmatrix}}^{\boldsymbol{K}\ 4\times4},\ \boldsymbol{V}=\begin{pmatrix}\boldsymbol{R}&\boldsymbol{T}\\\boldsymbol{0}&1\end{pmatrix} \tag{3}$$

其中,F 为相机到远裁剪平面的距离,N 为相机到近裁剪平面的距离,W,H 为图片的宽和高.

利用投影纹理的方法进行纹理替换,新旧纹理融合时,为使边界过渡比较平缓,主要利用差值和高斯融合的方法进行纹理的融合.直接进行纹理投影,会产生遮挡穿透的问题.图像中每个像素点对应空间中的一条射线,

纹理投影时,射线上的点均对应着相同的纹理.但实际中由于遮挡的存在,一条射线上只能有一个点能用该像素点进行着色.因此,实际绘制时,需借助点的深度信息[21]来判断是否着色,杜绝了投影纹理的穿透问题.如图 4 所示.

(a) 原模型 (b) 投影纹理映射

Fig.4 Result of projective texture mapping

图4 投影纹理映射的结果

4.2 多视频的融合

虚拟场景中可能会存在多个投影机,投影机之间可能存在投影区域的重叠,在重建重叠区域的纹理时,可能会有多个图像纹理参与.对于重叠区域内的三维点,我们需要考虑覆盖该三维点的每个投影机的贡献值,贡献值的大小主要取决于以下几个因素:虚拟视点位置、投影机位置到该三维点的距离、两直线的夹角、投影机图像的像素分辨率.距离越近,夹角越小,分辨率越高,投影机的贡献率越大.融合时,需要构建一个权值函数来体现上述关系.纹理贡献的权值为 $r = p/(\partial \times d)$.$p$ 为图像分辨率,∂为夹角,d 为距离.最后融合后得到的纹理值为

$$T = \left(\sum I_i \times r_i\right) \Big/ \sum r_i .$$

对于多个视频与虚拟场景的融合,主要是寻找一条合理的分割线,从不同视频中截取不同的部分,然后将这些部分融合起来,分割线周边的三维点的纹理则由不同视频的贡献值加权获得.目前系统使用两种融合策略:(1) 选择与虚拟视点最接近的视频作为主投影源,投影时先投影主视频源,再投影其他视频源.主投影源的确定主要根据投影机的位置、视角与虚拟视点的位置和方向的差异来确定,如果差异在阈值以内,则存在主投影源.(2) 如果不存在主投影源,换言之,多个视频的贡献率相仿,则先将这些视频投影变换到同一视点下,利用 GraphCut 方法[22]得到融合这些视频的分割线,然后将这些分割线反投影回原视频中,得到每个视频实际使用的区域.由于这些操作比较耗时,如果每次虚拟视点变化时都计算一次,那么很难达到实时.因此需要提前采样一组空间分布均匀的视点,计算出在这些视点下的每个视频对应的 Mask 图片(实际使用的区域),然后使用第 1 种策略选择相应的采样视点.如图 5 所示.

(a) 4 个视频 (b) 融合后的效果

Fig.5 Fusion of multiple videos

图5 多视频的融合

4.3 视频动态场景的融合

对于视频中的动态对象,在虚拟场景中可能没有对应的三维模型,实际投影时,这些动态目标物体只能投影到错误的模型上,例如将行人投影到地面上.如果虚拟视点与视频实际拍摄位姿差异比较大,那么这些动态物体的投影会产生较大的畸变问题,如图 6 所示.为了解决畸变问题,我们需要检测出这些动态对象的所在区域,然后以其他融合策略在虚拟场景中呈现这些动态物体.

(a) 虚拟视点在图像位置附近 (b) 虚拟视点与图像位置偏差很大

Fig.6 Distortion problem occurs when viewpoint changes larger

图6 视点变化较大时产生畸变问题

考虑到实现的难易程度以及时间效率的高低,我们利用背景差分的方法来检测视频中的动态物体,实时地构建视频的背景图片并逐帧更新,然后逐帧地与背景图片作差,利用阈值法得到二值图像,再对二值图像进行腐蚀(erode)或膨胀(dilate)操作,利用两遍扫描法(two-pass)寻找图像中的联通区域,每个联通区域认为是一个动态目标.系统中使用高斯混合背景建模法[23]动态更新视频的背景.关于动态物体的追踪,我们使用光流法[24]在相邻视频帧之间追踪,对于追踪丢失的物体,可以比较其颜色直方图来确定是否为同一个目标物体.

