



华夏英才基金学术文库

王涌天 陈 靖 程德文 著

增强现实技术导论



科学出版社

 华夏英才基金学术文库

增强现实技术导论

王涌天 陈 靖 程德文 著

科学出版社

北 京

内 容 简 介

增强现实技术的目标是将计算机生成的虚拟环境与用户周围的现实环境融为一体,使用户从感官效果上确信虚拟环境是其周围真实环境的组成部分。因此,增强现实技术具有虚实结合、实时交互、三维注册的特点。围绕这三方面,本书主要介绍增强现实系统的标定方法、实现虚拟与真实环境无缝融合的显示设备、有标识的六自由度跟踪注册技术、无标识的六自由度跟踪注册技术以及适用于移动终端的增强现实系统,并给出了典型的应用实例,包括基于增强现实技术的轻量级飞行模拟器开发和基于增强现实技术的圆明园数字重现定点观察设备。

本书可作为计算机、信息工程等相关专业高年级本科生、研究生的教材以及相关领域研究人员的参考用书。

图书在版编目(CIP)数据

增强现实技术导论 / 王涌天,陈靖,程德文著. —北京:科学出版社,2015.6
ISBN 978-7-03-044025-9

I. ①增… II. ①王…②陈…③程… III. ①数字技术—研究
IV. ①TP391.9

中国版本图书馆 CIP 数据核字(2015)第 064846 号

责任编辑:孙伯元 陈 婕 / 责任校对:桂伟利

责任印制:张 倩 / 封面设计:陈 敬

科学出版社 出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

中国科学院印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2015 年 6 月第 一 版 开本:720×1000 1/16

2015 年 6 月第一次印刷 印张:15 1/4

字数:290 000

定价:128.00 元

(如有印装质量问题,我社负责调换)

序

20 世纪 90 年代初,伴随着虚拟现实(virtual reality,VR)技术的发展,增强现实(augmented reality,AR)技术应运而生。增强现实技术是虚拟现实技术的一个重要分支。它借助显示技术、交互技术、传感技术和计算机图形技术将计算机生成的虚拟物体与用户周围的现实环境融为一体,使用户从感官效果上确信虚拟物体是其周围环境的组成部分。因此,增强现实技术具有虚实融合、实时交互、三维注册的特点,在军事、医疗、科研、工业、教育、培训、文化、娱乐等领域都有着广泛的应用前景。

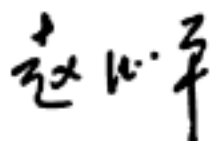
十多年前,增强现实技术对于大众来说或许还是一个抽象的概念,那时的增强现实技术还处于实验室研究阶段。在虚实融合方面,由于增强现实技术需要同时融合现实世界与虚拟世界,因此研究的重点主要是设计开发能够实现虚实融合显示的显示器设备,即不同种类的显示器原型系统,从头盔显示器、视网膜显示器、投影仪、手持式显示器到空间式显示器相继出现;在实时交互方面,电磁和光学人体动作跟踪器以及后期出现的语音、手势乃至体感交互技术,实现了人与虚拟和现实世界自然交互的愿望;在三维注册方面,重点是实现用户与自然场景间的实时六自由度位置和方位跟踪计算,更准确、更鲁棒、更实时的姿态计算算法是这一方向的研究重点。该技术经历了早期的基于标志物的六自由度跟踪注册,并逐渐过渡到基于自然特征的六自由度跟踪注册。从增强现实系统来讲,20 世纪 90 年代末期的增强现实系统将用户局限在室内较小的工作范围内,用户的运动范围受限。21 世纪初期开始出现了户外移动增强现实系统。虽然用户的运动范围得以扩展,但是受限于当时的技术条件,增强现实系统的重量以及便携性成为发展的瓶颈。而所有这些技术的进展基本都仅局限在实验室,不为大众所熟知。

近年来,随着增强现实技术以及移动互联网、高性能计算等技术的不断发展,结合自然场景的增强现实技术悄然走进了大众的视线。2009 年,荷兰的一家软件公司面向智能手机开发的首款增强现实浏览器问世,预示着增强现实技术开始从实验室阶段迈向了商用阶段。2012 年,谷歌眼镜的横空出世直接将增强现实技术这一陌生的科技名词推向大众,激发了人们对增强现实技术的研究热情。2015 年初,微软公司全息眼镜产品的发布,更让业界为之振奋,一时好评如潮。佩戴上该

款头盔显示器,用户可置身于奇妙的增强现实世界中,实现了虚拟与真实环境的良好融合、自然的用户交互以及精准的三维定位,从而将数字化的信息世界与真实的三维世界紧密地联系起来。相信在不远的未来,增强现实技术的应用将会涉及人类生活的诸多方面,直接影响人们观察世界和理解世界的方式。

北京理工大学是国内最早开展增强现实关键技术研究的单位之一,从20世纪90年代末期就已着手研发适用于增强现实系统的头盔显示器与跟踪注册技术;先后设计开发了视频透视式增强现实头盔显示器与光学透视式头盔显示器;跟踪注册技术的研发也经历了从有标识跟踪注册到无标识跟踪注册的阶段性发展。21世纪初期,在有关政府部门的支持下,设计开发了基于增强现实技术的定点观察设备,并成功地将其应用于圆明园的数字化场景重现中。

该书是作者在多年从事增强现实技术研究的基础上所著,是国内第一部系统介绍增强现实的专著,从增强现实技术的概念与起源、增强现实的显示设备、标定技术、注册定位技术以及移动增强现实系统应用研究等几个方面全面详细地介绍了增强现实技术,并以轻量级增强现实飞行模拟器、定点增强现实观景器为例介绍了增强现实的实际应用。相信该书会给从事增强现实研究的科技工作者带来有益的启示,并对我国增强现实技术研究、应用系统开发和人才培养起到积极的推动作用。



2015年4月9日

前 言

智能手机、智能可穿戴设备,特别是谷歌眼镜的横空出世,激起了人们对增强现实技术的好奇心。人们希望能够拨开增强现实技术的面纱,一探究竟。增强现实技术本身是一个跨学科的研究领域,它综合使用了不同研究领域的多种技术,如虚拟现实技术、计算机视觉技术、人工智能技术、可佩戴移动计算机技术、人机交互技术、生物工程学技术等。其目标是将计算机生成的虚拟环境与用户周围的现实环境融为一体,使用户从感官效果上确信虚拟环境是其周围真实环境的组成部分。由此可见,为构建一个成功的增强现实系统,需要综合使用不同研究领域的技术,这对初学者而言是非常困惑的事情。增强现实系统都有哪些技术支撑?都需要使用哪些硬件设备?怎样构建一个成功的增强现实系统?这些是本书所要阐述的主要内容。

本书共8章。第1章主要阐述增强现实技术的概念、起源以及构成增强现实系统的主要关键技术;第2章重点介绍增强现实系统中不同类型的融合显示设备;第3章介绍增强现实系统的标定,包括视频透视式和光学透视式两种不同的增强现实系统标定方法;第4章介绍增强现实系统中较常采用的基于标识的跟踪注册算法及其基本原理;第5章就无标识跟踪注册算法进行详细介绍;第6章主要阐述移动增强现实系统的研究进展以及其在Android系统中的应用;第7章和第8章介绍两个不同的增强现实应用实例。

本书是在作者在虚拟现实与增强现实方面的研究生课程讲义的基础上形成的,目的是为初学者介绍增强现实系统的基本概念、理论知识以及构建增强现实系统所需要的主要技术手段和方法。通过阅读本书,读者可以对增强现实技术有全面、深入的了解,并对进入该研究领域进行知识准备。

完成本书的合作教师还有刘越教授、翁冬冬副教授。北京航空航天大学的梁晓辉教授对本书提出了宝贵的建议,在此表示感谢。在本书写作过程中,对本书的内容撰写做出贡献的研究生有李玉、张运超、桂振文、林一、许晨、梁杏、蔡郡、严雷、黄秀杰、曹若琛、孙磊等,以及一些已毕业的研究生,在此谨向他们表示最诚挚的感

谢。此外,本书在写作过程中参考了有关书籍和文献(见本书列出的参考文献),在此也向这些作者一并致谢。

增强现实技术是一个多学科交叉的新兴研究课题,由于作者水平有限,书中难免存在不足之处,敬请读者不吝指正。

目 录

序

前言

第1章 引言	1
1.1 增强现实技术的概念和发展	1
1.1.1 虚拟现实技术	2
1.1.2 增强现实技术	3
1.1.3 增强现实技术与虚拟现实技术的比较	4
1.1.4 增强现实系统的性能评价	6
1.2 增强现实系统中的硬件与软件包	7
1.2.1 增强现实系统中的硬件	7
1.2.2 增强现实系统软件包	16
1.2.3 增强现实技术的国内外研究机构	17
1.3 增强现实技术的难点	19
1.3.1 增强现实系统中的头盔显示技术	20
1.3.2 增强现实系统中的跟踪注册技术	22
1.3.3 增强现实系统的交互技术	24
1.4 增强现实技术的应用领域	27
参考文献	38
第2章 增强现实系统的显示技术	43
2.1 显示设备的分类及概述	43
2.1.1 屏幕显示器	43
2.1.2 头盔显示器	48
2.1.3 投影式头盔显示器	52
2.1.4 视网膜直接成像显示器	55
2.1.5 投影显示设备	56
2.1.6 智能可穿戴显示设备	60
2.2 头盔显示器的分类及其核心技术	61

2.2.1	自由曲面头盔显示器	61
2.2.2	全息波导型头盔显示器	62
2.2.3	几何波导型头盔显示器	63
2.2.4	高性能头盔显示技术	64
2.2.5	真实立体感头盔显示技术	66
2.3	增强现实头盔显示器的设计要点	73
2.3.1	头盔显示器的组成与原理	73
2.3.2	头盔显示器的光学系统设计要点	74
2.3.3	部分国内外头盔显示器的设计	77
	参考文献	80
第3章	增强现实系统的标定	86
3.1	摄像机的几何模型和坐标变换	86
3.1.1	图像坐标系与像素坐标系	87
3.1.2	图像坐标系与摄像机坐标系	87
3.1.3	摄像机坐标系与世界坐标系	88
3.2	摄像机标定	89
3.2.1	线性模型的摄像机标定	89
3.2.2	基于平面方格点的摄像机标定	92
3.2.3	非线性模型的摄像机标定	95
3.3	基于透视式头盔显示器的增强现实系统标定	95
3.3.1	视频透视式增强现实系统的标定	95
3.3.2	光学透视式头盔显示器的标定技术	96
	参考文献	104
第4章	基于标识的增强现实跟踪注册技术	106
4.1	基于标识的跟踪注册技术概述	108
4.2	常用的标识点	111
4.2.1	ARToolkit 方形标识点	111
4.2.2	SCR 标识点系统	116
4.2.3	环形标识点	118
4.2.4	Sony 计算机实验室标识点	121
4.2.5	室内装饰性人工标识	122
4.2.6	红外标识系统	125

4.3 各种人工标识系统中的识别定位算法比较	127
4.3.1 人工标识的形状	128
4.3.2 标识的颜色	128
4.3.3 定位标识区域	129
4.3.4 标识的可识别范围	129
4.3.5 特征点提取	130
4.3.6 标识的编码信息量	130
4.3.7 理想特征的概要	130
参考文献	130
第5章 无标识增强现实三维跟踪注册	134
5.1 基于场景平面的增强现实跟踪注册	135
5.1.1 系统初始化	135
5.1.2 摄像机位姿估计	136
5.1.3 实验结果	137
5.2 基于模型和关键帧的注册方法	138
5.2.1 3D 场景模型与关键帧的创建	138
5.2.2 选择参考图像帧	139
5.2.3 宽基线匹配	140
5.2.4 摄像机位姿求解	141
5.2.5 实验结果	142
5.2.6 算法评价	142
5.3 基于图像匹配的无标识跟踪注册方法	144
5.3.1 关键帧匹配	144
5.3.2 2D 图像匹配算法	145
5.3.3 实验结果	147
5.4 基于 SIFT 特征的无标识点跟踪注册方法	148
5.4.1 SIFT 特征描述算子	149
5.4.2 SIFT 关键点的匹配	150
5.4.3 基于 SIFT 特征的摄像机位姿估计	151
5.5 基于运动结构重建的跟踪注册算法	152
5.5.1 MonoSLAM 算法	152
5.5.2 PTAM 算法	153

5.6 混合跟踪注册	157
5.6.1 根据传感器类型分类	158
5.6.2 根据传感器融合方式分类	160
5.7 无标识跟踪注册算法的未来研究方向	162
参考文献	163
第6章 移动增强现实系统	169
6.1 概述	169
6.2 移动增强现实系统的研究现状	171
6.2.1 移动增强现实技术的学术研究	171
6.2.2 移动增强现实技术的商业应用	174
6.3 面向智能手机的移动增强现实核心技术	178
6.3.1 主体结构框架	179
6.3.2 跟踪定位算法	181
6.3.3 海量场景的目标识别	183
6.3.4 人机交互方式	185
6.3.5 数据通信和传输	185
6.3.6 功耗	186
6.4 Android 系统开发环境	187
6.4.1 Android 操作系统	187
6.4.2 手机硬件环境和软件开发环境	189
6.4.3 OpenCV 库编译和基于 NDK 的开发	190
6.4.4 部分增强现实功能实现	191
参考文献	199
第7章 增强现实技术的应用实例——轻量级飞行模拟器	203
7.1 增强现实轻量级飞行模拟器	203
7.1.1 系统功能	203
7.1.2 系统结构	204
7.1.3 座舱视景的虚实融合原理	206
7.1.4 样机系统实现与测试	208
7.2 改进型增强现实轻量级飞行模拟器	212
7.2.1 改进型增强现实飞行座舱系统设计	212
7.2.2 座舱头部跟踪系统的实验模拟	213

7.2.3 跟踪器的实验对比结果	214
参考文献	215
第8章 增强现实技术的应用实例——圆明园定点观景器	216
8.1 第一代定点增强现实观景器	216
8.2 第二代增强现实定点观景器	219
8.2.1 第二代增强现实定点系统的优点	219
8.2.2 第二代系统的构建	220
8.2.3 增强现实定点观景器的用户使用情况调查	221
8.3 AR-View 增强现实定点观景器	224
8.4 随动式增强现实定点观景器	226
8.4.1 系统构成	229
8.4.2 样机设备与操作使用	229

第 1 章 引 言

1.1 增强现实技术的概念和发展

在信息时代,人们处于计算机日益普及的世界中,从当初几人共享一台计算机的模式转变为一人拥有多台计算机。人们的生活、工作已离不开计算机生成的信息数据。但是,目前计算机仍处于智能化的初级阶段,只是帮助人类进行逻辑思维的有力工具,要使计算机成为人类进行形象思维的帮手,仍有较长的路要走。这一切首先要求计算机能够适应人们所习惯的信息获取方式和思维过程,与真实的环境进行自然融合,这也是计算机技术发展的趋势。图 1.1 展示了增强现实的概念,将虚拟的三维模型与周围的真实环境完美地融为一体。增强现实系统可以将计算机生成的信息(包括文字、图像、三维物体等)以视觉融合的方式叠加至真实场景中,在用户眼前呈现出“增强”了的世界。在理想情况下,用户无法辨别哪个是虚拟物体,哪个是真实物体,看到的将是完整的融合显示场景。因此,增强现实从字面上来说就是对现实的增强,而不是对真实环境的取代。



图 1.1 圆明园的真实场景与圆明园虚实融合场景

增强现实(augmented reality, AR)是一种利用计算机产生的附加信息对使用者所看到的真实世界景象进行增强或扩张的技术,是跨学科的综合体。增强现实和虚拟现实(virtual reality, VR)联系非常紧密,是虚拟现实发展的扩展或者分支,也

是近年来的一个研究热点。增强现实系统综合使用了不同研究领域的多种技术,如虚拟现实技术、计算机视觉技术、人工智能技术、可佩戴移动计算机技术、人机交互技术、生物工程学技术等。因此,谈到增强现实时首先需要讨论虚拟现实。

1.1.1 虚拟现实技术

“虚拟现实”一词是由美国 VPL Research 公司的奠基人 Lanier 于 20 世纪 80 年代末首次正式提出的。他将虚拟现实描述为一种逼真地模拟人在自然环境中的视觉、听觉、运动等行为的人机界面技术。自从冯·诺依曼结构的计算机诞生以来,计算机的形式化、逻辑指令化和串行方式与人类与生俱来的形象化、模糊概念化、联想和并行接收处理信息的能力不协调。计算机的运行方式与用户的认识空间和信息空间之间存在鸿沟。从 1965 年 Sutherland 提出“计算机图形学”概念至今,人类一直致力于建立尽可能和谐理想的人机环境,这正是虚拟现实技术的历史渊源。

伴随着计算机软硬件技术和传感技术的发展,虚拟现实技术已经不仅以优化人机交互环境为目的,而且更多地结合了计算机仿真和图形学,并借助人类似妙的想象力,为开拓人类的信息空间和认识手段提供了强有力的途径。虚拟现实系统实现了人与计算机之间通过语言、手势等自然方式进行信息交互,使用户逐渐从“主动适应”发展到“被适应”的地位。更重要的是,用户不仅能够通过虚拟现实系统感受到在客观物理世界中所经历的“身临其境”的逼真感,而且能够突破物理空间、时间以及其他客观限制,感受到在真实物理世界中无法亲身经历的体验。因此,虚拟现实技术具有超越现实的虚拟性。

虚拟现实系统的基本特征被描述为如图 1.2 所示的“浸没-交互-构想”,即 immersion-interaction-imagination。上述三个特征在虚拟现实系统中相互影响并共同存在。虚拟现实的第一个特征是“浸没”,又称为临场感、存在感,强调 VR 境界的逼真性,即用户在计算机所创建的三维虚拟环境中处于一种全身心投入的状态。该环境中的一切看上去是真的,听上去是真的,动起来是真的,甚至闻起来、尝起来等一切感觉都是真的,如同在现实世界中的感觉一样行动自如,让参与者有身临其境的感觉。第二个特征是“交互”,虚拟现实的交互特性主要通过虚拟现实系统中的交互接口设备实现,通过交互设备人们以自然的方式与虚拟现实环境进行交互操作。第三个特征是“构想”,强调对人类认知范围的扩展,不仅可以再现真实存在的环境,而且可以随意构想客观世界中不存在的,甚至不可能存在的环境。从图形学的角度来看,传统的计算机图形学强调了三维场景在屏幕上的二维显示。虚拟现实技术则强调三维场景的立体显示,并为参与者创建了在这个立体的三维虚

拟世界中的想象空间。上述三个基本特征使得 VR 技术区别于传统的计算机仿真、遥现技术、多媒体技术和科学计算可视化技术等。归结起来,虚拟现实的概念包括以下含义。

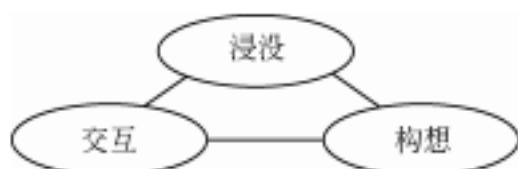


图 1.2 虚拟现实的三个特征图

(1) 虚拟环境是由计算机生成、实时动态、具有双眼立体视觉、三维听觉、触觉及嗅觉等感官信息表现的、逼真的三维图像环境,可以是某一特定现实世界的真实再现,也可以是虚构的世界。

