基于多视点视频的虚拟会议显示与合成

孙立峰,李 放,钟玉琢,杨士强
(清华大学计算机科学与技术系,北京 100084)

摘 要: 构造能够支持与会者群组间自然方式交流与交互的虚拟会议人机交互环境是虚拟会议要解决的核心问题.本文介绍了建立半沉浸的投影显示设备构造虚拟环境的方法,能够实现与会者之间空间位置的重构;提出了基于水线的视频对象分割算法,能够稳定地从虚拟会议静态背景环境实时提取与会者对象,以及基于位置的多视点视频重构合成方法.

关键词: 多视点视频; 交互; 虚拟会议; 合成
中图分类号: TP391 **文献标识码**: A **文章编号**: 0372-2112 (2005) 02-0193-04

Multiview Video Based Virtual Teleconferencing Synthesizing

SUN Li-feng, LI Fang, ZHONG Yu-zhuo, YANG Shi-qiang
(Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)

Abstract: Building a human-computer interaction environment that supports natural communication and interaction among participator groups is the key issue of virtual teleconferencing system. We introduced a method that uses multiple projection screens to construct semi-immerse virtual environment, which established the correct space relationship of participators, proposed a watersheds based algorithm for stably real time video object extracted from static background and implemented position based multi-view video reconstruction and synthesizing.

Key words: multi-view video; human-computer interaction; virtual teleconferencing; synthesizing

1 引言

虚拟会议系统[1~3]概念的提出,是为了研究实现一种能够支持不同领域的人们通过交互与协作解决复杂问题的协作环境.一方面,在虚拟会议空间中,建立了统一的空间坐标系,定义了会议进行的空间范围,对虚拟会场中的场景进行描述,构造出具有真实感的虚拟会场;另一方面,通过为与会者赋予空间属性,根据与会者在虚拟会议空间中的位置与观察方向合成本地虚拟会场,与会者以"人在回路"的方式加入到虚拟会议应用中,通过身体语言、眼神接触等自然、直观的交互行为进行交互.

构造能够支持与会者群组间自然方式的交流和交互的虚拟会议显示交互环境是虚拟会议要解决的核心问题.虚拟会议空间合成的特点与要求体现在以下几个方面:

® 空间感

自然的交互行为,如身体语言、眼神接触、凝视感知等,都必须依赖于特定的空间参照系来完成.虚拟会议人机交互环境必须能够在虚拟会场中为与会者建立正确的空间逻辑关系,使与会者能够获得对其他与会者的位置感和方向感.

® 真实感

虚拟会议是一种群组协作应用,为实现如同实际环境中面对面交流的效果,要求虚拟会议的人机交互环境提供真实的交流感受,如实际尺寸的与会者影象(Life Size Video)表情、行为等.

® 实时性

与会者之间的交互具有很强的实时性,虚拟会议空间的合成必须实现对实时交互的合成,才能支持与会者之间正常的交流.

TelePort[1]使用大屏幕投影仪来构造一个沉浸显示环境,该系构造了一个显示房间,有一面墙被整个地覆盖以大屏幕投影,使用视觉跟踪来实时改变虚拟环境. GreenSpace, Panorama, FreeWalk[4]在 CAVE 的环境中采用三维平面纹理贴图的合成方法支持沉浸与交互.上述原型系统存在硬件要求过高的缺陷,我们建立半沉浸的投影显示设备构造虚拟环境,提出基于水线的视频对象分割算法,实现稳定的虚拟会议静态背景环境下与会者对象实时提取;采用线性视点合成(View Synthesis)算法,实现了基于位置的多视点视频重构合成.

本文组织结构如下:第 2 部分介绍基于一般投影变换的虚拟会场构造与显示,第 3 部分介绍与会者视频替身的构造与合成的算法与实验结果,第 4 部分是系统原型介绍,最后给

收稿日期:2003-11-17;修回日期:2004-12-09
基金项目:国家"863"项目(No. 2002AA119040);国家自然科学基金(No. 60273008)

出是结论和对未来工作的想法.

2 虚拟会场构造与显示

2.1 虚拟会场构造

为了实现具有沉浸感的虚拟会议空间,我们设计了如图 1 所示的虚拟会议显示环境.

虚拟会议显示环境由三面大屏幕投影共同构造一个半沉浸的虚拟会议空间,这种设计不仅能够实现虚拟会议空间合成的空间感与真实感,同时能够允许本地多个与会者共享同一个虚拟会场,支持更多与会者的会议研讨应用.