给定空间三维点,我们根据公式(2)能够计算其在图像中的像素位置,反之,给定图像的像素位置,如果知道该像素位置的深度值,那么就可以求得其在空间中的三维点位置,即公式(2)的逆过程.给定相机的姿态信息($\boldsymbol{K},\boldsymbol{R},\boldsymbol{T}$),我们可以构建虚拟视点的视图矩阵和投影矩阵,然后绘制三维场景得到该视点下的深度图,由于三维场景中缺少动态物体的三维模型,因此图像中动态物体所在区域的深度值是错误的,但动态物体边界处的深度值是准确的,因此可以利用这些位置的深度值,估算出动态物体在空间中的位置.

对于图像中的动态物体,我们提取其 AABB 包围盒(x_i,y_i,w_i,h_i).(x_i,y_i)表示包围盒左上角位置,(w_i,h_i)表示包围盒的宽和高.那么物体在空间的位置由像素$(x_i+w_i/2,y_i+h_i/2)$决定,如果在空间立一薄片表示物体,那么薄片的宽由(x_i,y_i+h)和(x_i+w_i,y_i+h_i)决定,薄片的宽与高的比为w_i/h_i.薄片站立的方向与薄片所在位置的法向一致.

(a) Billboard (b) 高亮 Marker (c) 三维模型

Fig.7 Strateqies of representing moving objects

图7 动态物体的呈现策略

关于动态物体与在虚拟场景中的呈现方法,系统提供了 3 种策略:(1) 以 Billboard(半透明矩形薄片)的方式绘制,其始终面向虚拟视点的方向,即正面始终朝向观察者.(2) 当视点高度比较高时,用高亮的 Marker(例如球)表示,当鼠标悬停在 Marker 上时,弹出窗口显示图像中物体的 AABB 包围盒区域.(3) 放置相应的三维模型来替换运动的物体,目前系统仅考虑行人和车辆.

5 结果及讨论

5.1 实验结果

基于上述设计,利用 C++/OpenGL 进行原型系统的开发,软件系统主机配置如下:Intel Core(TM) i5-3470 CPU,主频 3.2GH;8GB 内存;NVIDIA GeForce GTX 650 显卡.系统的部分截图如图 4、图 5、图 8、图 9 所示.

在做多视频虚实融合的过程中,需要对视频进行解码,多视频同时解码,需要占用大量的资源,这也是系统的瓶颈所在.为了提高效率,我们利用多线程的技术,每个视频流用一个线程单独处理,并为每路视频分配一个缓冲队列,缓冲队列的大小主要由内存的开销与绘制帧率之间的权衡来决定.实验中我们发现,随着摄像头数目的增加,绘制的帧率会有所降低,具体见表 1.总体而言,系统能够达到实时交互的帧率.

Table 1 Frames per second of our system

表1 系统绘制帧率

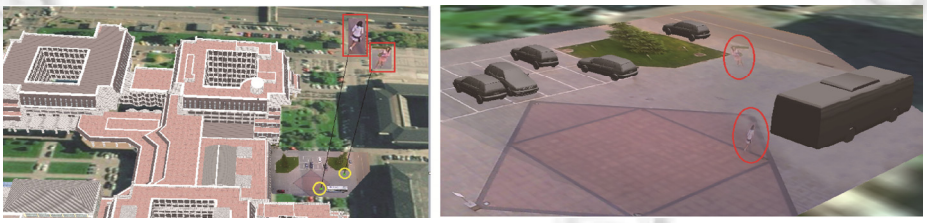
视频的数目	绘制帧率
0	60
1~5	25~35
6~10	10~22
11~15	1~9