(2) 理想虚拟现实技术的“实时感知”应该具有人的所有感知本领,包括视觉、听觉、触觉、嗅觉、味觉、力觉、运动等多种感知渠道。

(3) 使用“自然技能”是指计算机处理与人的头部转动、眼睛运动、手势和其他人体动作语言相对应的数据,对用户的输入作出实时响应,并分别反馈到感官。

(4) “传感设备”主要是指三维交互设备,包括立体头盔显示器、数据手套、三维鼠标和数据衣等。

虚拟现实技术综合了计算机仿真技术、计算机图形学、多媒体技术、人工智能技术、计算机网络技术、并行处理技术和多传感器技术模拟人的视觉、听觉、触觉等感官功能,旨在通过显示技术、交互技术、传感技术和计算机图形技术的手段,使用户从感官效果上沉浸在由计算机创造的虚拟环境中。目前,虚拟现实的应用已经涵盖了产品设计制造、模拟训练、远程医疗、太空探险、教育培训、科学研究、娱乐等各类军民领域。

1.1.2 增强现实技术

增强现实技术是借助光电显示技术、交互技术、多种传感技术和计算机图形与多媒体技术将计算机生成的虚拟环境与用户周围的现实环境融为一体,使用户从感官效果上确信虚拟环境是其周围真实环境的组成部分。因此,增强现实系统具有虚实结合、实时交互、三维注册的新特点^[1]。图 1.3 为 MIT 多媒体实验室开发的第六感装置,配合手势识别、人脸识别等先进技术,将计算机产生的虚拟信息直接投影至真实场景中^[2]。

图 1.4 为用户通过一个医疗教学增强现实系统可能看到的场景。该系统将人



图 1.3 MIT 多媒体实验室开发的第六感装置^[2]

脑解剖结构可视化,并准确地显示在观察对象的头部位置,用户将同时看到观察对象的头部外形与其大脑结构。系统不仅可以帮助医学院的学生学习大脑的解剖结构,还可以帮助手术医生准确定位手术部位,提高其临床技能^[3]。



图 1.4 医疗教学增强现实系统^[3]

从上述应用可以看出,增强现实技术借助计算机图形技术和可视化技术产生现实环境不存在的虚拟对象,并通过传感技术将虚拟对象准确地“放置”在真实环境中,并借助显示设备将虚拟对象与真实环境融为一体,呈现给用户一个感官效果真实的新环境。

1.1.3 增强现实技术与虚拟现实技术的比较

增强现实技术是随着虚拟现实技术的发展而产生的,因此二者存在不可分割的纽带关系。虽然它们所需要的基础技术相同,但二者存在较大的区别。虚拟现实技术综合了显示技术、交互技术、传感技术和计算机图形等技术,使得用户从感官效果上沉浸在一个与其周围环境完全不同的虚拟环境中。而增强现实技术借助显示技术、交互技术、传感技术和计算机图形技术将计算机生成的虚拟环境与用户周围的现实环境融为一体,使用户从感官效果上确信虚拟环境是其周围真实环境

的组成部分。因此,构建一个增强现实应用系统的目的并非以虚拟世界代替真实世界,而是利用附加的信息增强使用者对真实世界的观察和感知。增强的信息可以是虚拟的三维模型,也可以是真实物体的非几何信息。形象地说,虚拟现实系统试图将世界送入使用者的计算机,而增强现实系统却是把计算机带进使用者的真实环境。可以说增强现实技术在虚拟环境与真实世界之间架起了一座桥梁。具体而言,增强现实技术与虚拟现实技术的差异主要体现在以下四方面^[4]。

(1) AR 与 VR 最显著的差别在于两者对于沉浸感的要求不同。

VR 系统强调用户在虚拟环境中视觉、听觉、触觉等感官的完全沉浸,强调将用户的感官与现实世界绝缘,使其沉浸在一个完全由计算机控制的信息空间(cyberspace)之中。这通常需要借助能够将用户视觉与现实环境隔离的显示设备,一般采用沉浸式头盔显示器(immersive head-mounted display, IHMD)。而与之相反,AR 系统不仅不隔离周围的现实环境,而且强调用户在现实世界的存在性,并努力维持其感官效果的不变性。AR 系统致力于将计算机产生的虚拟环境与真实环境融为一体,从而增强用户对真实环境的理解。这就需要借助能够将虚拟环境与真实环境融合的显示设备,如透视式头盔显示器(see-through head-mounted display, STHMD)。

(2) AR 和 VR 关于注册(registration)的含义和精度要求不同。

在沉浸式 VR 系统中,注册是指呈现给用户的虚拟环境与用户的各种感官匹配。例如,当用户用手推开一扇虚拟的门时,用户所看到的场景就应该同步地更新为屋子里面的场景,一条虚拟小狗向用户跑过来,用户听到的狗吠声就应该是由远及近变化的,这种注册误差是视觉系统与其他感官系统以及本体感觉之间的冲突。而心理学研究表明,视觉往往占了其他感觉的上风。而在 AR 系统中,注册主要是指将计算机产生的虚拟物体与用户周围的真实环境全方位对准,而且要求用户在真实环境的运动过程中维持正确的对准关系。较大的注册误差不仅不能使用户从感官上相信虚拟物体在真实环境中的存在性及一体性,甚至会改变用户对其周围环境的感受,改变用户在真实环境中动作的协调性,严重的注册误差甚至导致完全错误的行为。

(3) AR 可以缓解对系统计算能力的苛刻要求。

一般来说,要求 VR 系统精确再现人们周围的简单环境,需要付出巨大的代价,而其结果在当前技术条件下也未必理想,其逼真程度总是与人的感官能力不相匹配。而 AR 技术则是在充分利用周围已存在的大量信息的基础上加以扩充,这就大大降低了对计算机图形能力的要求。

(4) AR 与 VR 应用领域的侧重不同。

虚拟现实系统强调用户在虚拟环境中的视觉、听觉、触觉等感官的完全浸没,对于人的感官来说,它是真实存在的。而对于所构造的物体来说,它又不存在的。因此,利用这一技术能模仿许多高成本的、危险的真实环境。其主要应用在虚拟教育、数据和模型的可视化、军事仿真训练、工程设计、城市规划、娱乐和艺术等方面。而增强现实系统并非以虚拟世界代替真实世界,而是利用附加信息增强使用者对真实世界的感官认识。因而其应用侧重于辅助教学与培训、医疗研究与解剖训练、军事侦察及作战指挥、精密仪器制造和维修、远程机器人控制等领域。图 1.5 是虚拟现实应用领域之一的虚拟教育示意,学生可以通过虚拟的人体内脏,形象化地理解生理学和解剖学的基本理论。加州大学的 Hoffman 研制的系统可以带领学生进入虚拟人体的胃部^[5],检查胃溃疡并可以“抓取”它进行组织切片检查。图 1.6 为增强现实领域的辅助教学与培训。不仅医生能够手持手术探针实时地对患者进行胸部活组织切片检查,而且系统可根据此时获得的切片组织情况决定手术探针的位置,指导医生完成手术。对比图 1.5 和图 1.6 可以看到,由于增强现实系统中真实环境的存在,用户不仅对融合环境的感知更具真实感,同时增强了对虚拟环境的感知。因而在某些应用领域,增强现实技术与虚拟现实技术相比更具感知优势。

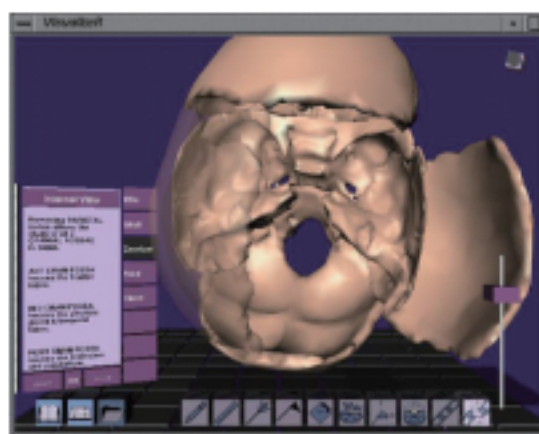


图 1.5 虚拟人体内脏^[5]

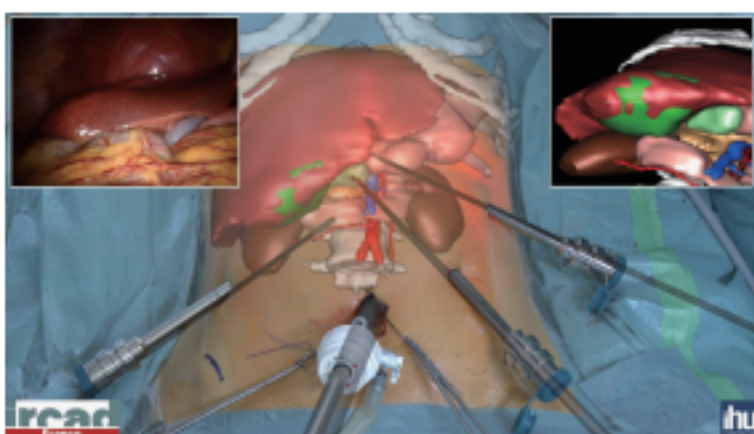
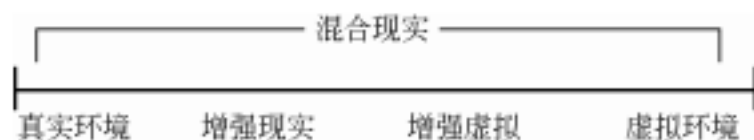


图 1.6 AR 辅助医疗手术过程^[6]

图 1.7 是 Milgram 等提出的从现实(reality)到虚拟(virtuality)连续变化的分类概念^[7],从图中可以很明显地区分 AR 和 VR 的关系。虚拟环境(virtual environment)包括虚拟现实、增强虚拟(augmented virtuality, AV)和增强现实。

1.1.4 增强现实系统的性能评价

与虚拟现实系统类似,增强现实系统是实时性系统。增强现实系统的用户期

图 1.7 从现实世界到虚拟世界的顺序图^[7]

望能够像在现实世界中一样,在增强信息空间中自由活动,用户所感觉到的信息变化应是协调一致的。衡量 AR 系统性能好坏的标准可以归纳如下。

(1) 增强场景图像的刷新频率,即图像的实时性。

(2) 虚拟场景图像和真实场景图像在照明、色彩、质感等物理特性方面的一致性。

(3) 虚拟场景和真实场景的注册精度。

(4) 增强场景与用户视点变化的一致性以及用户在增强场景中的运动精度和自由度。

关于实时性的主观性标准可以定义为:从用户视觉感受的角度看,增强场景的刷新频率以不使用户感觉到任何跳动感为依据。一般认为,没有跳动感的图像系统必须满足每秒 10 帧以上的绘制频率。基于现有图形系统的功能,绘制简单和中度复杂场景的能力基本没有问题。而且增强现实系统中的虚拟场景往往比虚拟现实系统简单,因此刷新频率基本上不是严重问题。但是图像质量距离“逼真”的标准还有差距,特别是有真实物体对照的情况下。人眼在这方面的分辨能力很强,虚拟与真实的一致性差会使用户产生视觉偏向,不能达到视觉增强的效果,甚至适得其反。在增强现实系统中,虚拟场景与真实场景的注册精度是系统成功与否的关键,根据人眼对相对位置差异的分辨能力,注册精度的目标是 0.5 弧分左右。而最后一条标准与虚拟现实系统的要求相同。

1.2 增强现实系统中的硬件与软件包

1.2.1 增强现实系统中的硬件

由于增强现实技术是在虚拟现实技术的基础上发展起来的,因此在硬件结构上同虚拟现实系统的硬件一样具有一定的继承性和一致性。增强现实系统将计算机生成的虚拟物体、文字和图形等信息重叠显示在用户所看到的真实物体上,为用户提供与现实环境有关的辅助信息,增强用户对真实环境的感受,其中关键技术涉及融合显示和跟踪注册。融合显示设备和方位跟踪器是增强现实系统的主要外部

设备。AR 系统中较常采用的融合显示设备是透视式头盔显示器。头盔显示器的性能、价格差别很大,有售价仅数百美元的 LCD 头盔显示器,也有售价上万美元的 CRT 头盔显示器。色彩分辨率和视场角是头盔显示器的两个至关重要的指标。方位跟踪器主要用于确定用户与其周围真实环境间的相对位置和方向,实现虚拟场景与现实环境之间的“对准”问题。与大多数 VR 系统一样,图形处理器也是 AR 系统所必不可少的。此外,AR 系统还包括如数据手套、6D 鼠标器、跟踪器、力反馈装置、语音识别与合成系统等在内的人机交互设备,每种设备品种繁多、性能各异。

尽管 AR 技术在不同应用领域所强调的重点有所不同,不同档次、不同用途的系统所需要的配置也各不相同,但它们在原理上存在很多共性。选配哪些硬件设备往往需要用户在经费投入、性能需求和市场效益等方面折中考虑。如图 1.8 所示,一个典型的 AR 系统由虚拟场景发生器、透视式头盔显示器、实现用户观察视线跟踪的头部方位跟踪设备、虚拟场景与真实场景对准的方位跟踪定位设备和各种交互设备构成。其中虚拟场景发生器负责虚拟场景的建模、管理、绘制和其他外设的管理。头部方位跟踪设备跟踪用户视线变化,实现用户观察坐标系与虚拟场景坐标系的匹配。交互设备用于实现感官信号以及环境控制操作信号的输入和输出。而这里的方位跟踪定位设备负责测量用户在真实环境中的六自由度位置和方位信息,从概念上讲,其与头部方位跟踪设备没有区别,但实际的测量精度要求有很大差别,将直接影响 AR 系统的注册精度。

本书将在第 2 章专门介绍增强融合显示技术,这里不再赘述,仅介绍常用的方

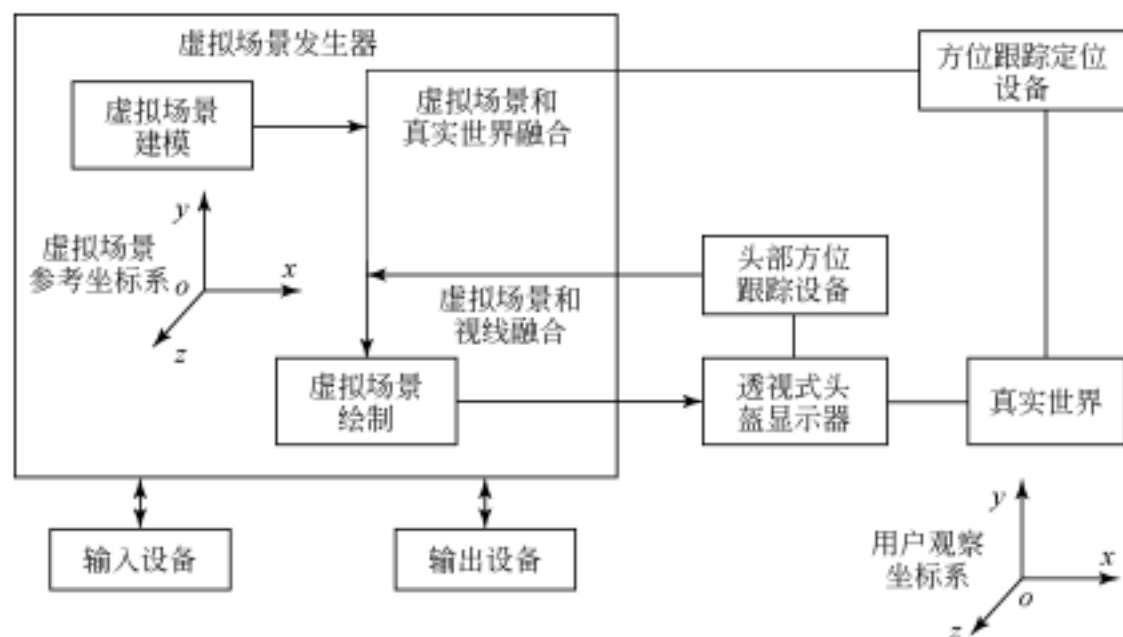


图 1.8 典型的 AR 系统结构图

位跟踪设备。

头盔显示器不仅要具备普通的三维图像显示功能,而且需要能够实时跟踪用户头部的位置和姿态,即跟踪视线方向,并及时将视线信息反馈给图像生成软件,实现视景图像与增强信息的同步刷新。这样可以使用户能够在真实世界中观察周围事物的方式在增强环境中畅游,从而增强用户在融合环境中的沉浸感和交互。获取头部方位信息的设备就是头部方位跟踪器。

从结构的角度看,方位跟踪系统可以分为有源跟踪系统和无源跟踪系统两类。有源跟踪系统包含发射器和接收器,将发射器和接收器两者之一固定在被跟踪的物体(如头、手等)上,而将另一个固定在相对静止的参考环境中,通过发射信号和接收信号之间的某种物理联系跟踪用户的位置和方向。无源跟踪系统没有主动信号源,只利用接收器测量某种被动信号的变化来跟踪其位置和方向的变化。例如,可将摄像机与计算机相连,采用基于计算机视觉的方式,实时计算被跟踪目标的位置和方向信息。基于计算机视觉的方位跟踪方式造价低廉、跟踪定位的精度可达到亚像素级,因此被广泛采用。基于计算机视觉的跟踪定位方式将在第4章、第5章详细介绍,这里不再赘述。

基于硬件跟踪设备获取被跟踪目标位置和方向信息的方式,也常被应用于增强现实系统中。这些硬件跟踪设备包括机电跟踪器、电磁跟踪器、超声波跟踪器、光电跟踪器和惯性跟踪器,它们的实现方法各不相同,各有优缺点,而且在现有的增强现实系统中都有应用实例。

1) 机电跟踪器(probe)

机电跟踪器是一种绝对位置传感器。通常由体积较小的机械臂构成,将一端固定在一个参考机座上,另一端固定在待测对象上。采用电位计或光学编码器作为关节传感器测量关节处的旋转角,再根据所测得的相对旋转角以及连接两个传感器之间的臂长进行动力学计算,获得六自由度方位输出。这种跟踪器性能较可靠,潜在干扰源较少,延迟时间短。但其缺点是,跟踪器测量精度受环境温度变化影响,关节传感器的分辨率低,跟踪器的工作范围受限。在一些特定的应用场合(如外科手术训练),用户的活动范围不是重要指标时这种跟踪器才具有优势。

Fake Space 实验室研制的 BOOM(binocular omni-orientation monitor)显示系统采用机电式跟踪器,其有效角分辨率小于 0.1° ,跟踪频率为 60Hz,工作范围为 $1.5\text{m} \times 0.75\text{m}^{[8]}$,如图 1.9 所示。此外,还包括 NASA 用于远程操作的 FaroArm、Phantom 和 Spidar 等。



图 1.9 Fake Space BOOM 原理示意及效果图^[8]

2) 电磁跟踪器

电磁跟踪器是应用较为广泛的一类方位跟踪器,它利用一个三轴线圈发射低频磁场,用固定在被测对象上的三轴磁接收器作为传感器感应磁场的变化信息,利用发射磁场和感应信号之间的耦合关系确定被跟踪物体的空间方位。根据三轴励磁源的形式不同,电磁跟踪器分为交流电磁跟踪器和直流电磁跟踪器。

交流电磁跟踪器的励磁源由三个磁场方向相互垂直的、交流电流产生的双极磁源构成,磁接收器由三套分别测试三个励磁源的线圈构成。磁接收器感应励磁源的磁场信息,根据从励磁源到磁接收器的电磁能量传递关系计算磁接收器相对于励磁源的空间方位。受计算性能、反应时间和噪声等因素的影响,励磁源的工作频率通常为 $30 \sim 120\text{Hz}$ 。为了保证不同环境条件下的信噪比,通常使用 $7 \sim 14\text{kHz}$ 的载波对激励波进行调制。直流电磁跟踪器的发射器(相当于励磁源)由绕立方体芯子正交缠绕的三组线圈组成,依次向发射器线圈输入直流电流,使每一组发射器线圈分别产生一个脉冲调制的直流电磁场。接收器也是由绕立方体芯子正交缠绕的三组独立线圈构成的,直流磁场方向的周期性变化在三向接收器线圈中产生交变电流,电流强度与本地直流磁场的可分辨分量成正比。可在每个测量周期获得九个数据,它们表示三组接收器线圈所感应发射磁场的大小,由电子单元执行一定的算法即可确定接收器相对于发射器的位置和方向。