2.2 虚拟场景投影显示

通过多屏幕构造虚拟会场显示环境,需要保持场景在不同投影平面的连接处空间上的连续性,为此,我们提出并实现了基于一般透视投影变换的多屏虚拟会场拼接方法,消除不同屏幕连接处的空间畸变.

首先,我们引入了虚拟视点的概念,表示与会者当前在虚拟会议空间中的位置、观察方向,其视角范围是每一屏幕对应视角范围的总合,如图 1 所示. 图中观察体 M 的投影参考点位于 z 轴上,观察体 L 和观察体 R 的投影参考点不在 z 轴上,即一般透视投影变换,观察体 L、M、R 共同构成了与会者虚拟视点所对应的观察体,决定了与会者在其空间位置与方向所能观察到的虚拟会场中的场景.

我们采用针孔摄象机模型进行透视投影. 对于中间的观察平面,设投影参考点位于 z 轴的 z_{prp} 处,且观察平面在 z_{vp} 处,如图 2 所示. 其中, $d_p = z_{prp} - z_{vp}$,是投影参考点到观察平面的距离. 对于左右观察平面,我们通过以下两个步骤获得其对应的一般透视投影变换矩阵[5]:

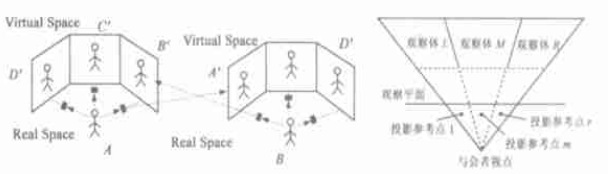

图 1 虚拟会议人机交互环境与多屏虚拟会场构造

(1) 错切观察体使得棱台的中心线垂直于观察平面;

(2) 用相应的 $1/z$ 缩放因子缩放观察体.

使一般透视观察体与投影窗口对齐的错切操作如图 2 所示,此变换将棱台中心线上所有点,包括窗口中心错切到与观察平面垂直的线上.

基于一般透视投影变换的多屏虚拟会场绘制能够实现相邻屏幕边缘显示场景的空间连续性,做到虚拟会场的无缝拼接.

3 与会者视频合成

3.1 视频替身模型

多用户虚拟环境中用户模型的建立是感知用户存在及其行为的基础. 研究表明,在虚拟环境中,即使是引入简单的用户模型也会在很大程度上增强虚拟环境的表现力和可感知性[5,6]. 我们引入替身的概念模型来描述与会者实际形象和行为在虚拟会议空间中的表现,包括三个组成部分:

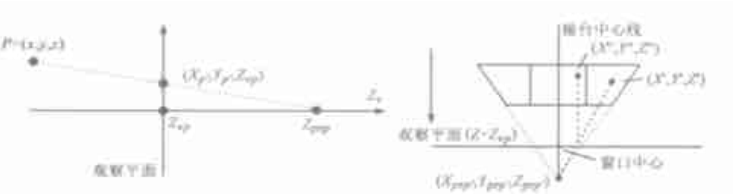

图 2 透视投影与一般透视观察体的错切操作

(1) 空间标识

与会者平面是表征与会者空间状态的三维平面,该平面通过以下方式对与会者进行空间标识:

® 与会者平面的空间位置表示在与会者虚拟会议空间中的位置;

® 与会者平面的法矢量表示与会者的观察方向;

® 该平面的大小反映合成到虚拟会议空间中的与会者显示比例.

(2) 视频对象

视频对象是从多路视频流的背景环境中分离出的仅包含与会者的视频图像.

(3) 视频替身 (Video Avatar)

视频替身是根据与会者空间标识对与会者视频对象进行空间变换,如旋转、缩放、变形等得到的用于在虚拟会议空间中进行合成的与会者影象. 与会者视频合成就是将与会者视频替身以透明纹理的方式映射到虚拟会议空间中的与会者平面,实现真实感、空间感的合成,如图 3 所示.