(a) 静态三维场景 (b) 高空俯视的效果 (c) 两个视频融合的场景

圆圈内为投影区

Fig.8 Virtual-Real fusion sence of two videos

图8 两个视频的虚实融合场景



(a) 高亮 Marker 显示运动物体,圆圈内为高亮的 Marker,矩形框内为实际运动物体图像

(b) Billboard 显示运动物体,即用半透明的竖立的薄片来显示运动的物体,如图中圆圈所示

(c) 用放置三维模型的方式来表示动态目标 (d) 三维虚拟环境中增加的三维模型

Fig.9 Display results of the actual system

图9 实际系统的显示效果

与现有的 Video Flashlights[2],AVE[3,4],3D Town[10]等系统相比,本文系统主要有以下几个优势:(1) 现有系统的 3D 虚拟环境多采用 3D Laser 的方式构建三维模型,成本高,而本文使用无人机采集图像进行三维重建,成本低而且可面向大场景的重建.(2) 本文的 3D 虚拟环境是利用多视图几何方法重建出来的,因此图像的注册可以利用特征点的匹配来求解,求解方法简单、有效,而其他系统不具备这样的条件.(3) 本文采用多种方式显示动态目标物体,可以在二维和三维之间自由切换,让用户能够快速捕捉并分辨不同的物体,而其他系统显示方式单一,不够灵活.(4) 多数系统考虑单幅图像或单个视频与虚拟场景的融合,即使考虑多视频,视频间无重叠或重叠区融合效果一般.而本文考虑了具有一定重叠区域的多视频之间的融合,并且取得了不错的融合效果.

5.2 局限性讨论

目前系统能够支持多图像、多视频与虚拟场景的融合,对于视频与虚拟场景的融合能够达到实时交互的帧率,一个很重要的应用前景是监控摄像头视频流与虚拟场景的融合,实时反映某区域的监控情况.但目前,系统仅支持视点固定不变的视频流,对于这种视频流,相机的位姿只需要求解一次就可以了.但对于视点发生变化的视频流,例如无人机采集的视频流、PTZ(pan/title/zoom)云台摄像机、手持设备采集的视频流等,相机的位姿信息是一直变化的,目前系统只能将这些视频流拆解成图像,逐个融合到系统中,难以达到实时交互.如果能实时求解出每一帧图像的位姿并进行融合,系统将会具有更加广泛的应用.

6 结束语

本文提出了一种基于多视频的虚实融合可视化系统的构建方法,提供一种增强的虚拟环境,用来在虚拟环境中呈现视频的真实性和动态性.系统借助多视频与虚拟场景的融合,使得用户能够在三维模型上下文的基础上浏览视频中的动态事件,并借助视频丰富的三维模型的纹理细节,增加虚拟场景的真实感,降低建模的复杂度.但目前系统仍有待优化且效果提升的空间比较大,下一步的工作如下:构建一套比较完备的三维模型库,根据视频中前景物体的图像,检索出相应的模型,在虚拟场景中不仅要恢复三维模型的纹理信息,也要恢复其运动的姿态信息.另外,还需要考虑光照一致性的问题,使得多视频以及虚拟场景的光照条件近似吻合,平缓过渡,获得更佳的融合效果.最后,尝试对无人机获取的实时信息进行配准,并实时获取纹理和动态信息融合到三维场景中.

References:

[1] Debevec PE, Taylor CJ, Malik J. Modeling and rendering architecture from photographs: A hybrid geometry-and image-based approach. In: Proc. of the 23rd Annual Conf. on Computer Graphics and Interactive Techniques. New York: ACM Press, 1996. 11–20. [doi: 10.1145/237170.237191]

[2] Sawhney HS, Arpa A, Kumar R, *et al*. Video flashlights: Real time rendering of multiple videos for immersive model visualization. In: Proc. of the ACM Int'l Conf. Aire-la-Ville: Eurographics Association Press, 2002. 157–168.

[3] Neumann U, You S, Hu J, *et al*. Augmented virtual environments (AVE): Dynamic fusion of imagery and 3D models. In: Proc. of the IEEE Virtual Reality. Los Alamitos: IEEE Computer Society Press, 2003. 61–67.

[4] Sebe IO, Hu J, You S, *et al*. 3D video surveillance with augmented virtual environments. In: Proc. of the 1st ACM SIGMM Int'l Workshop on Video Surveillance. New York: ACM Press, 2003. 107–112. [doi: 10.1145/982452.982466]

[5] Wang Y, Krum DM, Coelho EM, *et al*. Contextualized videos: Combining videos with environment models to support situational understanding. IEEE Trans. on Visualization and Computer Graphics, 2007,13(6):1568–1575. [doi: 10.1109/TVCG.2007.70544]

[6] Kim K, Oh S, Lee J, *et al*. Augmenting aerial earth maps with dynamic information from videos. Virtual Reality, 2011,15(2/3): 185–200. [doi: 10.1007/s10055-010-0186-2]

[7] Chen YY, Huang YH, Cheng YC, *et al*. Integration of multiple views for a 3-D indoor surveillance system. Information-An Int'l Interdisciplinary Journal, 2010,13(6):2039–205.