交流电磁跟踪系统的接收器通常体积小,适合安装在头盔显示器上,但这种跟踪器最致命的缺点是易受环境电磁干扰。发射器产生的交流磁场对附近的电子导体特别是铁磁性物质非常敏感,交流旋转磁场在铁磁性物质中产生涡流,从而产生二级交流磁场,使得由交流励磁源产生的磁场模式发生畸变,这种畸变会引起严重的测量误差。

直流电磁跟踪器最大的优点是只在测量周期开始时产生涡流,一旦磁场达到

稳态状态,就不再产生涡流。只要在测量前等待涡流衰减就可以避免涡流效应,从而可以减小畸变涡流场产生的测量误差。

Polhemus 公司是交流电磁跟踪系统的主要研制者,该公司的系列产品有 Patriot、Scout 和 Fastrack 等^[9]。以 Fastrack 跟踪器为例,其测量范围为 4 ~ 6ft,测量精度为 0.03in,角度精度为 0.15°,数据刷新频率为 120Hz/接收器数目,延迟为 4ms。采样频率会随着接收器数目的增多而有所下降。图 1.10 所示为该公司的 Patriot、Scout 以及 Fastrack 产品。另外,Ascension Technology 公司也是一家著名的研制电磁跟踪器的公司^[10],该公司于 1991 年开始研制直流电磁跟踪器,目前主要的跟踪器有 trakSTAR、Class B Flock 等。



图 1.10 Polhemus 公司的 Patriot、Scout、Fastrack 电磁跟踪器产品^[9]

3) 超声波跟踪器(acoustic tracker)

利用不同声源的声音到达某一特定地点的时间差、相位差或者声压差可以进行定位与跟踪,一般有脉冲波飞行时间(time-of-flight, TOF)测量法和连续波相位相干测量法两种方式。TOF 测量法是在特定的温度条件下,通过测量声波从发射器到接收器之间的传播时间来确定传播距离的一种方法,大多数超声波跟踪器都采用这种测量方法。此方法的数据刷新率受到几个因素的限制,声波的传输速度约为 340m/s,只有当发射波的波阵面到达传感器时才可以得到有效的测量数据。而且必须允许发射器在产生脉动后发出几毫秒的声脉冲,并且在新的测量开始前等待发射脉冲消失。因为每个发射器-传感器组都需要单独的脉冲飞行序列,测量所需要的时间等于单组飞行时间乘以组合数目。这种飞行时间测量系统的精度取决于检测发射声波到达接收器准确时刻的能力,环境中诸如钥匙叮当响的声音都会影响测量精度,空气流动和传感器闭锁也会导致测量误差产生。

连续波相位相干测量法通过比较参考信号和接收到的发射信号之间的相位来确定发射源和接收器之间的距离。此方法测量精度较高,数据刷新频率高,可通过多次滤波克服环境干扰的影响,而不影响系统的精度、时间响应特性等。

与电磁跟踪器相比,超声波跟踪器最大的优点是不会受到外部磁场和铁磁性物质的影响,测量范围较大。基于声波飞行时间法的跟踪器易受伪声音脉冲的干

扰,在小工作范围内具有较好的精度和时间响应特性。但是随着作用距离的增大,这类跟踪器的数据刷新频率和精度降低。而基于连续波相干测量法的跟踪器具有较高的数据刷新频率,因而有利于改善系统的精度、响应性、测量范围和鲁棒性,且不易受伪脉冲的干扰。不过,上述两种跟踪器都会因为空气流动或者传感器闭锁产生误差。但如果采用适当的调制措施,就可以改善连续波相位测量法的环境特性,有望实现高精度、高数据刷新率和低延迟的声学跟踪器。

1966年,美国 MIT 林肯实验室的 Roberts 研制了一种超声式位移跟踪器 Lincoln Wand,该跟踪器基于声波飞行时间测量法,使用四个发射器和一个接收器,跟踪精度和分辨率只达到 5mm。Logitech 开发了另一种基于 TOF 的超声波跟踪系统,又称为 Red Baron,其跟踪精度和分辨率也只达到几毫米。

4) 光电跟踪器(electro-optical tracker)

光电跟踪器(又称为视觉跟踪器)是利用环境光或者控制光源发出的光,在图像投影平面上的不同时刻或者不同位置的投影,计算出被跟踪对象的方位。在有控制光源的情况下,通常使用红外光,以避免跟踪器对用户的干扰。

从结构方式的角度看,光电跟踪器分为“外-内”(outside-in, OI)和“内-外”(inside-out, IO)两种结构方式。对于“外-内”方式而言,传感器固定,发射器安装在被跟踪对象上,这意味着传感器“向内注视”远处运动的目标,这种系统需要极其昂贵的高分辨率传感器。对于“内-外”方式而言,发射器固定,传感器安装在运动对象上,这意味着传感器从运动目标“向外注视”。在工作范围内使用多个发射器可以提高精度,扩展工作范围。

内-外式光电跟踪器的时间响应特性良好,具有数据刷新频率高,适用范围广,相位滞后小等潜在优势,更适合于实时应用。但光学系统存在虚假光线、表面模糊或者光线遮挡等潜在误差因素,为了获得足够的工作范围而使用短焦镜头,系统测量精度降低。多发射器结构是一种解决方案,却以复杂性和成本为代价。因此,光电跟踪器必须在精度、测量范围和价格等因素之间作出折中选择,而且必须保证光路不被遮挡。

比较著名的“内-外”光电跟踪器是由北卡罗来纳大学(UNC)研制的 IO 型 Hi-Ball 跟踪器^[11,12],如图 1.11 所示,将大约 1000 个红外 LED 安装在边长 4m 的天花板上,三个摄像器件固定在头盔上,通过图像信息处理获取头盔的方位。这种系统根据采用的计算机不同,刷新频率为 25 ~ 200Hz,位相滞后约 5ms。

商业上可用的 OI 光电跟踪器有英国 Vicon 公司开发的使用多个摄像机传感器的 Vicon 系列运动跟踪系统^[13],德国 A. R. T 公司的 AR Track 红外跟踪器系

图 1.11 UNC Hi-Ball 跟踪器^[11]

列^[14], Motion Analysis 公司的 Rhptor 以及便携式 Hawk 系列产品^[15], 加拿大 NDI 数字公司提供的 Optotrak 和 Polaris 系列产品^[16], 如图 1.12 ~ 图 1.15 所示。

图 1.12 Vicon 公司的光电跟踪器^[13]图 1.13 德国 A. R. T 公司的 AR Track 产品^[14]

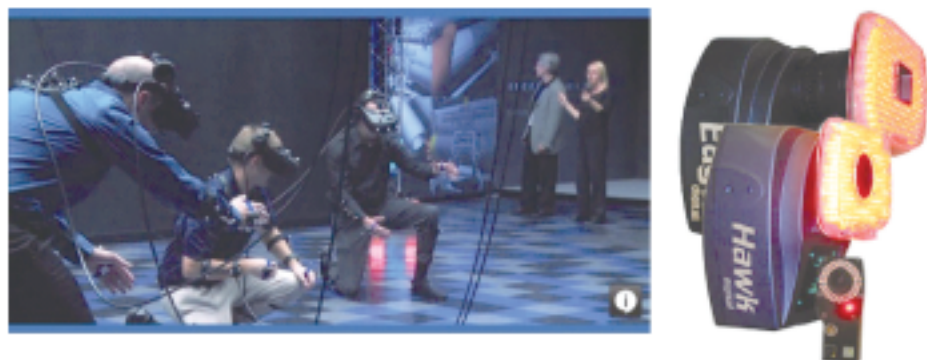


图 1.14 Motion Analysis 公司的光电跟踪器^[15]

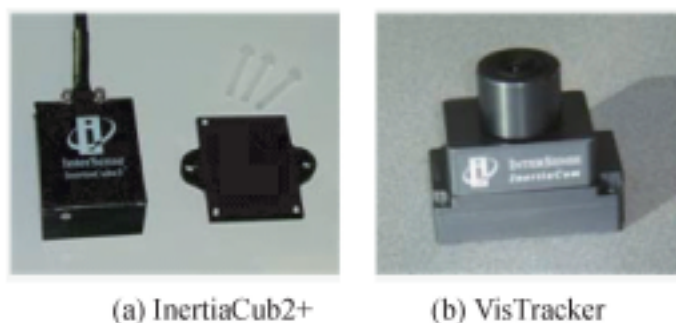
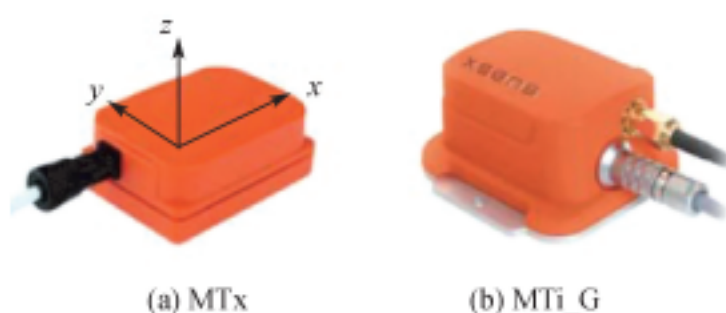


图 1.15 加拿大 NDI 数字公司的 OI 光电跟踪器系列^[16]

5) 惯性跟踪器

惯性跟踪器利用陀螺的方向跟踪能力,测量三个转动自由度的角度变化;利用加速度计测量三个平动自由度的位移。以前这种方位跟踪方法常被用于飞机和导弹等飞行器的导航设备中,比较笨重。随着陀螺和加速度计的微型化,该跟踪方法在民用市场也越来越受到青睐。不需要发射源是惯性式跟踪器最大的优点,然而传统的陀螺技术难以满足测量精度的要求,测量误差易随时间产生角漂移,受温度影响的漂移也比较明显,需要有温度补偿措施。新型压电式固态陀螺在上述性能方面有大幅度改善。

GyroEngine 是美国 Gyration 公司研制的一种采用光纤陀螺的三自由度惯性跟踪器,其最大数据采样频率为 1000Hz,偏移约 $2^{\circ}/\text{min}$ 。1993 年,MIT 的 Foxlin 等设计实现了一种基于惯性导航的方向跟踪器^[17],分辨率达到 0.008° ,精度达到 1° 。1996 年,Foxlin 创建了 InterSense 公司,陆续研制了 3DOF IS-300、6DOF IS-600、惯性-超声波混合的 IS-900、惯性-视觉混合的 IS-1200 VisTracker 等系列产品。图 1.16 为 InterSense 公司的 InertiaCub2+ 和 VisTracker 跟踪传感器^[18]。此外,另一个惯性跟踪器制造商是欧洲的 XSense 公司,其所生产的惯性传感器如图 1.17 所示^[19]。

图 1.16 InterSense 公司的惯性跟踪器产品^[18]图 1.17 XSense 公司的惯性跟踪器产品^[19]

6) 全球卫星定位系统(GPS)

普通 GPS 的定位精度为 30m,通过采用差分技术可以达到 3m,典型 GPS 的刷新率为 1Hz。新型载波相位 GPS 具有厘米级精度,但价格昂贵且刷新速度慢。由于 GPS 通过接收卫星的时间基准信号进行定位,要求导航卫星可视。而在城市、峡谷、山脚或其他复杂地形的情况下,GPS 信号经常无法正常接收,这限制了 GPS 在户外增强现实系统中的应用。

针对有无发射源、便携性、测量精度以及价格等方面,表 1.1 给出了上述跟踪技术的性能对照表。

表 1.1 方位跟踪器技术性能对照表

跟踪器	发射源	便携性	精度	价格	数据刷新 频率	延迟	活动范围	潜在 干扰源
机电	无	差	很高	较高	高	小	很小	无
电磁	有	好	较高	较低	较高	较大	单传感器较小, 多个传感器较大	金属及其 他磁场

续表

跟踪器	发射源	便携性	精度	价格	数据刷新频率	延迟	活动范围	潜在干扰源
超声波	有	较好	TOF 法较低, 相位差法较高	低廉	TOF 法较低, 相位差法较高	小	较小	空气流动和超声波噪声源
光电	有	较好	较高	昂贵	高	小	大	光反射及其他红外光源
惯性	无	可能好	较低	昂贵	很高	未知	未知	随时间和温度漂移
GPS	有	好	较低	昂贵	快	小	大	突变地形

当然,基于硬件设备的跟踪定位也存在以下问题:机电跟踪器活动范围小,便携性差。技术比较成熟的电磁跟踪器的时间延迟大,不能满足实时交互的要求,而且对环境电磁场的敏感限制了这类跟踪器的测量精度和应用范围。性能较好的光电跟踪器价格昂贵。惯性跟踪器需要克服器件测量精度随温度和时间变化的漂移。声波传输速度限制了基于 TOF 方法的超声波跟踪器的数据刷新频率,抗干扰能力低,而且目前尚无基于相位差的超声波跟踪器产品。

1.2.2 增强现实系统软件包

多年来,通过众多研究人员对增强现实技术的不懈努力,目前已有多种可用于 AR 系统开发的工具包和 API(application programming interface)面市。早期最具代表性的软件包为基于计算机视觉技术采用人工标识跟踪定位的 ARToolkit^[20]软件包以及日本混合实境实验室开发的 MR 平台^[21]。ARToolkit 是一套开放源代码的工具包,它主要由日本大阪大学的 Hirokazu 开发,用于快速编写 AR 应用 ARToolkit。受到了华盛顿大学人机界面实验室和新西兰坎特伯雷大学人机界面实验室的支持,已成为在 AR 领域使用最广泛的开发包之一。许多早期的 AR 应用都使用 ARToolkit 或在其基础上改进的版本来进行开发。ARToolkit 基于人工标记跟踪定位,其工具包中包含摄像头校准和人工标记的制作工具。支持 Direct3D、OpenGL 和 VRML 的三维场景格式。同时支持显示器和 S-HMD 等多种显示设备。MR 平台由日本的混合实境实验室开发,其中包含能减少人眼与头盔上摄像机之间平行度误差的 S-HMD 和运行于 Linux 环境下的用 C++语言开发的 SDK,该工具包中提供了摄像机校正工具、视频捕捉、图像检测和控制六自由度传感器等基本功能。

随着跟踪定位技术的不断发展,美国高通公司开发了采用自然特征进行跟踪定位的 AR 软件包 Vuforia^[22]。ARSDK 是由美国高通公司推出的基于移动设备的增强现实软件开发工具包,开发者可以利用该工具包构建应用程序,让用户体验到结合网络与现实空间的移动增强现实的奥妙。目前该版本 SDK 支持 Android 以及 IOS 系统设备,开发人员可以免费下载。当然,目前面世的增强现实商业软件开发包也有很多,例如,Total Immersion 公司推出的 AR 解决方案——D'Fusion^[23],为主题公园、零售业、移动电话运营商等行业提供增强现实解决方案,德国 Metaio 公司开发的移动增强现实软件包 MobileSDK 已更新至 V3.0 版本^[24]。此外,瑞士联邦理工学院 Cvlab 实验室开发的基于自然特征的 AR 实时跟踪定位软件 BAZar^[25],英国牛津大学机器视觉组开发的实时运动结构重建软件 PTAM^[26,27],都非常优秀且极具参考价值,同时也是完全开源的。

1.2.3 增强现实技术的国内外研究机构

增强现实技术的研究起始于 20 世纪 90 年代初,发展至今已经成为集计算机图形学、计算机视觉、图像处理、模式识别、光电技术、多传感器技术等学科的综合性科学技术,其重要研究成果也陆续发表于相关的国际学术期刊和学术会议论文集上。比较有影响力的国际期刊包括 *Computer and Graphics*, *Computer Graphics and Applications*, *Internal Journal of Computer Vision*, *Journal of Real Time Image Processing*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *Presence* 等,有代表性的国际学术会议包括 SIGGRAPH 年度会议、Internal Conference of Computer Vision (ICCV)、Proceedings of Computer Vision and Pattern Recognition (CVPR)、International Conference on Pattern Recognition (ICPR)、Proceedings of Intelligent Robots and Systems、IEEE International Symposium on Mixed and Augmented Reality (ISMAR)、Virtual Reality (VR) Conference、Designing Augmented Reality Environments Workshop 等。2008 年,新加坡南洋理工大学的研究人员通过统计最近十多年发表在 ISMAR 会议上的文章^[28],总结出在跟踪、交互、标定注册、AR 应用和显示技术等五方面的研究文章各占 20.1%、14.7%、14.1%、14.4% 和 11.8%,而在评估测试、便携式 AR、AR 授权、可视化、多模 AR 以及绘制等六方面的研究文章各占 5.8%、6.1%、3.8%、4.8%、2.6% 和 1.8%。这展现了 AR 研究在最近十多年的发展重点和路线,也为从事 AR 的研究人员在选择课题和研究方向上提供了帮助。国际上,许多著名大学和跨国科研机构建立了从事 AR 技术研究的实验室,为 AR 技术的研究和发展做出了巨大贡献,表 1.2 罗列了从事 AR 技术研究的主要

大学和商业研究机构。

表 1.2 从事 AR 技术研究的大学或者商业研究机构

研究机构	国家和地区	研究重点
北卡罗来纳大学	美国	Hi-Ball 光学跟踪器、HMD、医学应用
南加利福尼亚大学	美国	混合跟踪、自然跟踪、室外 AR 系统
麻省理工学院	美国	机器视觉、摄像机标定、可视化
中佛罗里达大学	美国	头盔显示器设计、医学可视化
华盛顿大学	美国	ARToolkit、人机交互
科罗拉多矿业大学	美国	医学增强现实
慕尼黑大学	德国	工业规划设计、医学可视化
日内瓦大学	瑞士	混合跟踪
皇家理工学院	瑞士	无标识跟踪定位
剑桥大学	英国	机器视觉、混合跟踪定位
牛津大学	英国	摄像机自标定、运动结构重建
格拉茨理工大学	奥地利	无标识跟踪定位、混合跟踪
东京大学	日本	ARToolkit、HMD 设计
南澳大利亚大学	澳大利亚	便携式增强现实
新加坡国立大学	新加坡	无标识跟踪、标识点跟踪
InterSense 公司	美国	微型惯性跟踪器、混合跟踪
XSense 公司	荷兰	微型惯性跟踪器、混合跟踪
MATRIS	欧盟	无标识跟踪、视觉-惯性传感器融合
Sony 公司	日本	增强现实

北卡罗来纳大学计算机科学系的跟踪器项目组于 1990 ~ 2001 年研制了大场景跟踪的 Hi-Ball 六自由度光学跟踪器。南加利福尼亚大学(USC)的集成媒体系统中心的项目组从事混合跟踪以及自然特征跟踪技术研究。中佛罗里达大学(UCF)的 ODA (optical diagnostic and applications) 实验室主要从事头盔显示器光学元件的设计研发以及医学可视化研究。MATRIS(markerless real-time tracking for augmented reality image synthesis)是隶属于欧盟的从事 AR 研究的项目组,其研究目标是通过分析摄像机图像和利用惯性传感器,开发可实时跟踪摄像机位置、方向以及摄像机焦距的增强现实系统,主要应用领域包括虚拟演播室、维修、建筑设计、文物遗产保护等^[29]。XSense 公司是总部位于荷兰的从事微型 MEMS 惯性传感器研究的商业公司,主要产品包括 MTi 系列和 MTx 系列的三自由度和六自由度惯性