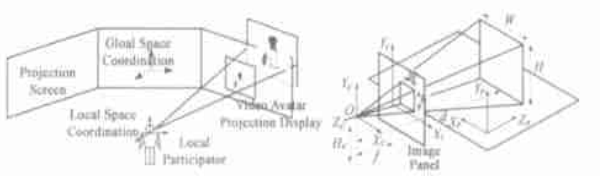

图 3 与会者视频替身及其透视投影

3.2 基于水线分割的视频对象提取与跟踪算法

稳定的视频对象提取是视频替身合成的基础,由于虚拟会议实际应用环境大部分是在静态场景中,我们研究了基于形态学计算的视频对象提取与跟踪算法. 图 4 所示是算法流程图,分为帧内分割和帧间跟踪两个部分. 帧内分割首先利用

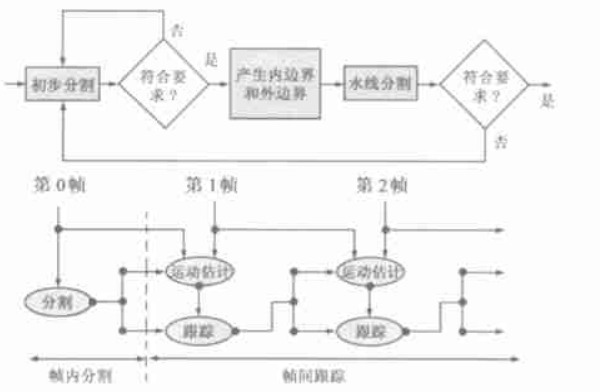

图 4 视频对象提取与跟踪算法流程图

多帧视频序列的差值计算和平滑滤波,获得视频对象的大致区域;然后运用形态学的腐蚀、膨胀运算构造视频对象内外边界,通过改进的多值水线分割算法精确提取对象边界;帧间跟踪利用前一帧分割的结果进行鲁棒性运动估计,再与当前帧的图像进行比较,在当前帧中再次使用形态学运算进行修正提取精确的对象轮廓.

3.2.1　初始轮廓分割　初始轮廓分割的策略是直接比较动态视频中相邻两帧图像的亮度信息:

对于前景视频对象加入造成的光线强弱变化,在计算亮度差值时进行了均衡处理,计算背景帧的平均亮度以及当前帧的平均亮度;

由于观察噪声的存在,为了把由于观察噪声和对象运动所引起的非零偏差区分开,引入偏差阈值 T, T 可以选择固定值,也可以自适应地进行调整,我们采用由与会者交互调节的方法对阈值进行动态调整. 则有:

$$dY(i,j) = Yn(i,j) - Yb(i,j) * \overline{Yn}/\overline{Yb}$$
$$Y0(i,j) = \begin{cases} Yn(i,j) & \cdots dY(i,j) > T \\ 0 & \cdots\cdots dY(i,j) \leq T \end{cases} \quad (1)$$

根据初始轮廓分割获得封闭区域,得到视频对象掩膜图像(Mask Image),这个封闭的掩膜区域记作 R_{user},其边界记作 B_{user}. 对二值区域 R_{user} 做形态学的膨胀、腐蚀运算[8],得到了外区域 R_{outer} 和内区域 R_{inner},其边界分别为 B_{outer}, B_{inner}. 由形态学中膨胀、腐蚀的性质有:

$$R_{inner} \subset R_{user} \subset R_{outer} \quad (2)$$

3.2.2　水线分割　在进行分类之前,必须先对区域内的图像进行图像的平滑简化,使得属于同一物体区域的像素点平滑,消除噪声,并且使区域与区域之间的边界得到保持,我们采用 Salembier[9] 提出的基于形态学运算的滤波器:

开-闭重构运算:　$\Psi_k^{(rec)} = \gamma_k^{(rec)} \phi_k^{(rec)}$　(3)

序列滤波器(结构元素的最大尺寸为 n):

$$\Psi_n^{(rec)}(\Psi_{n-1}^{(rec)} \cdots \Psi_k^{(rec)} \cdots \Psi_2^{(rec)}(\Psi_1^{(rec)})) \quad (4)$$

考虑到运算复杂度,我们采用简化的序列滤波器 $\Psi = \Psi_n^{(rec)}(\Psi_{n/2}^{(rec)})$ 进行区域的平滑简化.

由式(2)知,区域 $R_{outer} - R_{inner}$ 是不确定区域(Uncertain Region),以下简称 UR,我们所采用改进的多值水线彩色图像分割算法(Chuang Gu 和 Ming-Chieh Lee[9])进行区域分割.