[8] De Camp P, Shaw G, Kubat R, *et al*. An immersive system for browsing and visualizing surveillance video. In: Proc. of the 18th ACM Int'l Conf. on Multimedia. New York: ACM Press, 2010. 371–380. [doi: 10.1145/1873951.1874002]

[9] Abrams AD, Pless RB. Webcams in context: Web interfaces to create live 3D environments. In: Proc. of the Int'l Conf. on Multimedia. New York: ACM Press, 2010. 331–340. [doi: 10.1145/1873951.1873997]

[10] Corral-Soto ER, Tal R, Wang L, *et al*. 3D Town: The automatic urban awareness project. In: Proc. of the 9th Conf. on Computer and Robot Vision (CRV). Los Alamitos: IEEE Computer Society Press, 2012. 433–440. [doi: 10.1109/CRV.2012.64]

[11] De Haan G, Scheuer J, de Vries R, *et al*. Egocentric navigation for video surveillance in 3D virtual environments. In: Proc. of the IEEE Symp. on 3D User Interfaces. Los Alamitos: IEEE Computer Society Press, 2009. 103–110. [doi: 10.1109/3DUI.2009.4811214]

[12] Lowe DG. Distinctive image features from scale-invariant keypoints. Int'l Journal of Computer Vision, 2004,60(2):91–110. [doi: 10.1023/B:VISI.0000029664.99615.94]

[13] Snavely N, Seitz SM, Szeliski R. Photo tourism: Exploring photo collections in 3D. ACM Trans. on Graphics (TOG), 2006,25(3):835–846. [doi: 10.1145/1179352.1141964]

[14] Bleyer M, Rhemann C, Rother C. PatchMatch stereo-stereo matching with slanted support windows. In: Proc. of the British Machine Vision Conf. Dundee: BMVA Press, 2011,11:14.1–14.11.

[15] Kazhdan M, Hoppe H. Screened poisson surface reconstruction. ACM Trans. on Graphics (TOG), 2013,32(3):29. [doi: 10.1145/2487228.2487237]

[16] Waechter M, Moehrle N, Goesele M. Let there be color! Large-Scale texturing of 3D reconstructions. In: Proc. of the European Conf. on Computer Vision. Switzerland: Springer International Publishing, 2014. 836–850. [doi: 10.1007/978-3-319-10602-1_54]

[17] Zhang Z. A flexible new technique for camera calibration. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2000,22(11):1330–1334. [doi: 10.1109/34.888718]

[18] Aubry M, Russell BC, Sivic J. Painting-to-3D model alignment via discriminative visual elements. ACM Trans. on Graphics (TOG), 2014,33(2):14. [doi: 10.1145/2591009]

[19] Irschara A, Zach C, Frahm JM, *et al*. From structure-from-motion point clouds to fast location recognition. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2009. 2599–2606. [doi: 10.1109/CVPR.2009.5206587]

[20] Sattler T, Sweeney C, Pollefeys M. On sampling focal length values to solve the absolute pose problem. In: Proc. of the European Conf. on Computer Vision. Switzerland: Springer International Publishing, 2014. 828–843. [doi: 10.1007/978-3-319-10593-2_54]

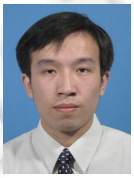
[21] Segal M, Korobkin C, Van Widenfelt R, *et al*. Fast shadows and lighting effects using texture mapping. ACM SIGGRAPH Computer Graphics, 1992,26(2):249–252. [doi: 10.1145/133994.134071]

[22] Kwatra V, Schödl A, Essa I, *et al*. Graphcut textures: Image and video synthesis using graph cuts. ACM Trans. on Graphics (ToG), 2003,22(3):277–286. [doi: 10.1145/1201775.882264]

[23] Zivkovic Z. Improved adaptive Gaussian mixture model for background subtraction. In: Proc. of the 17th Int'l Conf. on Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2004,2:28–31. [doi: 10.1109/ICPR.2004.1333992]

[24] Bouguet JY. Pyramidal implementation of the affine Lucas Kanade feature tracker description of the algorithm. Microprocessor Research Labs, Intel Corporation, 2001,(5):4.

潘成伟(1989－),男,安徽合肥人,博士,CCF 会员,主要研究领域为计算机图形学,虚拟现实,计算机视觉.



王少荣(1974－),男,博士,讲师,主要研究领域为计算机图形学,可视化,虚拟现实.

张建国(1990－),男,硕士,CCF 学生会员,主要研究领域为计算机图形学,计算机视觉.

汪国平(1964－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为计算机图形学,虚拟现实,人机交互.