跟踪器等^[19]。InterSense 公司是总部位于美国的从事运动跟踪器技术研究的专业商业公司,主要研究项目涉及微型惯性跟踪器、视觉-惯性混合跟踪技术,主要产品包括 InertiaCube2、InertiaCube3 系列 3DOF 跟踪器、IS-900 和 IS-1200 系列 6DOF 跟踪器等^[18]。

在国内,“十一五”期间国家在高技术研究发展计划(863 计划)信息技术领域设立了虚拟现实专题,极大地促进了国内虚拟现实和增强现实技术的发展,取得了一批突破性的进展。例如,北京理工大学自主完成了基于光学跟踪的头部位跟踪器和轻便型头盔式显示器,实现了自由曲面的目视光学系统,将光学系统的重量和体积大大减小,在视场(水平 45°)和出瞳直径(8mm)等主要参数上超越了国外同类产品,填补了我国相关领域的空白。该项目组研制了复杂户外环境下的增强现实定点观测系统、随动观测系统以及移动增强现实系统,部分成果已实际用于圆明园现场数字重建,其中定点观察系统效果很好,在“十一五”期间已成功首次在国内进行户外增强现实系统的商业试运营。浙江大学研究了面向大尺度运动结构重建的非连续特征自动跟踪方法,通过提高特征的区分度和降低噪声的敏感性,延长特征点在连续帧上的生命周期,同时降低了特征匹配的复杂度,高效地解决了循环回路序列的高效稳定跟踪和多视频序列的自动匹配问题,极大地提高了场景空间结构重建的精度和规模。此外,针对光照一致性和虚实遮挡问题也进行了系统研究。以北京航空航天大学为主要单位,围绕中国商用飞机有限责任公司支线和干线飞机设计、维护和训练中的实际需求,建立了在一定物理空间支持多人自由参与,具有视、听、触觉多种交互方式和沉浸感的虚实融合协同工作集成环境,在高精度定位装置、虚实融合头盔显示器、触力觉交互装置、虚实融合绘制软件、多人协同网络软件和便携交互装置等方面取得了技术突破。中国科学院自动化所、华中科技大学、上海交通大学等单位也在增强现实领域取得了较大的研究成果。

增强现实技术必将使人类的生活方式和生产方式发生重大改变,发展增强现实技术,创建增强现实产业是高科技领域的又一次重大机遇。

1.3 增强现实技术的难点

从 AR 和 VR 分析对照的结果不难看出,除了与 VR 技术面临共同的难点,诸如真实感虚拟场景绘制、传感设备和交互设备等以外,AR 技术具有挑战性的难点在于:能够将虚拟物体和真实环境精确对准的定位手段;能够将虚拟场景与真实环境融为一体的显示设备。

这里对定位手段的要求不仅是 VR 系统中所要求的方位跟踪系统的精度、数据刷新频率以及延迟,更重要的是理解虚拟环境坐标系、真实环境坐标系以及用户本身视觉坐标系之间的关系,并实现它们之间的精确对准以及对准关系在运动过程中的维持,要求虚拟物体在虚拟环境坐标系中运动的精确性及其相对真实环境坐标系运动的精确性。

对显示设备的要求不再局限于对于虚拟环境的浸没感,也不是虚拟环境与真实环境的简单叠加,关键是虚拟环境深度细节、照明条件、分辨率与真实环境诸因素的匹配,不相匹配的环境会导致感觉的倾向性,甚至产生误导。

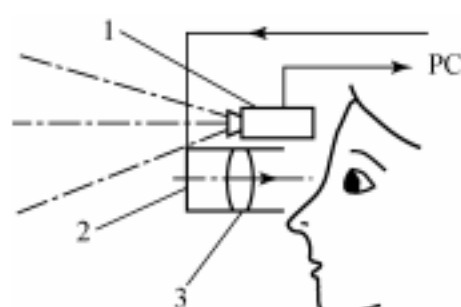
目前阻碍增强现实系统得以广泛应用的技术难点主要表现在以下两方面:一方面是增强现实的显示技术;另一方面是增强现实的跟踪注册技术。以下仅对两种常用的头盔显示设备进行较为详细的对比和论述,投影显示、全息显示以及桌面显示等将后面章节作详细介绍。

1.3.1 增强现实系统中的头盔显示技术

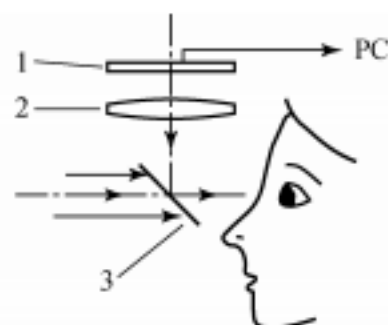
视觉通道是人类与外部环境之间最重要的信息接口,人类从外界所获得的信息有近 80% 是通过眼睛得到的。因此,增强现实系统中的头盔显示技术就显得尤为重要,是整个增强现实系统的关键问题之一。

目前,有许多 AR 系统采用透视式头盔显示器实现虚拟环境与真实环境的融合。透视式头盔显示器由三个基本环节构成:虚拟环境显示通道、真实环境显示通道、图像融合与显示通道。其中虚拟环境的显示原理与浸没式头盔显示器基本相同,图像融合与显示通道是与用户的最终接口,基本取决于真实环境的表现方式。因此,根据真实环境的表现方式划分,目前主要有视频透视式头盔显示器和光学透视式头盔显示器两种。

视频透视式头盔显示器原理如图 1.18 所示,由安装在头盔上的微型 CCD 摄取外部真实环境的图像,计算机通过计算处理将所要添加的信息或图像信号叠加在摄像机的视频信号上,实现计算机生成的虚拟场景与真实场景的融合,最后通过类似于浸没式头盔显示器的显示系统呈现给用户。基于光学原理的透视式头盔显示器原理如图 1.19 所示,真实场景直接通过安装在眼前的一对半反半透镜呈现给用户,经过光学系统放大的虚拟场景经半反半透镜反射而进入人的眼睛,真实场景和虚拟场景的融合通过该光学元件来实现。视频透视式头盔显示器与光学透视式头盔显示器各有优缺点。视频透视式头盔显示器的优点体现在以下几方面。



1-摄像机；2-LCD；3-中继投机组



1-LCD；2-中继透镜组；3-光学组合器

图 1.18 视频透视式头盔显示器原理 图 1.19 光学透视式头盔显示器原理

(1) 合成景象的灵活性。光学透视式头盔显示器的一个主要问题是计算机生成的虚拟物体不能够完全遮挡真实环境中的物体,这是由于光学透视式头盔显示器前方的光学透镜组允许来自真实和虚拟环境中的光线同时通过。这一问题的存在,使得注册的虚拟物体呈现半透明状态,从而破坏了真实景象与虚拟场景相融合的真实感以及由于遮挡而提供人们的深度信息。对于视频透视式头盔显示器而言,却不存在以上问题。因为无论真实或虚拟环境中的物体均是由计算机处理后在头盔显示器的液晶屏上显示输出。由于真实场景的分辨率降低到了摄像机能够提供的水平,可以适当降低对于注册精度的要求,而且呈现给用户的增强场景完全受系统控制,真实场景和虚拟场景的匹配很容易通过电子融合方式实现,来自系统其他部分的潜在延迟也可以得到补偿,可以提供更强烈的浸没感,也可以同时利用摄像机实现头部方位跟踪。

(2) 较宽的视野。光学系统通常具有畸变,要开发视野较宽的 AR 系统,就必须考虑光学系统的畸变问题。视频透视式头盔显示器可以通过图像处理的方法校正光学系统存在的畸变。

(3) 延时的匹配。对于光学透视式头盔显示器,用户可以实时地看到真实世界的景象,而虚拟场景往往要经过一系列的延时才能够添加在真实场景之上,容易造成系统的动态注册误差,使增强信息在真实场景中产生游移现象。而采用视频透视式头盔显示器可以将拍摄到真实场景消耗的系统时间与描绘虚拟场景的系统延时进行匹配,以减小动态注册误差,提高注册精度。

(4) 光强度的匹配。自然界光照强度的变化范围跨度大,显示器光强度的变化范围无法与之匹敌。因此,光学透视式头盔显示器存在真实与虚拟场景光强度难以达到最佳匹配的缺点。视频透视式头盔显示器却不存在这方面的问题,因为无论真实还是虚拟场景均在头盔显示器上显示输出。

光学透视式头盔显示器的优点体现在以下四方面。

(1) 结构简单,价格便宜。光学透视式 AR 系统在工作过程中只需考虑由计算机生成的虚拟物体,周围真实世界景象可由用户直接透过安置在眼前的光学器件感知。而视频透视式头盔显示器需要分别处理摄像机拍摄到的真实场景图像以及由计算机生成的增强信息,这一过程一般存在几十毫秒的时间滞后。因而光学透视式头盔显示器在接口电路以及显示器件方面结构均较简单,对计算机的图形运算速度要求相对较低,价格也较便宜。

(2) 分辨率较高。视觉科学的研究结果表明,人眼在视网膜中心的分辨率高达 1 弧分,这是 AR 系统特别是基于光学透视式头盔显示器对于注册精度要求的极限。而视频透视式头盔显示器将人眼的分辨率降低至显示器的分辨率范围。

(3) 安全性高。视频透视式头盔显示器是一种封闭式的显示设备,一旦断电用户无法看到外界的景象,在某些需要考虑安全性的特殊领域,光学透视式头盔显示器更具优势。

(4) 不存在视点的偏差。对于视频透视式头盔显示器,为了使拍摄的外部世界景象和用户感受相重合,摄像机应尽可能与人眼视点重合。但是由于摄像机与用户的视点难以保持完全的重合,用户看到的视频景象与真实景象存在偏差。

综上所述,光学透视式头盔显示器与视频透视式头盔显示器各有利弊,究竟采用何种形式的头盔显示器,应根据所开发 AR 系统的应用领域确定。

1.3.2 增强现实系统中的跟踪注册技术

增强现实系统中,跟踪注册是增强现实系统构建的基础技术,也是决定增强现实系统性能的关键。跟踪注册技术一直以来是增强现实系统研究的重点,也是主要的技术难点。跟踪注册精度是衡量 AR 系统性能、影响其实用性的关键指标。AR 系统中的注册误差可分成两大类:静态注册误差和动态注册误差。当用户的视点与真实环境中的物体均保持静止时,系统产生的误差称为静态注册误差。而只有当用户的视点或环境中的物体发生相对运动时才会出现的误差称为动态注册误差。动态注册误差是造成增强现实系统注册误差的主要来源,也是限制增强现实系统实现广泛应用的主要因素。

增强现实系统的静态注册误差源主要有以下几方面:光学系统畸变、图像噪声、机械未校准、非正确的系统参数、算法本身以及摄像机标定时引入的误差。大多数摄像机光学透镜均存在畸变,使用光学透视式头盔显示器时,使用者可以看到一个未发生畸变的真实世界环境。然而由于摄像机和计算机显示器的非线性畸

变,所添加的增强图像产生畸变,引入了静态注册误差。图像噪声主要为系统和传感器的固有噪声。机械未校准指摄像机光轴与 CCD 像平面不垂直、镜头中各个镜片的偏心、倾斜误差;头盔的机械结构设计不够紧凑,当头部运动时各元器件的位置发生微小改变;这些都可能引入静态注册误差。非正确的系统参数是指瞳距、视场中心的位置误差, X 轴与 Y 轴的尺度因子不相等,这些也是造成静态注册误差的原因。William 就静态误差造成目标三维空间的位置偏差进行了详细的研究,并给出了误差的三维空间分布图形^[30]。

系统时延是造成动态注册误差的主要原因,它包括用户头部位置和方向的跟踪计算、图像数据的传输与 A/D 转换、图形引擎产生虚拟增强信息以及扫描输出至头盔显示器正确位置所需的时间。对于光学透视式头盔显示器而言,用户可以实时地看到周围真实环境中的景象。而对真实景象进行增强的虚拟物体或信息要经过以上一系列的系统延时后才能显示到头盔显示器上。当用户的头部或周围景象物体未发生运动时,系统的延时不会造成动态注册误差。但是当用户的头部或周围景象物体发生变化时,系统延时会使增强的信息或物体在真实环境中发生游移现象,造成动态注册误差。目前国外先进的增强现实系统,其时间延时一般在 25ms 以下。对于一个系统延时为 100ms 的增强现实系统,若用户头部的平均运动速率为 $50^{\circ}/s$,则系统动态注册误差为 5° ,而人眼的角分辨率为 $1'$ 。若摄像机放置在离真实世界中的物体大约 68cm 的地方,那么真实与虚拟场景间的注册误差高达 60mm,显然无法满足增强现实高精度注册的要求。

目前,增强现实系统中主要采用以下几种跟踪注册方法。

(1) 基于跟踪器跟踪用户头部的方位,主要采用的跟踪器包括电磁跟踪器、惯性跟踪器、测距仪、超声波定位仪、GPS 等,用以探测和跟踪真实环境中目标的位置和方向。所有这些跟踪器各有优缺点。为了弥补各跟踪器的缺点,许多研究者采用混合跟踪的方法以取长补短,满足增强现实系统高精度注册的要求。

(2) 基于视觉的跟踪注册。目前 AR 领域基于计算机视觉的注册技术研究处于主导地位,它是通过对一幅或多幅视频图像进行处理,获得跟踪信息,判断固定在使用者头部的摄像机在真实环境中的位置和方向。仅根据头部跟踪系统提供的信息进行三维注册属于开循环控制,系统没有反馈增强信息与真实环境间的匹配程度,因而难以取得最佳匹配效果。而基于视觉的 AR 系统可利用图像处理和计算机视觉的方法协助注册,因此能够使注册误差局限在像素级(以像素为单位,而不是以世界空间坐标系中的米、英寸为单位)。

(3) 混合跟踪注册。目前单一的跟踪技术无法全面地解决增强现实应用系统

中的方位跟踪问题。为此,国外著名大学和研究机构的研究人员提出采用混合跟踪(hybrid tracking)的方法对增强现实系统进行跟踪注册。所谓混合跟踪是指采用不同种类的跟踪设备,取长补短共同完成增强现实系统的注册任务。目前常采用的硬件跟踪设备包括机电跟踪器、电磁跟踪器、光电跟踪器、惯性跟踪器、超声波跟踪器、陀螺仪、GPS 等。

现有的跟踪注册方法各有优缺点和适用性。一般需要使用头部跟踪器的增强现实系统,系统跟踪较复杂,并且会因此引入系统固有误差。基于计算机视觉的跟踪注册是一种综合性能比较好的注册方法,其通用性较强,仅通过图像分析的方式就可以得到头盔显示器的方位信息,系统构造简单,减少了系统误差(不存在摄像机和跟踪器之间的误差),尤其适用于使用光学透视式头盔显示器的增强现实系统,但是计算量较大,系统的延迟性问题较严重。

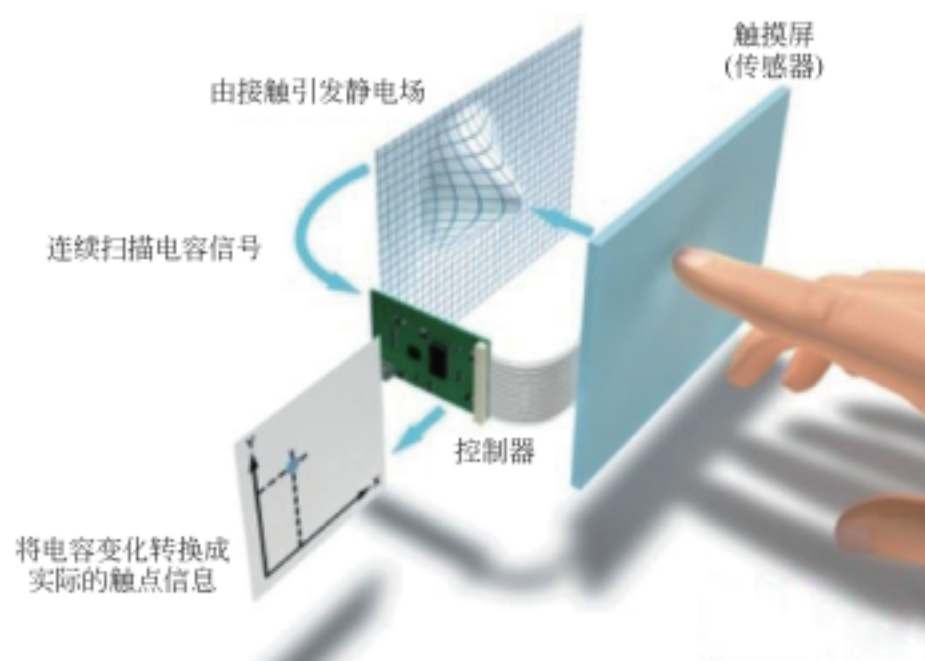
跟踪注册技术是实现增强现实技术的基础,是表征增强现实系统性能的重要指标。一直以来都是增强现实研究领域的热点。国外众多著名学府和研究机构投入了大量的人力和物力进行研发,致力于研制与应用系统相配套的注册算法。增强现实的跟踪定位算法将在后续章节进行详细介绍。

1.3.3 增强现实系统的交互技术

交互技术是增强现实系统中与显示技术和注册技术密切相关的技术,满足了人们在虚拟和现实世界自然交互的愿望。随着人机交互技术的发展,用户可以摆脱传统的交互方式,如键盘和鼠标,实现更加自然和人性化的交互方式。如图 1.20 所示的 iPhone 多点触摸交互技术,可以精确识别多种手指运动方式并获取手指触摸位置坐标,达到人手动作手势的自然交互。增强现实技术本身作为一种新型的人机交互接口,基于动态手势识别的人机交互可为增强现实系统带来更广泛的应用。因此,交互技术具有很高的科学研究价值。

在早期的 AR 研究中,研究重点主要集中于跟踪、注册和显示,只需简单地将虚拟物体叠加在真实场景内,并通过显示设备观看虚实效果,没有太多与外界的交互。但是随着计算机性能的提高,显示设备的微型化、便携化,仅“显示”的增强场景不再能满足用户的需求,从而促使多种交互技术在 AR 系统中应用,如利用语音识别技术、手势和人体姿态识别技术等。

交互技术是一些有共同特征的交互任务的抽象表达方式,研究人机交互过程中的共性,实现在不同环境下的交互。其研究目的是达到人与机器交互的自然和高效,例如,传统的鼠标采用单击方式对该交互进行抽象。新型的手势识别技术可

图 1.20 iPhone 多点触摸交互技术^[31]

以用特定的手势来实现单击功能。如图 1.21 所示,医生只需挥动手就可以翻看患者的 X 射线照片,而不需要用传统的遥控器上下翻阅,这为医生在手术过程中对病情资料的查看带来巨大方便。如图 1.22 所示,用户只需舞动手,就可以利用微软公司研发的 Kinect 控制虚拟场景中的人物打乒乓球。

图 1.21 利用手势识别进行翻阅观看^[32]图 1.22 微软 Kinect 手势识别乒乓球游戏^[33]

增强现实系统中的交互方式大体可分为以下几类。

(1) 基本命令式交互:通过交互完成选择、漫游、旋转、操控等功能。该技术通过指定动作状态对应指定操作,如选择、移动等。可以通过获取人手的空间位置特征信息,触发相应命令。

(2) 双手交互:该技术在认知心理学上逐渐受到关注,该交互方式能够给人带来直接、高效的交互体验。双手交互的研究主要集中于双手操作的行为学原理、心理学基础和双手交互在人机交互中的应用。图 1.23 是基于增强现实的汽车维修,