在已经得到的经过平滑的 UR 中,内外谷地的边界就是所分割对象的边界. 把 B_{outer} 和 B_{inner} 的像素点作为谷地的中心,分别作为标记点,从两个边界向内部进行区域扩张. UR 中间的点如果被扩张到,则计算它与那个标记过的区域在某种意义上最近,将其加入该区域. 直到所有的点都被扩充完毕. 我们用区域中各个点的 RGB 分量的平均值来作为该区域的 RGB 值. 当扩张到一个像素时,计算该像素点与该区域在色彩空间上的距离,由式(5)所示,考虑到开方为浮点运算,故采用 d_2 作为距离值.

$$d_i = \sqrt{(r-\overline{r}_i)^2 + (g-\overline{g}_i)^2 + (b-\overline{b}_i)^2} \quad (5)$$

这样,当内边界扩展区域和外边界扩展区域相交时,相交的边界就是视频对象的边界.

3.2.3　基于滑窗的快速算法　为了达到实时处理的要求,我们引入滑动窗口的概念,可以快速实现以矩形结构元素进行腐蚀、膨胀的形态学运算. 滑动窗口是附加在图像的行与列上的小缓冲区,每个缓冲区包含若干可存储一个象素值的单元. 通过记录已经比较过的局部信息,在以后的计算中可以直接读取并参与比较,以达到减少并消除重复比较的计算冗余. 与基于定义的算法相比较,滑窗算法的计算时间与结构元素的周长是线性关系,用 MPEG-4 测试序列,该快速算珐比传统的形态学运算在不同结构元素尺寸下速度均有显著提高,能够实时进行处理.

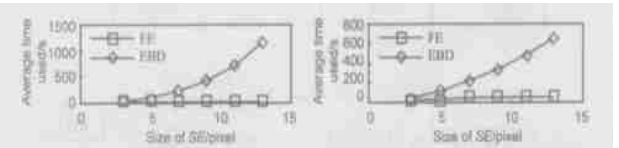

图 5 灰度图象与视频序列条件下腐蚀运算时间比较

3.3　基于空间位置的多视点视频替身合成

虚拟会议空间的合成是面向与会者的,即在不同虚拟会议终端按照不同与会者在虚拟会议空间中的位置、方向,将与会者视频替身以透明纹理的方式映射到虚拟会议空间中的与会者平面进行合成,如图 6 所示.

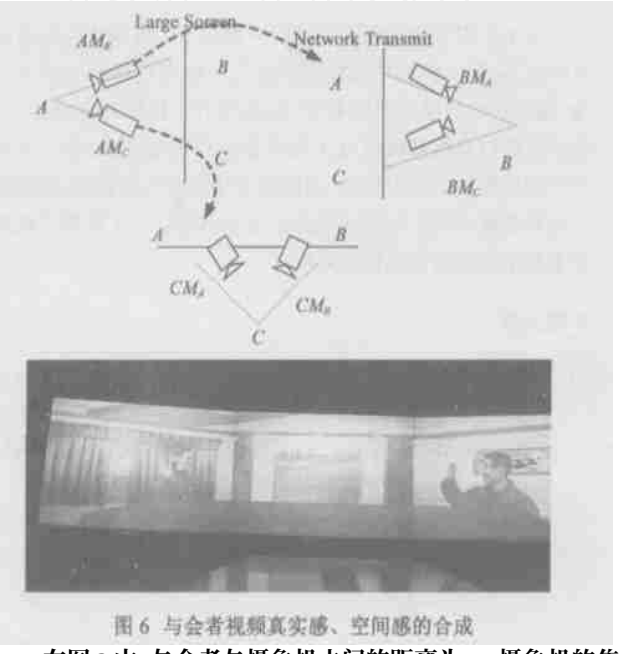

图 6　与会者视频真实感、空间感的合成

在图 3 中,与会者与摄象机之间的距离为 d,摄象机的焦距为 f,沿 x 轴的缩放因子为 α_x,沿 y 轴的缩放因子为 α_y,与会者视频对象在摄象机成像平面中的宽高以像素表示分别为 w, h,根据摄象机针孔透视投影模型,与会者视频替身的尺寸由下式计算:

$$W = w * d/\alpha_x \quad H = h * d/\alpha_y \quad d = Hc * \alpha_y/|v-h| \quad (6)$$

Hc 为摄象机光心距地面的高度,与会者成像区域边界矩形顶点坐标表示为 (u,v), $(u+w,v)$, $(u,v-h)$, $(u+w,v-h)$. 计算过程中用到的摄象机参数预先通过摄象机定标(Camera Calibration)[11] 获得的.