用户可以通过观看虚拟动画提示进行正确的操作维修和学习。

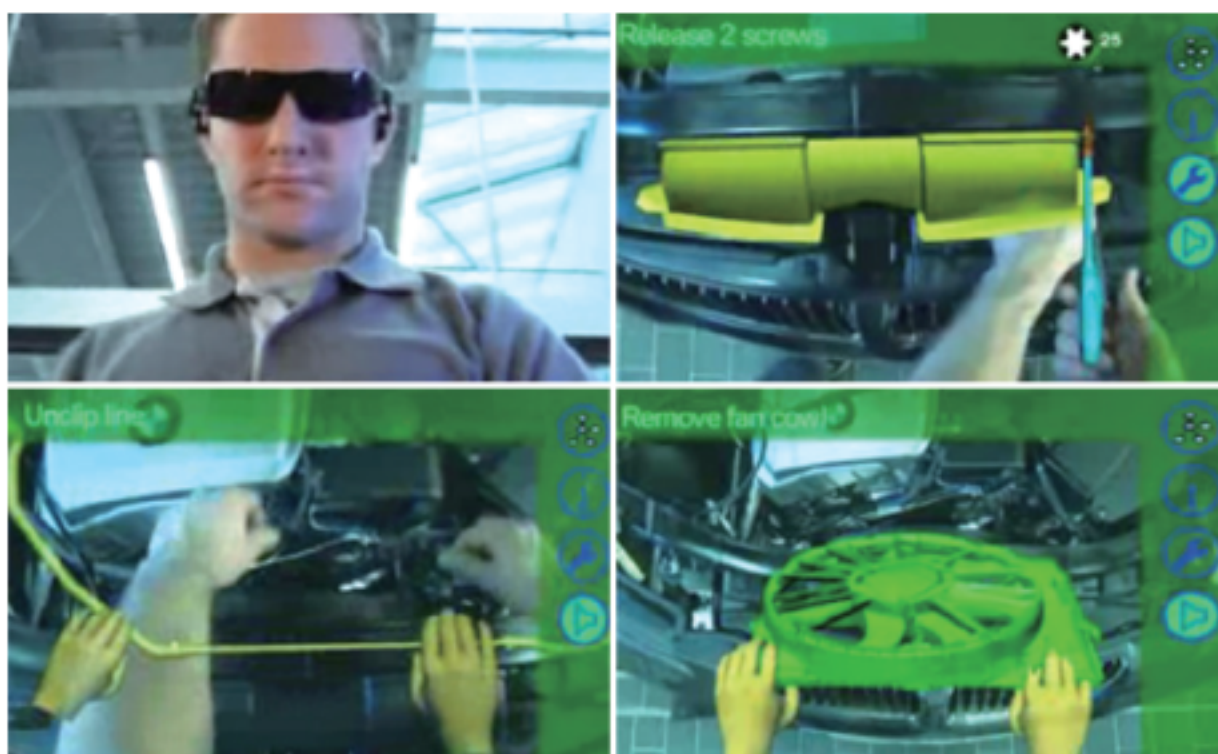


图 1.23 基于增强现实的汽车维修^[34]

(3) 多通道交互:人自身具有多种感官感知功能,虚拟环境可以为用户提供真实的高浸没的感官体验。通过手势、身体姿态、语音甚至对眼睛视点捕捉都可以作为增强现实系统中的交互方式。此外,还可以将触觉、嗅觉、听觉、力反馈等作为输出,从而实现多通道的增强现实交互。其研究的关键技术是如何实现多种交互方式与用户意图的结合。图 1.24 所示为 Tempest 在 TED 大会上分享的基于增强现实技术的投影追踪和绘图系统,该系统包括手势跟踪、脸部追踪以及基于 Kinect 深



图 1.24 基于增强现实的虚拟魔术交互表演^[35]

度图像控制的“global magic dust”^[35]。

(4) 特殊工具的交互方式: 可以使用简单的、易于识别的工具或标识点作为交互设备, 通过识别不同工具或动作的命令, 实现不一样的增强现实交互体验。如图 1.25 所示为 Lee 使用简单自制的红外笔, 外加投影跟踪系统, 就可以实现简单的电子白板, 实现只要花费 40 美元, 就可以等同拥有上千美元的数位白板、多点触控大屏幕和头戴式 3D 浏览器。

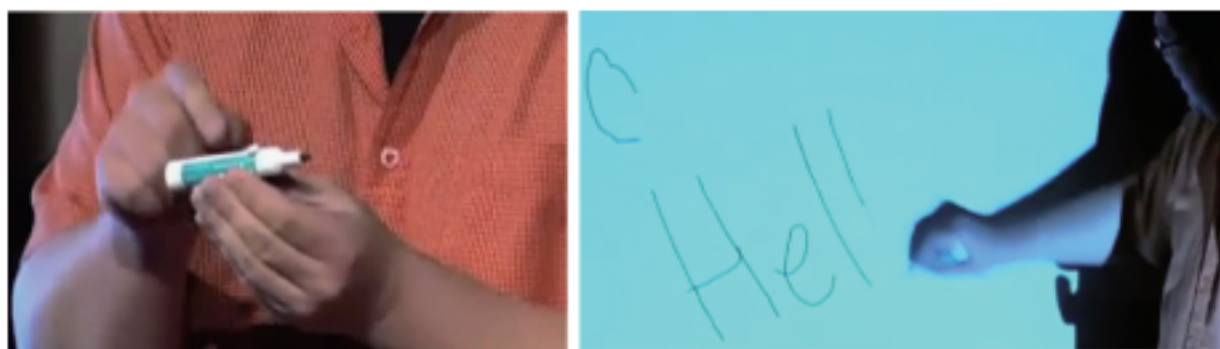


图 1.25 基于红外跟踪的投影交互电子书写板^[36]

1.4 增强现实技术的应用领域

AR 技术不仅继承了 VR 技术的特点, 而且具有虚实结合的独特性。因此, AR 技术不仅具有与 VR 技术类似的应用领域, 诸如尖端武器、飞行器的研制开发、战场虚拟训练、太空探险、样机设计、制造与装配、教育培训等, 而且更显著地突出其对于真实环境的增强优势, 在医疗研究、解剖训练、精密仪器制造、装配与维修、军事训练、工程设计和远程机器人等领域具有比 VR 技术更加明显的优势。

得益于国内外科研机构对增强现实技术的研究贡献, 增强现实系统中所必需的头部跟踪技术逐步从单一传感器向高精度、鲁棒的混合跟踪发展, 显示器技术从透视式头盔显示器向投影式显示器、轻巧便携的可穿戴眼镜等方向发展。这些发展和进步已经促使增强现实技术在工业、医学、军事、文物古迹保护、建筑、游戏等多领域得到广泛应用。

1) 医疗研究

各种成像技术在医学领域的盛行, 为 AR 技术在医学领域的应用奠定了良好的基础。目前比较成熟的 CT 技术、MRI 扫描、X 射线透视以及 B 超技术为人体内部解剖结构可视化提供了大量图像数据。因此, AR 技术在解剖训练、教学、诊疗和图像引导手术等医疗研究方面的应用几乎是顺理成章的。AR 技术能够将患者的

MRI 或 CT 磁共振图像叠加在患者相应的身体部位,使医科学生能够更加直观、具体地了解患者的病情,赋予医生“透视功能”。此外,AR 在协助手术过程可视化应用方面也有相当出色的表现。例如,在做脑部手术的过程中,AR 系统为医生提供患者头部需要开刀的精确手术位置,帮助医生准确定位手术部位。也就是说,可利用 AR 系统提供的精确文字、图像帮助医生一步步地完成手术全过程。早在 20 世纪 90 年代,科罗拉多矿业大学研究小组就利用超声波传感器扫描孕子宫,得到胎儿的三维图像,并将其显示在医生佩戴的头盔显示器上,从而使医生看到正在运动的胎儿图像(图 1.26)^[37]。MIT 多媒体实验室也在 20 世纪 90 年代初期就开始从事 MRT 和 CT 可视化显示方面的工作^[38],并在基于 AR 技术的医学治疗方面有出色的表现(图 1.27)。美国华盛顿大学开发的 AR 系统能够帮助帕金森综合征患者逐步学习、恢复走路的功能^[39]。德国慕尼黑大学开发的 AR 系统可实现 CT 体数据与病患病灶部位的实时配准,系统采用视频透视式头盔显示器与光学跟踪系统对增强结构进行可视化(图 1.28)^[40]。此外,在利用增强现实技术辅助心脏手术导航以及增强现实辅助脊椎成形手术方面也作了大量的应用研究。



图 1.26 孕子宫中的虚拟胎儿^[37]

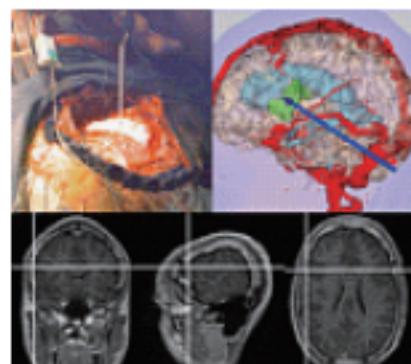


图 1.27 AR 用于医疗^[38]



图 1.28 德国慕尼黑大学开发的 AR 系统的融合显示效果^[40]

2) 军事应用

多年来,军用飞机和直升机利用平视显示器或透视式头盔显示器将矢量图形和文字符号叠加到飞行员的视野中,提供导航和瞄准信息。AR 系统的透视式头盔显示器可以让战士在看到战场实际景象的同时,显示红外或夜视图像;也可以让指挥部用远程通信的方式传来的各种侦察信息、指挥命令、对特殊情况的帮助(如战场医疗急救、排雷指导)等。

美军提出未来城市作战的关键在于能提供单兵级的环境位置及协同信息,战场增强现实系统(BARS)的提出满足了这一需求,该项目包括穿戴式增强现实系统和三维交互命令环境。图 1.29 为美国海军研究中心研制的 BARS 原理样机模型框图,它已实现了指挥中心与各个战士之间的战略和战术信息传输,可作为多人使用的分布式增强现实系统。配备有该平台的战场指挥员和士兵可以通过透视式头盔显示器在观察战场实际景物的同时,看到叠加在其上的各种信息^[41]。

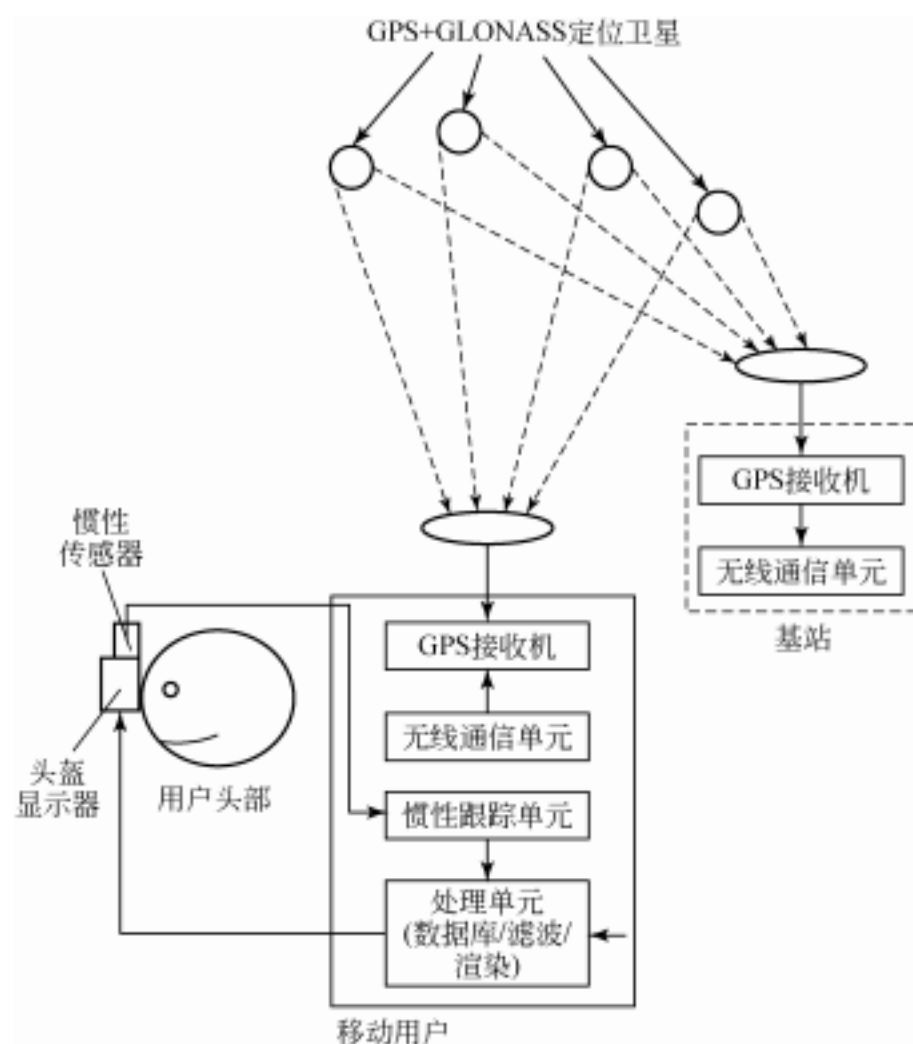


图 1.29 BARS 的原理样机模型框图^[41]

利用 AR 技术也可以为军事训练提供比实兵演习更加真实的战场环境。例如,士兵在训练时通过随身携带的 AR 系统,不仅可以看到真实的场景,而且可以观察到场景中各种虚拟添加的人或物,使军事训练更加实战化,如图 1.30 所示。

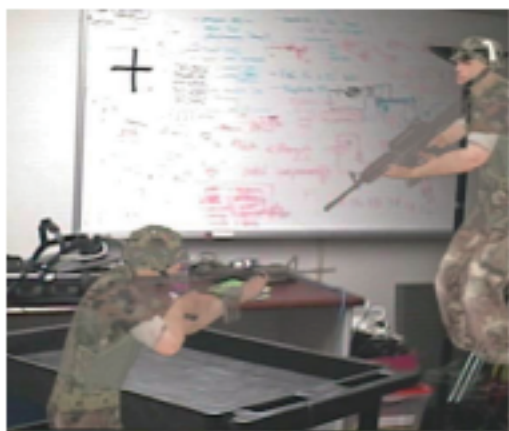


图 1.30 基于 AR 技术的士兵军事训练^[42]

利用增强现实技术不同地域、不同单位的设计人员可通力配合为军方完成复杂装备的研制,提升装备研制的效率以及装备的实用性。大量复杂装备的维修一直是部队面临的棘手问题,而战场上的装备快速维修更是决定战争胜负的要素之一。传统的装备说明书一般是文本和图形形式的,不方便技术人员参考阅读。若将增强现实技术用于装备维修,则可以直接在实际设备中添加 3D 画面,一步一步地提示技术人员应该做什么以及如何做,从而方便装备的维修,极大地提高装备保障的效率^[43],如图 1.31 所示。



图 1.31 武器装备维修与组装^[43]

20 世纪 90 年代以来,信息化武器装备已经成为提高现代军队作战能力的关键。增强现实技术就是对人类信息功能的全面拓展,从大型机械设备维护到成像

系统,乃至城市交通辅助和军事侦察应用,增强现实系统都能大显神通。

3) 装配与维修

在复杂机械、仪器的组装维护和修理方面,增强现实系统同样有着出色的表现。装配一台复杂而陌生的精密仪器时,结合 CAD/CAM/CAE 技术,AR 系统将装配流程指南按照工作进度准确地显示给用户,指导用户顺利地完成任务。对于用户而言这些附加的文字、图像较之厚厚的安装或维修手册更加生动且易于理解,如图 1.32 所示^[44]。早在 20 世纪 90 年代初,美国波音公司的计算机服务研究和技术组就将增强现实技术应用于飞机制造中的电力线缆的连接和接线器的装配中。在 AR 系统的帮助下,一个未受过训练的工人亦可胜任此项工作,从而缩短了员工培训的周期(图 1.33)^[45]。Reiners 研究组将 AR 系统用于汽车门锁的安装,该系统能够一步步地指导毫无经验的用户完成复杂汽车门锁的安装任务(图 1.34)。在安装过程中用户还可通过语音识别系统与计算机进行交互式的对话^[46],该样机在 1998 年的工业交易会上一经展出就受到了众多媒体的关注。

美国哥伦比亚大学计算机科学系的图形与用户接口实验室(Graphics and User Interfaces Lab in the Computer Science Department)和建筑学研究生院的建筑技术工作组(Building Technologies Group in the Graduate School of Architecture)合作的一个项目也颇有成效^[47]。在 1996 年的美国民用工程师学会的第三次民用工程学中的计算机应用会议(American Society of Civil Engineers Third Congress for Computing in Civil Engineering)中,他们展示了将增强现实技术应用于建筑的一个示例,通过叠加在真实环境上的信息,指导工作者从一堆外形和尺寸相似的钢件中挑选出合适的构件(图 1.35),并且正确组装整个构架。



图 1.32 添加增强信息在几何物体上^[44]

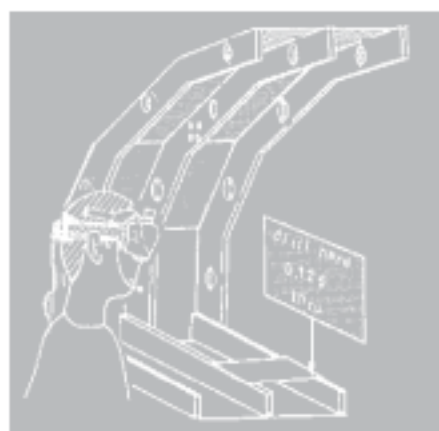


图 1.33 波音公司的 AR 系统^[45]



图 1.34 汽车门锁安装^[46]



图 1.35 使用者挑选构件^[47]

此外,2003 年德国的 Starmate 系统^[48]和同期的 Arvika 系统^[49]的成功研制,再



图 1.36 利用 Starmate 系统进行装配^[48]

次展示了 AR 技术在复杂机电系统维修、装配领域的巨大应用潜力。Starmate 项目由欧洲共同体资助,由六个公司和一个德国研究所共同实现。Starmate 系统主要有两个功能,即帮助使用者实现组装和维修程序以及培训使用者(图 1.36)。Arvika 项目由德国教育研究部资助,鼓励发展 AR 技术在工业中的应用的项目。该项目主要是针对飞机、汽车的增强现实装配与维修。欧洲航空防务与航天公司(EADS)利用 Arvika 系统对欧洲某型战斗机布线问题的解决成为

其经典案例(图 1.37)。装配工人可以通过语音调用计算机生成的提示信息,轻松地完成 1m×6m 板上的高密度布线工作。空客公司利用 Arvika 系统进行了“干净水系统的转配”和“配电盘连接”的装配工作。



图 1.37 利用 Arvika 系统进行装配^[49]

4) 教育培训

在教育培训方面,AR 系统可用来对物体或环境进行注释说明,提供公共或私人信息。早在 20 世纪 90 年代初,欧洲计算机工业研究中心(ECRC)将 AR 应用于机械模型的注解。当 AR 系统用户指向该模型的某一部位时,用户将看到由计算机生成的叠加在该部位的名称信息图(图 1.38)^[50]。与建筑物结构信息的数据库配合,AR 可以显示建筑物内部的结构信息,如墙内管道、电线和支撑结构的位置。多伦多大学的研究人员开发了一套增强现实系统,该系统通过在立体视频图像上添加虚拟图形来实现增强效果。图 1.39 显示的是航天飞机上增强显示的框架线,这些框架线使得观察航天飞机的几何结构变得更加直观、容易理解。与此类似,虚拟增强图形图像也能够用于在观察条件非常差的情况下(如水下、大雾中),辅助导航或加强对场景的理解^[51]。

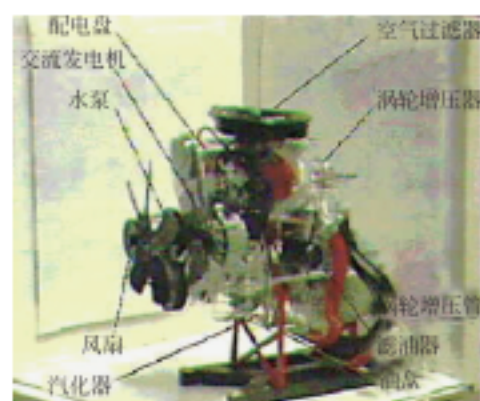


图 1.38 机械模型的注解^[50]



图 1.39 航天飞机的内部结构^[51]

Kaufmann 对 AR 应用于教学进行了多年的研究,开发了名为 Construct 3D 的增强现实工具进行空间几何教学^[52],如图 1.40 所示。该 AR 工具软件不仅能够为学生提供一种面对面的沟通与合作平台,而且通过 AR 可以增加学生对复杂空间中概念及其相对关系的理解。Kaufmann 通过 3 次分析和评估收集了 100 多名学生的 500 多个教学单元资料。结果显示,应用 Construct 3D 工具教学可以提高学习效率,节约大量的学习时间,同时能使學生保持较高的学习热情,而且这种模拟情境的学习效果具有一致性。

5) 机器人远程操作

机器人远程操作是一项非常艰巨的任务,特别是在机器人位于遥远的地方的情况下,信息处理的时间滞后较为严重。传统的解决方法是利用基于虚拟现实的预测图形仿真技术来消除其影响,即建立虚拟的仿真环境,实时地将操作结果反馈给操作者。而远端的机器人在一定的时延后重复仿真结果(图 1.41),使操作者面

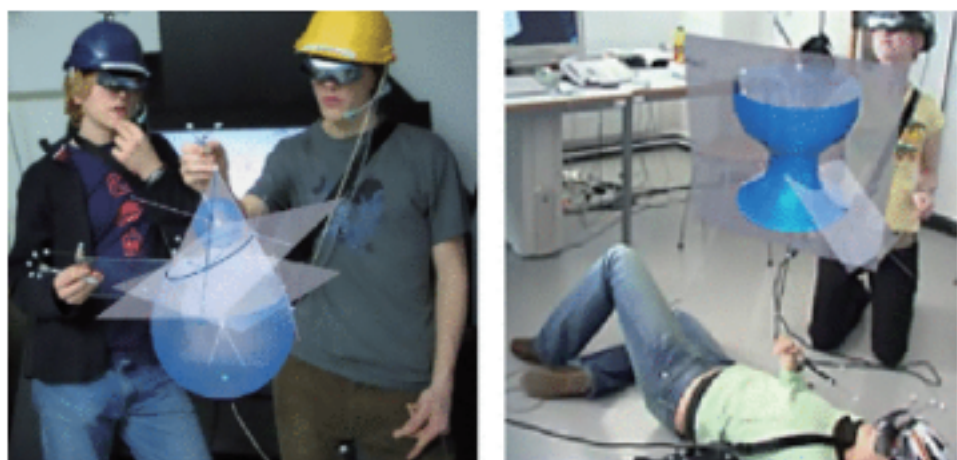


图 1.40 学生利用 Construct 3D 进行几何空间的学习^[52]

对虚拟仿真模型进行连续的实时操作,有效克服时延的影响。然而该方法也会带来一些问题,因系统不可避免地存在建模误差与运动累积误差,会导致仿真环境中的虚拟机器人与实际机器人的运动存在偏差,而这些误差的存在不但降低了预测的效果和精度,还可能導致在实际操作中出现误操作。

增强现实技术具有的虚实结合的特点,可以满足遥操作系统的需求,弥补预测图形仿真系统的不足。工程师可以将远端机器人的真实情况视频传送到本地操作端,与仿真模型在同一窗口中叠加显示,并对虚拟机器人和真实机器人的运动情况进行对比分析,找出它们之间的差异,在必要的时候对仿真模型进行及时修正,或对远端机器人发出修正命令,使得它们的运动相吻合。



图 1.41 机器人远程操作^[53]

6) 娱乐

娱乐也是 AR 的主要应用方向之一。在手机端运行的 AR 娱乐应用近年来大

幅增加,从屏幕效果到用户体验都带给人耳目一新的感觉。iButterfly 是一个有趣的基于 AR 技术的捉蝴蝶应用^[54]。运用运动传感器和 GPS 功能,利用手中的 iPhone 捕捉虚拟蝴蝶。用户可以在不同地域捕捉到不同品种的虚拟蝴蝶,捕捉蝴蝶的同时还能意外获得折扣优惠券(图 1.42)。

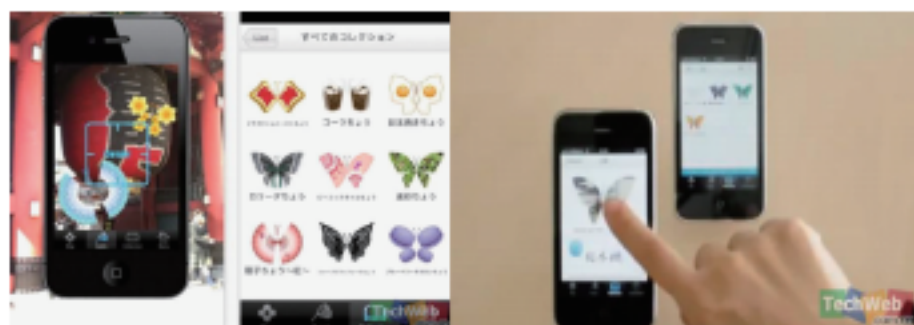


图 1.42 iButterfly 应用界面和用户使用过程^[54]

Sky Chart 和 Star Walk 吸引了众多天文爱好者,这两个令人惊叹的应用可以让用户身临其境地观看星星。对于生活在繁华大都市的人来说,这是很难做到的。只要启动应用,关掉灯,用户就可以抬头观看各种星座了(图 1.43)。这两个应用都提供了室内和室外模式的夜视选择,其中 Star Walk 的信息特别丰富^[55]。

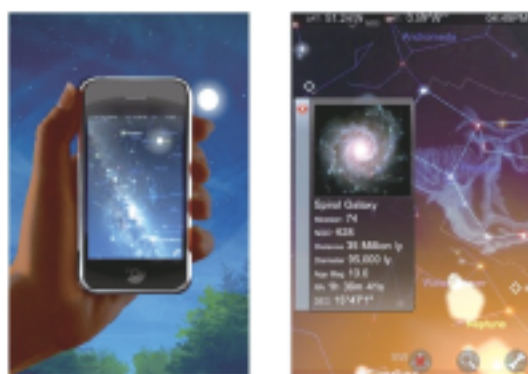


图 1.43 Sky Chart 和 Star Walk^[55]

微软研究部门研发了一款增强现实投影器,该科技利用多达四台的 Kinect 让用户与房间内的物品进行交互。简单地说,用户可以用手与投影仪投影出的影像进行交互。房间内的 Kinect 传感器可以将整个房间转换为一个数字化的空间,然后就可以用手与该投影仪投影的虚拟物体进行互动。用户可以在墙上画画,或是拖动投影仪投射的虚拟物件,还可以克隆真实的物件并投射,如图 1.44 所示^[56]。

7) 商业应用

在商业上,近年来,增强现实技术已被应用于实时视频转播领域,主要的应用实例有虚拟广告、比赛项目的文字图像注释。图 1.45 显示的是 2008 年奥运会游



图 1.44 微软的增强现实投影器^[56]

泳比赛实况转播过程中,实时添加的虚拟注释文字和图像,以便观众能够实时地了解选手的国别和当前的比赛状况^[57]。2010 年 CCTV5“豪门盛宴”在解读 2010 年世界杯的时候,节目现场直播中利用了增强现实技术,将南非世界杯各个球场的 3D 模型显示在演播大厅的舞台中央,而且 3D 模型可以随着镜头的视角和纵深的变化而发生相应的变化,着实令观众兴奋了一次(图 1.46)^[58]。增强现实技术无疑为赛事的转播增添了几许科技气息,在人们讨论世界杯球赛的同时,又多了一个谈论的话题。



图 1.45 实况转播中的虚拟注释^[57]



图 1.46 CCTV5 转播中的增强现实技术^[58]

在 TED2010 上,微软展示了全新的结合增强现实体验的 Bing 街景地图技术(图 1.47),震撼全场。Bing 街景地图整合了全景图、Photosynth 等技术,并结合地理信息标签实现了影像和地图街景的动态匹配^[59]。

8) 文化遗址保护

利用增强现实技术实现数字化古迹遗址保护技术的核心是实现遗产的原地保护以及遗产所在地环境风貌的整体保护,通过虚拟与真实场景的无缝融合,将数字化的文化遗产融合到当前真实的自然与社会环境中。该技术跨越时间和空间,将

图 1.47 Bing 街景地图^[59]

数字化的虚拟场景模型通过多种方式显示于真实的背景环境中,体现了城镇风貌和自然景观变化下的文化遗产保护理念,目前这项新技术已引起学术界的普遍关注。

欧洲、美国的一些著名研究机构在这方面进行了大量的探索和研究。例如,美国 IBM 公司与埃及多家单位合作开发的“永恒的埃及”项目,该项目利用虚拟现实、增强现实以及无线网络技术实现了埃及金字塔的数字化参观导览^[60]。2001 年希腊多家研究机构和政府部门联合成立的研究小组设计开发了可穿戴式增强现实系统 Archeoguide^[61],用于对希腊古遗址(如奥林匹亚神庙)的数字增强再现[图 1.48 (a)]。通过该系统游客可以看到希腊奥林匹亚神庙在 2000 多年前的情况,从而使游客在游览文物古迹时可以更好地理解这些古迹及发生在其中的重大历史事件。图 1.48(b)所示为德国 Fraunhofer 研究所的 Mobile AR 项目^[62],该项目利用增强现实技术将柏林墙的时代变迁呈现于用户手持超便携 PC 上。用户只要将手持的 UMPC 对准柏林墙拍摄,显示屏便呈现出几十年前柏林墙的旧貌。图 1.48(c)展示了瑞士日内瓦大学的 Miralab 实验室在古遗址以及古代人物情景的数字化增强再现方面的突出的研究成果^[63]。



图 1.48 增强现实技术在文化遗产保护中的应用

目前国内将增强现实技术应用于文物古迹遗址保护方面的工作也在相继展开。例如,北京理工大学的圆明园数字化重建项目^[64,65],利用增强现实技术将昔日圆明园辉煌的景观叠加到目前的遗址上,以固定式和移动式两种方式实时展现,见图 1.49 和图 1.50。



图 1.49 圆明园定点增强现实系统^[64]



图 1.50 移动增强现实系统^[65]

参 考 文 献

- [1] Azuma R, Baillot Y, Behringer R, et al. Recent advances in augmented reality. IEEE Computer Graphics and Applications, 2001, 21: 34 ~ 47.
- [2] MIT Sixth Senses. <http://www.pranavmistry.com/projects/sixthsense> [2012-5-20].
- [3] Grzeszczuk R, Chin S, Fahrig R, et al. A fluoroscopic X-ray registration process for three-dimensional surgical navigation. Medical Image Computing and Computer-Assisted Intervention, 2000, 1935: 551 ~ 556.
- [4] Azuma R T. A survey of augmented reality. Presence: Teleoperators and Virtual Environments,

- 1997,6(4):355 ~ 385.
- [5] What is Anatomic VisualizeR®. http://meded.ucsd.edu/index.cfm/edcom/web_app_development/grantsupported_projects/virtual_reality/what_is_anatomic_visualizer[2010-1-5].
 - [6] SurVis-1st International Workshop on Surgical Vision. <http://ranger.uta.edu/~gianluca/survis> [2011-12-2].
 - [7] Milgram P, Kishino F. A taxonomy of mixed reality visual displays. IEICE Transactions on Information and Systems, Special issue on Networked Reality, 1994, E77-D(12):1321 ~ 1329.
 - [8] Boom Displays. <http://www.fakespacelabs.com/tools.html> [2009-3-24].
 - [9] Polhemus Navigation Systems. <http://www.polhemus.com> [2009-5-12].
 - [10] Ascension Technology Corporation. <http://www.ascension-tech.com> [2009-5-24].
 - [11] The UNC Tracker Project. <http://www.cs.unc.edu/~tracker/>[2010-7-7].
 - [12] Welch G, Bishop G, Vicci L, et al. High-performance wide-area optical tracking; The Hi-Ball tracking system. Presence: Teleoperators and Virtual Environments, 2001, 10(1):1 ~ 21.
 - [13] Vicon 公司的光学跟踪系统. <http://www.vicon.com> [2010-6-28].
 - [14] Motion Analysis 公司的光学跟踪系统. <http://www.motionanalysis.com.cn> [2010-6-28].
 - [15] 德国 A. R. T 公司的 AR Track 产品. <http://www.ar-tracking.com>[2010-6-29].
 - [16] 加拿大 NDI 数字公司的 OI 跟踪器系列. <http://www.ndigital.com>[2010-6-29].
 - [17] Foxlin E, Durlach N. An inertial head orientation tracker with automatic drift compensation does use with HMD's//Proceedings of Virtual Reality Software and Technology, Singapore, 1994.
 - [18] InterSense 公司的惯性跟踪器产品. <http://www.intersense.com> [2010-7-2].
 - [19] XSense 公司的惯性跟踪器产品. <http://www.xsens.com> [2010-7-2].
 - [20] Kato H. ARToolkit CP/OL. <http://www.hitl.washington.edu/projects/ARToolkit>[2008-4-24].
 - [21] Uchiyama S, Takemoto K, Satoh K, et al. MR platform; A basic body on which mixed reality applications are built//Proceedings of International Symposium on Mixed and Augmented Reality, Darmstadt, 2002.
 - [22] Augmented Reality (Vuforia). <https://developer.qualcomm.com/mobile-development/advanced-features/augmented-reality-vuforia>[2012-2-15].
 - [23] D'fusion Studio. <http://www.t-immersion.com/products/dfusion-suitedfusion-studio> [2012-4-28].
 - [24] Metaio SDK. <http://www.metaio.com/products/sdk>[2012-3-12].
 - [25] BazAR; A Vision Based Fast Detection Library. <http://cvlab.epfl.ch/software/bazar>[2011-2-22].
 - [26] Klein G, Murray D. Parallel tracking and mapping for small AR workspaces//The 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, Nara, 2007.
 - [27] Klein G, Murray D. Parallel tracking and mapping on a camera phone//International Symposium

- on Mixed and Augmented Reality, Orlando, 2009.
- [28] Zhou F, Duh H B L, Billinghurst M. Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR//Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, Cambridge, 2008.
- [29] Chandaria J, Thomas G A, Stricker D. The MATRIS project: Real-time markerless camera tracking for augmented reality and broadcast applications. Journal of Real-Time Image Proc, 2007, 2: 69 ~ 79.
- [30] William A H. Fusion of data from head-mounted and fixed sensors//Proceedings of IWAR, San Francisco, 1998.
- [31] iPhone 触摸屏工作原理. <http://news.ipadown.com/352> [2012-8-28].
- [32] PointGrab. <http://www.pointgrab.com/274/consumer-electronics> [2012-8-28].
- [33] Kinect Sports. <http://www.it.com.cn/games/tv/qz/2010/06/15/14/823933.html> [2010-6-15].
- [34] BMW Augmented Reality- The Extension of Reality. http://www.bmw.com/com/en/owners/service/augmented_reality_workshop_1.html [2012-6-28].
- [35] Marco Tempest: A Magical Tale (with augmented reality). <http://www.ted.com/topics/magic> [2013-5-20].
- [36] Johnny Lee: Free or Cheap Wii Remote Hacks. http://www.ted.com/talks/johnny_lee_demos_wii_remote_hacks [2013-8-6].
- [37] State A, Chen D T, Tector C, et al. Case study: Observing a volume-rendered fetus within a pregnant patient//Proceedings of IEEE Visualization, Washington D C, 1994.
- [38] Grimson W, Lozano-Pérez T, Wells W, et al. An automatic registration method for frameless stereotaxy, image guided surgery and enhanced reality visualization//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Seattle, 1994.
- [39] Weghorst S. Augmented reality and parkinson's disease. ACM Comm, 1997, 40(8): 47, 48.
- [40] Improving Depth Perception and Perception of Layout For in-situ Visualization in Medical Augmented Reality. <http://campar.in.tum.de/Chair/ProjectDepthPerceptionMedicalAR> [2009-4-3].
- [41] Augmented Reality for Maintenance and Repair. <http://kejiao.cntv.cn/20111208/100194.shtml> [2011-12-8].
- [42] Livingston M A, Brown D G, Julier S J, et al. Mobile Augmented Reality: Applications and Human Factors Evaluations. <http://wenku.baidu.com/view/60bc4dc408a1284ac850436b.html> [2014-8-18].
- [43] <http://kejiao.cntv.cn/20111208/100194.shtml> [2011-12-8].
- [44] MAN Diesel: Close-Quarter Engine Replacement. <http://www.metaio.com/customers/case->

- studies/man-replacement-of-a-ship-engine-in-cramped-spaces[2013-9-12].
- [45] 波音的 AR 系统. <http://esto.sysplan.com/ESTO/Displays/HMD-TDS/Factsheets/Boeing.html>[2009-3-16].
- [46] Reiners D, Stricker D, Klinker G, et al. Augmented reality for construction tasks; Doorlock assembly. Proceedings of IEEE and ACM IWAR, 1998, 98(1): 31 ~ 46.
- [47] Augmented Reality for Construction. <http://monet.cs.columbia.edu/projects/arc/arc.html> [2008-9-20].
- [48] Schwald B, Figue J, Chauvineau E, et al. STARMATE: Using augmented reality technology for computer guided maintenance of complex mechanical elements. E-work and ECommerce, 2001 (1): 196 ~ 202.
- [49] Arvika 系统. <http://www.arvika.de/www/index.html>[2010-4-20].
- [50] Navab N, Bascle B, Appel M, et al. Scene augmentation via the fusion of industrial drawings and uncalibrated images with a view to markerless calibration//Proceedings of 2nd International Workshop Augmented Reality, San Francisco, 1999.
- [51] Drascic D, Grodski J J, Milgram P, et al. ARGOS: A display system for augmenting reality//Video Proceedings of INTERCHI '93: Human Factors in Computing Systems, New York, 1993.
- [52] Kaufmann H. Construct 3D: An augmented reality application for mathematics and geometry education//Proceedings of the 10th ACM international Conference on Multimedia, New York, 2002.
- [53] Telerobotic Control with Stereoscopic Augmented Reality. <http://etclab.mie.utoronto.ca/people/anu/spie96/spie96a.html>[2013-9-24].
- [54] iButterfly. <http://ibutterfly.co.id/id/index.html>[2012-12-21].
- [55] Star Chart. <https://itunes.apple.com/us/app/star-chart/id345542655?mt=8> [2012-12-21].
- [56] Molyneaux D, Izadi S, Kim D, et al. Interactive environment-aware handheld projectors for pervasive computing spaces. Pervasive Computing, 2012, 73(19): 197 ~ 215.
- [57] 2012 伦敦奥运会孙杨 1500 米决赛夺冠视频. <http://gz.bendibao.com/news/201286/content102912.shtml>[2012-8-6].
- [58] 运用 AR 技术的直播画面. <http://www.heacn.net/2013/0221/55062.html>[2013-5-28].
- [59] Bing 街景地图. <http://blog.ted.com/2010/02/13/augmentedrealit>[2010-2-13].
- [60] Eternal egypt. http://www.eternalegypt.org/EternalEgyptWebsiteWeb/HomeServlet?ee_website_action_key=action.display.splash&ee_messages=0001.clientrequired.text[2009-9-21].
- [61] Vlahakis V, Karigiannis J, Tsotros M, et al. ARCHEOGUIDE: First results of an augmented reality, mobile computing system in cultural heritage sites//Virtual Reality, Archaeology and Cultural Heritage International Symposium, Glyfada, 2001.
- [62] Zöllner M, Keil J, Wust H, et al. An augmented reality presentation system for remote cultural

- heritage sites//The 10th International Symposium on Virtual Reality, St. Julians, 2009.
- [63] Papagiannakis G, Schertenleib S, O' Kennedy B, et al. Mixing virtual and real scenes in the site of ancient Pompeii. *Computer Animation and Virtual Worlds*, 2005, 16(1): 11 ~ 24.
- [64] 王涌天, 林惊, 刘越, 等. 亦真亦幻的户外增强现实系统——圆明园的数字重建. *中国科学基金*, 2006, 20(2): 76 ~ 80.
- [65] Chen J, Wang Y T, Guo J W, et al. Augmented reality registration algorithm based on nature feature recognition. *Science in China Series F: Information Sciences*, 2010, 53(8): 1555 ~ 1565.