4 原型与实验

我们的原型系统由多个虚拟会议应用单元与多点合成服务器单元(Multiparty Composting Server,以下简称 MCS)组成,会议应用单元与 MCS 采用 100/10M 局域网相连,如图 7 所示.目前该系统集成了基于水线的视频对象抽取,基于感知的多点传输机制与视频替身的位置合成.

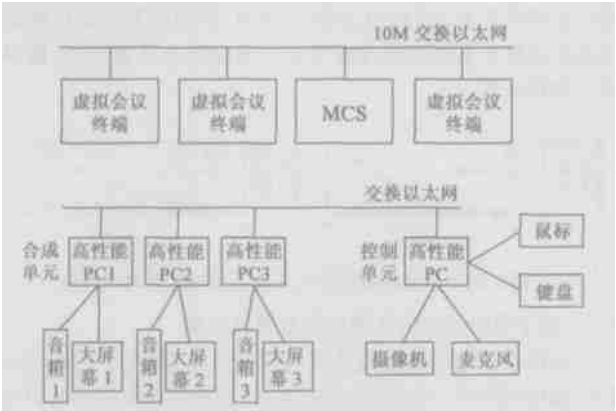

图 7 多视点视频虚拟会议原型系统结构

5 结束语

本文介绍了通过联网的微机驱动多个大屏幕拼接构造具有空间感、真实感的半沉浸虚拟会议人机交互环境的实现方法.原型系统的应用实验表明,在这样的人机交互环境中,与会者能够以自然直观的方式进行交流和研讨协作.进一步的工作将研究脱离投影屏幕,利用真实环境中的投影面,将虚拟会场和实际真实会场融合的算法和实现技术,以及基于视线的多视角视频的合成快速算法.

参考文献:

[1] Christian J Breitender, et al. TELEPORT - an augmented reality teleconferencing environment [A]. Proc. 3rd Eurographics Workshop on Virtual Environments Coexistence & Collaboration [C]. Monte Carlo, Monaco, Springer-Verlag, London, UK, February 1996. 41 - 49.

[2] Vali Lalioti, et al. Virtual meeting in cyberStage [A]. ACM VRST '98 [C]. TAIPEI, TAIWAN, New York, NY, USA: ACM Press, 1998. 2 - 5.

[3] Hideyuki Nakanishi, et al. FreeWalk: A 3D virtual space for casual meetings[J]. IEEE Multimedia, 1999, 6(2): 20 - 28.

[4] Donald Hearn, et al. Computer Graphics[M]. USA: Prentice Hall Inc, 1997.

[5] Maia Garau, et al. The impact of eye gaze on communication using humanoid avatars[A]. ACM SIGCHI '01[C]. Seattle, WA, New York, NY, USA: ACM Press, 2001. 309 - 316.

[6] Emilee Patrick, et al. Using a large projection screen as an alternative to head-mounted displays for virtual environments[A]. ACM SIGCHI '00[C]. Hague, Netherlands, New York, NY, USA: ACM Press, 2000. 478 - 485.

[7] J Serra. Image Analysis and Mathematical Morphology [M]. USA: Academic Press Inc, 1982.

[8] Chuang Gu, Ming-Chieh Lee. Semiautomatic segmentation and tracking of semantic video objects[J]. IEEE Trans on Circuits and systems for video technology, 1998, 8(5): 572 - 584.

[9] P Salembier, M Pardas. Hierarchical morphological segmentation for image sequence coding[J]. IEEE Trans on Image processing, 1994, 3(5): 639 - 651.

[10] Tsai R Y. An efficient and accurate camera calibration technique for 3D machine vision[A]. Proc of IEEE Conference of Computer Vision and Pattern Recognition CVPR '86[C]. Miami, FL, USA, 1986. 364 - 374.

作者简介:

孙立峰 男,1972 年生,2000 年毕业于国防科技大学,获博士学位,现为清华大学计算机科学与技术系讲师,主要研究方向为交互多视点视频,异构网络流媒体.

李 放 男,1982 年生,现为清华大学计算机科学与技术系硕士研究生,主要研究方向为视频合成与视频分析.

钟玉琢 男,1938 年生,现为清华大学计算机科学与技术系教授,博士生导师,主要研究方向为视频编码与流媒体,数字家电网络平台.

杨士强 男,1952 年生,现为清华大学计算机科学与技术系教授,博士生导师,主要研究方向为视频分析与网络多媒体.