第2章 增强现实系统的显示技术

人类的大脑通过视觉、听觉、触觉、嗅觉等感官功能与客观存在的真实世界形成信息通道,而构建增强现实系统的目的是通过扩展这些信息通道,使得进入人脑的信息得到增强。视觉通道是人类与外部环境之间最重要的信息接口,研究表明,人类从外界所获得的信息大约 80% 是通过眼睛得到的^[1]。因此,增强现实系统中的显示技术就显得尤为重要,是构建一个成功的增强现实系统的关键技术之一。

2.1 显示设备的分类及概述

2.1.1 屏幕显示器

屏幕显示是最传统,也是最常见的显示方式。因为增强现实技术是在虚拟现实技术的基础上发展起来的,所以虚拟现实系统中的三维屏幕显示器也可用在增强现实系统中。三维显示能够提供直观、自然的交流模式,因而有助于人们综合运用各种深度暗示,获得真实、丰富、可靠的感知体验。三维显示器按照工作原理可以分为视差型三维显示、透视三维显示、全息三维显示和真三维显示等几类。

视差型三维显示是最早实现的立体显示方式,可分为需要助视工具和不需要助视工具两大类。需要佩戴偏振眼镜、互补色眼镜或液晶眼镜的视差型三维显示器是最早出现并产业化的立体显示装置。该类显示器通过时间分割或者空间分割的方式实现左右眼分别观看一幅视差图,再通过大脑融合后形成立体影像。现在一些大型影院或科普单位依然采用这类显示方式。例如,北京天文馆采用这类显示设备播放“史前海洋生物”、“恐龙时代”等节目获得了较好的科普效果。这类显示装置由于需要佩戴专用的眼镜,成本较高,用户使用起来也不是很方便。

随着 LCD 和 PDP 的出现,不需要助视工具的自由立体显示器随之问世,这类显示装置采用光栅或柱面栅分离左右眼图像,目前该类型显示装置也已产业化。荷兰的 Philips 公司、美国的 DTI 公司、德国德雷斯頓技术大学以及日本的 Sony 公司都各自研制了 26.4 ~ 53.3cm (10.4 ~ 21in) 不等尺寸的立体液晶显示器或立体 PDP 显示器。国内的欧亚宝龙公司也已推出了自己的自由立体显示器。视差型三

维显示虽然能够带给用户具有深度空间感觉的虚拟立体感,但是它仅提供分离的视区和有限的视点个数。因此用户在长时间观看时,无论是否需要戴特殊眼镜,这种靠大脑融像产生的非自然深度感都会导致用户出现头痛、恶心等反应。

透视三维显示利用阴影和遮挡等多种心理暗示,采用计算机图形学技术对深度信息进行编码,将三维场景以透视形式显示在二维平面上。实际上,利用这种方法生成的三维图像是显示在平面显示器上的二维图像,但它能够提供深度错觉。透视三维显示只能提供部分单眼深度暗示,缺乏运动视差以及双眼深度暗示,只适用于对深度估计精度要求不高的场合,整体感知能力的提高是以牺牲细节为代价的。

全息三维显示利用光的干涉原理,将物体散射或发射的特定光波以干涉条纹的形式记录下来,然后利用光的衍射原理,在一定的条件下再现。全息术的特点在于记录了完整的波前信息,能够提供一种与观察原物时相同的视觉效果。但也正因为如此,其信息量比其他三维图像大了好几个数量级。大场景全息三维显示信息量之大,对空间光调制器、计算机的处理速度、存储容量和传输带宽的要求之高,都是目前软硬件技术所无法实现的。

真三维显示是根据屏幕的形状、运动方式和位置,将三维物体切片并解析成二维图像。当屏幕运动时,能在指定位置将解析的图像显示于屏幕上,利用人眼视觉暂留重构三维图像。真三维显示是三维显示的最终目标,它是一种能够实现 360° 视角观察的三维显示技术,是现实景物最真实的再现。与二维显示器相比,由之再现的三维场景真实地存在于用户所在的物质世界中,能提供几乎所有的视觉深度暗示,具有更宽的视场和更大的视距范围,完全符合人类对真实场景的观看方式,能够更好地传达物体间的相互空间位置关系信息,进而降低发生误判或错判的可能性,协助用户更快地作出判断。与 2D 显示器相比,真三维显示器可作为 3D 设计、可视化的显示终端。在模具设计和外形设计等领域,如飞机、汽车或者手机外形设计方面,在未生产样品前就能够获得几乎与实际物体相同的直观效果,有助于提高设计和制造人员的工作效率,缩短产品的试制周期。在飞机风洞实验中,配合各种传感器,利用该系统可以帮助设计人员更加准确地完成空气动力学的设计和实验。由于真三维显示图像无遮挡,具有透视效果,该系统尤其适用于医疗中 CT 图像的显示,可以帮助医生更好地确定病灶的位置,从而进行有效治疗。除此以外,真三维显示在机械装配、气象分析和空中交通管理、军事模拟以及广告娱乐等领域也具有广阔的应用前景。

1996 年,美国海军海洋系统中心(NOSC)在美国国防部(DOD)和海军(Navy)

的资金资助下推出了第二代基于氦氖离子激光器的真三维显示装置。美国空军和美国航空航天局(NASA)也非常重视真三维显示技术,在20世纪90年代投入了大量的资金资助研究。得州仪器公司(TI)在美国空军的支持下,大约在1990年提出一项基于DMD的真三维投影技术^[2]。GENEX等多家美国公司受NASA的委托,致力于将具有真实物理深度和全视景观察的真三维旋转面技术应用于航空、宇航以及地球科学等研究领域。2002年,美国Actuality公司推出了基于高频单投影机和平旋转屏的真三维显示器Perspecta™ 3D系统^[3]。2005年4月,美国拉什大学医学中心(Rush University Medical Center)首次试行将Actuality公司的Perspecta™ 3D系统用于辅助癌症患者的放射治疗诊断(图2.1)。2005年5月,美国空军和世界上最大的石油和天然气公司相继购买Perspecta™ 3D系统显示器用于战场可视化和石油天然气勘探。同期德国不伦瑞克工业大学的飞行器导航和控制研究所开发出了球状旋转面显示的原型化显示器FELIX(图2.2)、FELIXII^[4,5]。

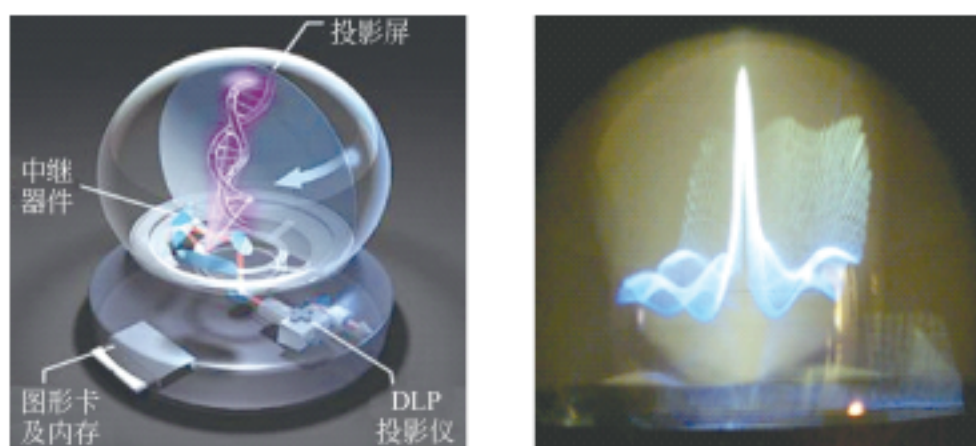


图2.1 Perspecta™ 3D系统^[3]

浙江大学于2005年建立了图2.3所示的基于二维LED旋转屏的体三维显示装置(图2.3),实现了直径为305.6mm,高76mm的体显示空间,体分辨率为235904,1bit灰阶^[6]。南京航空航天大学于2008年建立了如图2.4所示的高帧频DMD投影机和螺旋屏的真三维显示装置^[7,8]。北京理工大学于2008年6月研制了图2.5所示的基于多投影机旋转屏的真三维显示系统^[9,10],该系统采用多投影机方式突破了分辨率、灰阶和帧频相互制约的瓶颈,配合数据并行传输技术,使投影机的刷新率和分辨率有了很大的提升。

图2.6所示为丹麦ViZoo公司设计的360°全息显示器Cheoptics360^[11],该显示器为倒金字塔形状,四面由透明材料做成,观众可以从各个角度观看。实际上,该显示器并不是真正的三维显示器,仅是能够显示四个面的图像。但是,在广告、产

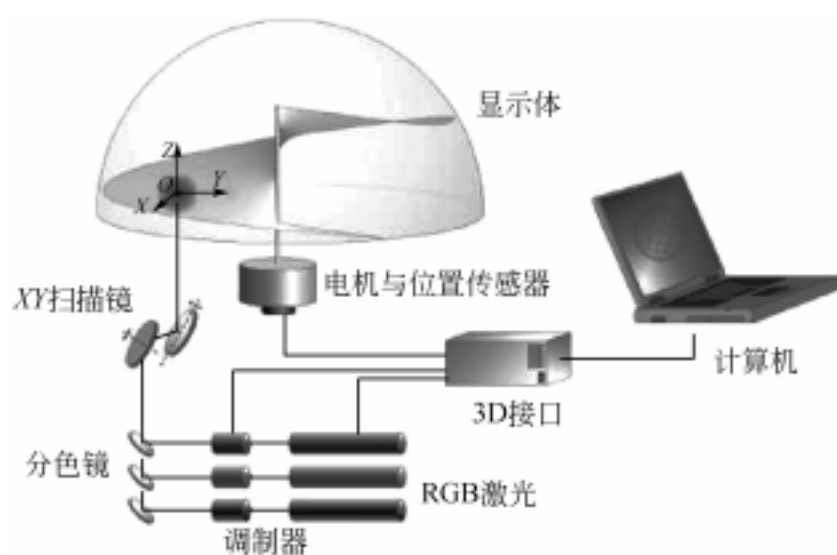


图 2.2 FELIX 3D 显示器

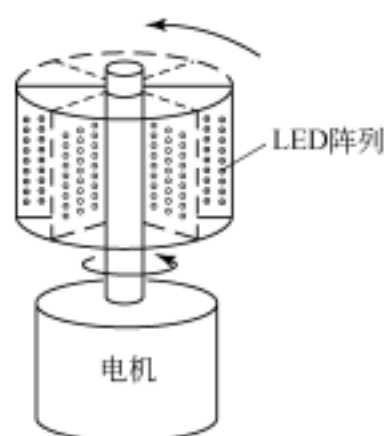


图 2.3 基于二维 LED 旋转屏的体三维显示装置^[6]

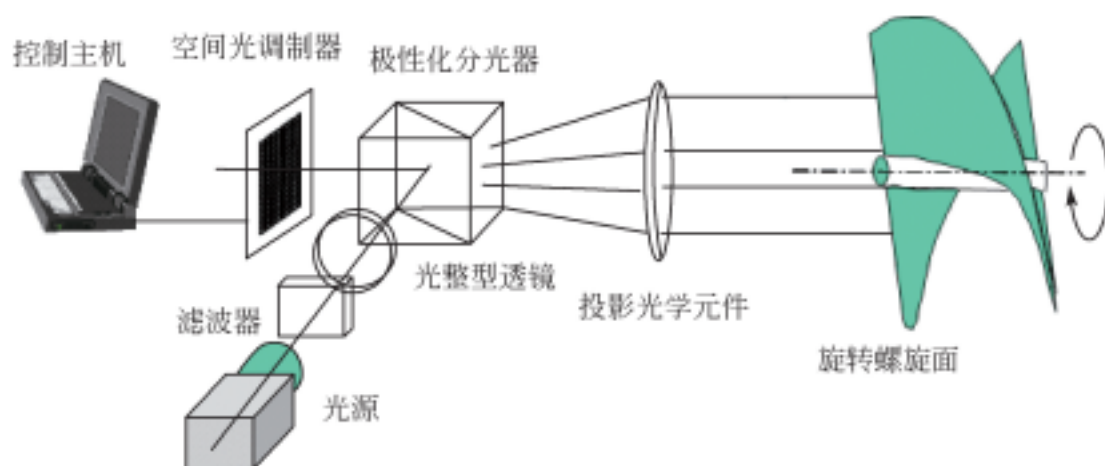


图 2.4 采用螺旋屏的真三维显示装置^[8]

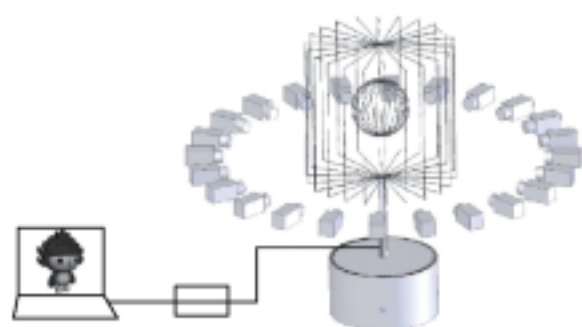
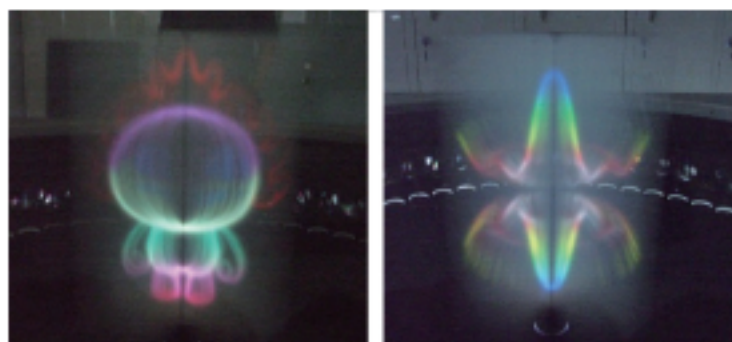
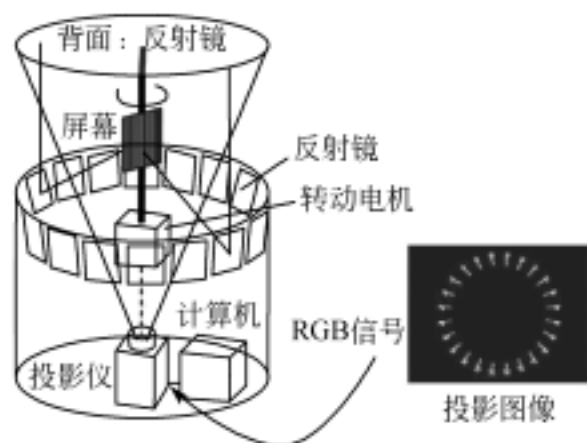
(a) 基于多投影机旋转屏的真三维显示系统^[9](b) 基于多投影机旋转屏的真三维显示装置显示效果^[10]

图 2.5 基于多投影机旋转屏的真三维显示系统与显示装置

品展示中该显示器还是显示出了巨大的吸引力,尤其是在 2008 年佛罗伦萨春季服装表演上使得海洋生物与演员同台演出,起到了很好的宣传效果(图 2.7)。Hitachi 公司开发了 360°全视角 3D 显示系统,该系统由单投影机、24 面反射镜和定向反射屏组成,可以同时显示 24 个面的图像,如图 2.8 所示。

图 2.6 ViZoo Cheoptics360^[11]图 2.7 2008 年佛罗伦萨春季服装表演^[12]图 2.8 360°全视角 3D 显示系统^[13]

尽管不能提供真正的三维显示效果,上述大多数显示设备都有很好的立体显示效果,在很多应用系统中都得到了成功的应用。由于增强现实的主体是真实环境,不像虚拟现实那样要求用户完全沉浸在虚拟环境中,需要添加的虚拟信息只是起到增强辅助的作用。因此,包括真三维显示设备在内的各类屏幕显示器在增强现实系统中都得到了广泛应用。

2.1.2 头盔显示器

头盔显示器(head-mounted display, HMD)是增强现实系统中最典型的显示设备。顾名思义,HMD 是安装在头盔上的显示装置。用户佩戴上 HMD 后,在其头部运动时,HMD 将随用户的头部运动而运动。增强现实系统的运动跟踪装置一般都安装在 HMD 上,从而可以实时跟踪用户头部的位置和方向。根据这些方位信息,就可以计算出对应于用户当前姿态的虚拟对象的位姿,并将其显示在头盔显示器的屏幕上。

VR 系统强调用户在虚拟环境中的视觉、听觉、触觉等感官的完全浸没,强调将用户的感官与现实世界绝缘而沉浸在一个完全由计算机所控制的信息空间(cyberspace)之中。通常需要借助能够将用户视觉与现实环境隔离的显示设备,一般采用浸没式头盔显示器(immersive head-mounted display)。由于 AR 系统致力于将计算机产生的虚拟环境与真实环境融为一体,增强用户对真实环境的理解,因此需要借助能够将虚拟环境与真实环境融合的显示设备,通常采用透视式 HMD。



图 2.9 Sutherland 设计的 HMD^[14]

透视式 HMD 最初发明于 20 世纪 60 年代,Sutherland 1965 年设计的光学透视式 HMD 和 1968 年设计的立体 HMD(图 2.9)是最早的使用微型 CRT 的基于 HMD 的计算机图形显示设备^[14,15]。几乎所有的后续透视式 HMD 都是光学透视式的,如 VCASS 系统^[16,17]、Tilted Cat HMD^[18]和 CAE Fiber-Optic HMD^[19]都是光学透视式 HMD 的例子,Kaiser Electro-Optics 公司和 McDonnell Douglas 公司研制了其中的一些系统^[20]。

图 2.10 给出了头盔显示器发展的里程碑图,列出了头盔显示器发展史上的一些重要进展。

目前,国际市场上生产头盔显示器的厂家及其主要产品如下。

1) Kaiser Electro-Optics

Kaiser Electro-Optics 公司是最早、最著名的头盔显示器生产厂商之一。



图 2.10 头盔显示器发展里程碑图

ProView™ XL 系列头盔显示器是该公司用于虚拟现实的高档头盔显示器(图 2.11), 它根据其对角线视场分为 XL35 和 XL50 两种, 支持全彩色 XGA [1024 (H) × 768 (V)] 输入, 垂直刷新频率为 60Hz, 瞳孔间距为 55 ~ 75mm 可调, 亮度可调, 对比度较高, 质量为 35oz (1oz = 28.35g), 售价为 19500 美元。

图 2.11 ProView™ XL 头盔显示器^[21]

2) Virtual Research

Virtual Research 公司是研制中档头盔显示器的著名公司之一, 该公司生产的 VR4、V6 和 V8 头盔在我国用户较多。图 2.12 所示的 V6 和 V8 头盔显示器采用 640×480×3 像素的 LCD, 支持 VGA 输入, 刷新频率为 60Hz, 对比度达到 200 : 1, 对角线视场为 60°, 配有立体声耳机, 其中 V8 的瞳孔间距为 52 ~ 74mm 可调, 质量为 1.0kg, V6 质量为 821g。



图 2.12 V6 和 V8 头盔显示器^[22]

3) Virtual I/O

图 2.13(a)和(b)所示为 Virtual I/O 公司生产的 i-glasses™ LC 和 i-glasses™ LCB,其中 i-glasses™ LCB 是低端产品的代表,使用 18 万像素的 LCD,支持 S-video, NTSC 输入,视场角为 30°,质量为 226g,售价低于 500 美元。



图 2.13 Virtual I/O 头盔显示器^[23,24]

图 2.13(c)所示的 i-glasses™ SVGA 是 Virtual I/O 公司推出的高清晰度、大视场头盔显示器,支持 SVGA、PAL、NTSC 输入,24 位颜色深度,分辨率为 800×600 像素,对角线视场角为 27°,刷新频率为 120Hz,出瞳直径为 6mm,出瞳距离 17mm,质量小于 200g。i-glasses™ SVGA 2D 售价为 699 美元,i-glasses™ SVGA 3D 售价为 999 美元。

4) 5DT(Fifth Dimension Technologies)

5DT 公司是专攻虚拟现实领域的高科技公司,研发并生产虚拟现实用的硬件设备以及开发系统软件等。其头盔显示器产品主要包括以下几款。

(1) 5DT HMD 800:支持 SVGA、PAL、NTSC 输入,单眼分辨率达到 800×600×3 像素,视场角为 28°(H) × 21°(V),配有高保真立体声耳机,重 594g。不带立体视觉的 HMD 售价为 2950 美元,带立体视觉的 HMD 售价为 4950 美元,如图 2.14(a)所示。

(2) Cy-Visor 系列是第一个采用反射式基于硅片的液晶器件(LCOS)的 SVGA 头盔显示器。其中 DH-4400VP 使用 144 万像素的 0.49in LCD,支持 VGA/SVGA、

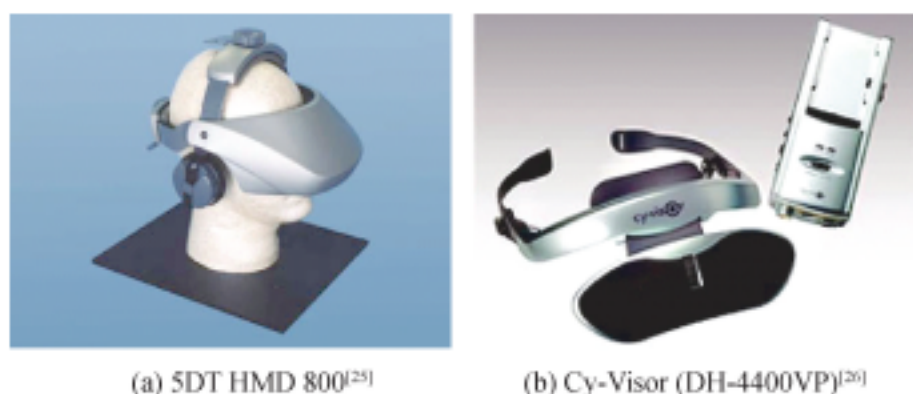
(a) 5DT HMD 800^[25](b) Cy-Visor (DH-4400VP)^[26]

图 2.14 5DT 公司的头盔显示器

NTSC/PAL、S-VHS 输入,瞳孔间距可调,对角线视场角为 31° 。其显示效果就像人在 2m 的距离观看 44in 画面一样,可调节亮度、对比度、锐度以及色调等,售价为 1150 美元,如图 2.14(b)所示。

5) Sony

Sony 公司在 LCD 微显示器上拥有多项专利。为将 HMD 应用于消费类电子产品,不惜投巨资研究开发 HMD 专用 LCD 微显示器。Sony 公司的 HMD 专用 LCD 微显示器有 0.55in ($1\text{in} = 2.54\text{cm}$),分 18 万像素和 24 万像素两种,并在 1998 ~ 1999 年推出一系列基于上述微显示器的 HMD, Sony 公司将它们命名为 Glasstron。Sony Glasstron PLM-A35 及 PLM-A55(图 2.15)支持 S-Video、VGA 输入,7in 18 万像素 LCD,分辨率为 $800(\text{H}) \times 255(\text{V})$ 像素,感觉就像从 6.5ft 远处观看 52in 电视屏幕。配备可调亮度、色彩和音量控制装置,质量仅 100g。其中,PLM-A55 可以利用 LCD Shutter 实现透视功能。PLM-A35 售价为 295 美元,PLM-A55 售价为 495 美元。Sony 公司在 2011 年推出家庭影院 HMZ-T1,2012 年 9 月又在德国柏林 IFA 国际消费类电子产品展上发布了 HMZ-T2,如图 2.16 所示。

图 2.15 Sony 公司的 Glasstron PLM-A35 和 Glasstron PLM-A55^[27]



图 2.16 Sony 公司的家庭影院 HMZ-T1 和 HMZ-T2^[28]

6) Olympus

作为光学产品和消费类电子产品的领导厂商, Olympus 公司从 1996 年开始积极进行 HMD 的研究开发, 现已形成比较完整的产品线。其中包括 Eye-Trek FMD-01/200/220/250/700 等产品。为了将 Eye-Trek 设计成容易佩戴的 FMD (face mounted display), Olympus 完成了一个技术上的突破, 即利用自由曲面棱镜 (free-shaped prism) 设计 HMD 的光学系统。Eye-Trek FMD-200 支持 PAL 和 NTSC 输入, 具有 18 万像素 TFT 液晶片, 视场角水平 35° 、垂直 26.6° (如同人在 2m 的距离观看 62in 画面的效果)。该 HMD 亮度高, 不受外界光线强弱的制约, 质量约为 85g, 售价为 499 美元。Eye-Trek FMD-700 与 Eye-Trek FMD-200 (图 2.17) 最大的区别在于, 引入了光学超分辨率 (optical super resolution) 技术, 能使分辨率从 18 万像素提高到 72 万像素, 质量约为 105g。



(a)Eye-Trek FMD-200

(b)Eye-Trek FMD-700

图 2.17 Olympus 公司的 Eye-Trek FMD-200 和 Eye-Trek FMD-700^[29,30]

2.1.3 投影式头盔显示器

投影式头盔显示器 (head-mounted projective display, HMPD) 的概念是 1997 年由 Ferguson 提出的^[31]。Rolland 等^[32]在医学应用中探索了投影式显示器的应用潜力。Hua 等^[33]采用双高斯透镜结构, 用市场上能买到的光学元件做了第一代样

机,并在论证 HMPD 成像概念的可实施性和量化回复反射物质在成像系统中的特性上做了很多工作。此后 Hua 和 Rolland 等在引入衍射光学元素(diffractive optical element)和塑料元件的基础上研制了超轻型、高投影质量的镜头,并结合用户自定义的镜头完成了一种紧凑样机^[34],该样机能达到 50° 的视场角,质量仅 750g。Kawakami 等研发了名为 X'tal Vision 的类似光学结构,并提出了面向对象的显示(object-oriented)和视触觉显示(visual-haptic display)^[35]。

投影式头盔显示器用一对小型投影透镜结合取代了传统的光学目镜,这样的结合使得投影式头盔显示器较传统的头盔显示器在许多方面都具有优势。例如,采用目镜的头盔显示器的尺寸会随着视场角的增大而增大,而采用投影透镜的头盔尺寸不会随着视场角的增大而增大。这样的特性可以为光学透视式头盔显示器设计出具有大视场角、超轻与小型化的光学器件。如图 2.18 所示,Hua 等研制的投影式头盔显示器的结构组成包含一对小型投影透镜、一组分光镜、固定在头盔上的显示器(LCD、CRT)以及直接放置在环境中的回复反射物质。

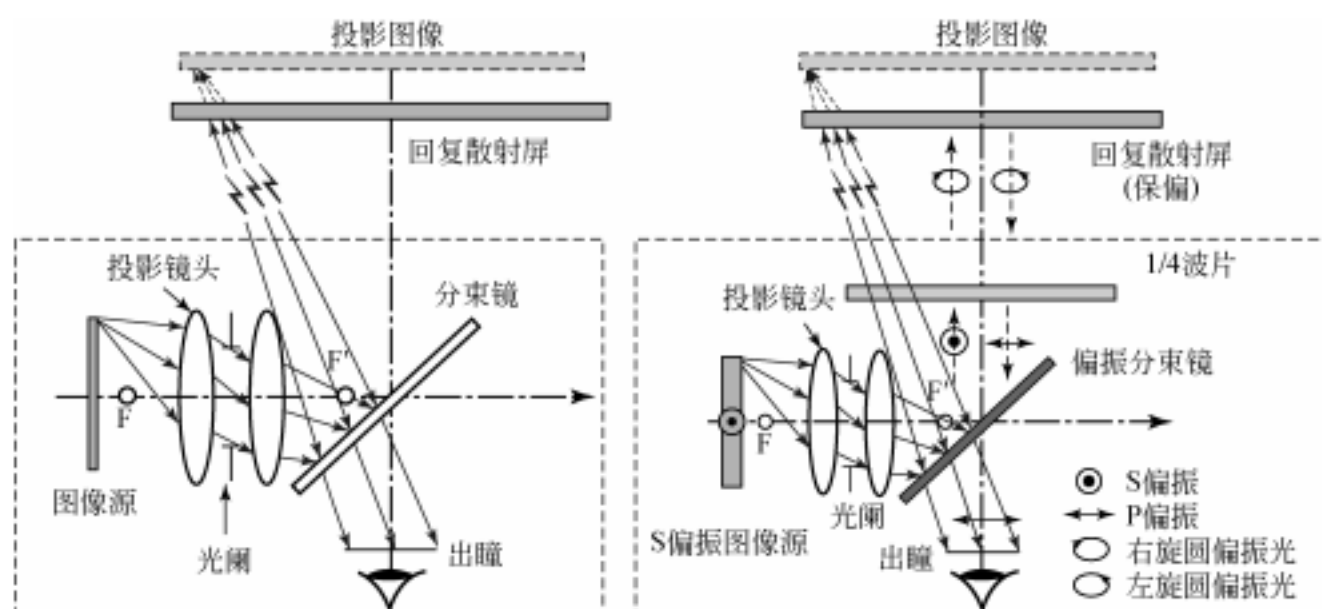


图 2.18 投影式头盔显示器的工作原理(第一代和第二代)^[36]

由图 2.18 可以看出,系统用投影透镜取代传统头盔显示器的目镜,用回复反射屏幕来取代立体投影系统中的漫射屏幕。与传统头盔显示器的不同之处在于,用来显示计算机生成图像的小型显示器被放置在透镜焦点以外的位置,而不是在透镜及其焦点之间。尽管投影透镜和分光镜与光轴成 45° 角,真实图像还是被投影到实际的空间当中。系统可在投影图像的前面或者后面放置一个回复反射屏幕。由于回复反射物质具有特殊性质(图 2.19),投影图像发出的光被反射到了光学出

瞳,使用者就可以获得一个由计算机生成的图像。在理想情况下,图像的位置和尺寸与反光屏幕的位置和形状没有关系。此外,与基于目镜的光学透射式头盔显示器相比,投影光学的引入还可以在获得更大视场角的同时降低畸变。为了满足眼睛间隙大以及不对称设计中限制目镜孔径的要求,使用分束镜可以使光学透射式头盔显示器获得的最大视场角,大约为 40° ,但同时边缘视场的畸变大于 15%。而如果使用投影透镜的对称设计,投影式头盔显示器的最大视场角可以达到 90° (一般为 $50^\circ \sim 70^\circ$),而且边缘视场的畸变能够降低到 15% 以下。

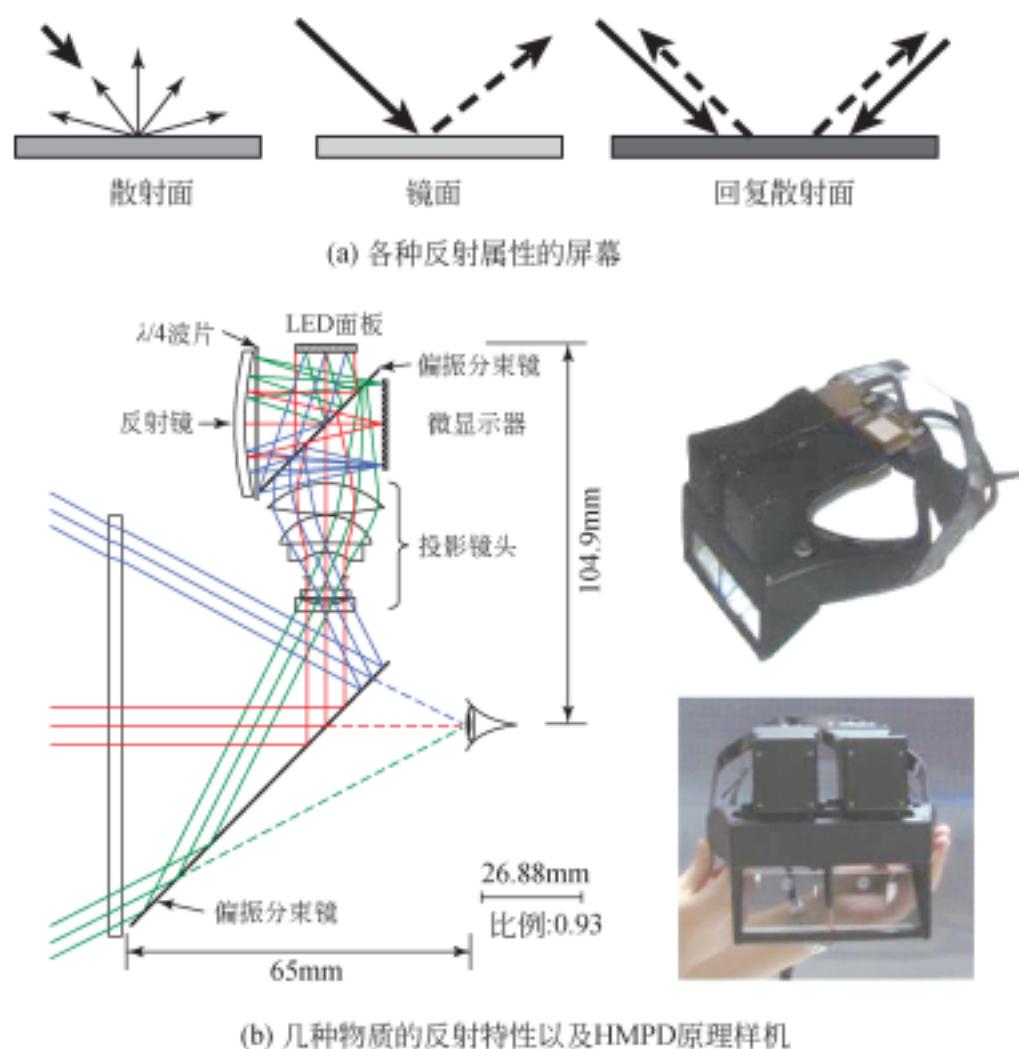
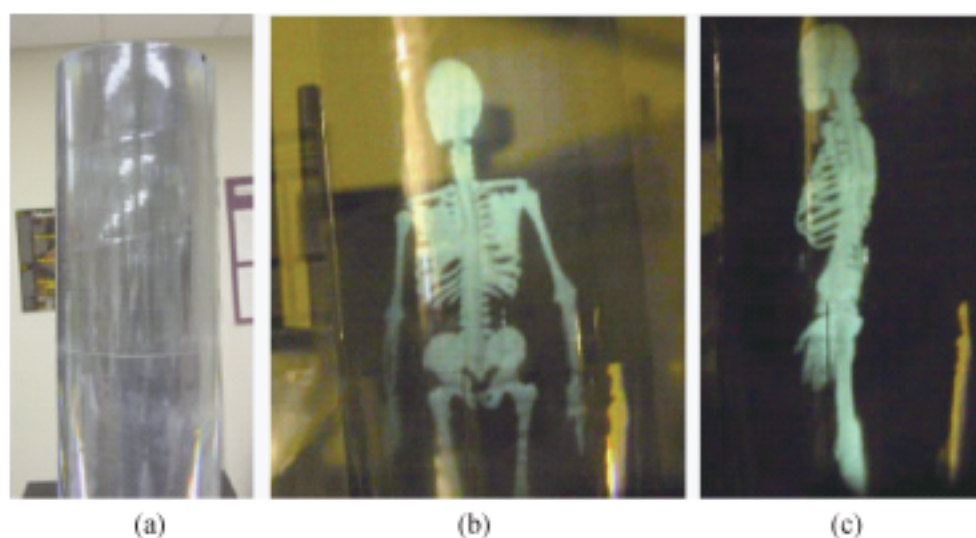


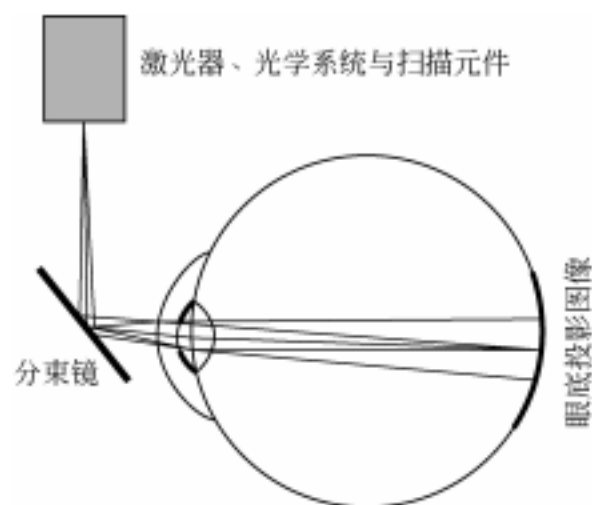
图 2.19 不同反射属性的屏幕与 HMPD 原理及样机图^[37]

图 2.20 为投影式头盔显示器的应用实例。在这个应用中,通过投影式头盔显示器和圆柱体,用户能够感知到一个完整的骨架,而且它好像站在圆柱里面一样。用户可以绕着圆柱体或直接转动圆柱体从不同的视点观察骨架。图 2.20(b)和图 2.20(c)是从头盔的左侧出瞳截取的,它清楚地展示了弯曲的屏幕,并没有扭曲感知到的投影图像。

图 2.20 圆柱屏幕及其应用^[38]

2.1.4 视网膜直接成像显示器

视网膜直接成像显示器也是一种头戴式显示设备,它利用低功率半导体激光器扫描并产生调制光,直接成像于人体视网膜上。这种显示器能生成高亮度、高分辨率的宽场图像,其成像原理如图 2.21 所示。视网膜显示技术的主要优势在于,其所成图像的亮度与对比度高且设备功耗低,这些优势使该设备能很好地满足户外机动使用的条件。未来可能发展为一种实现自动聚焦、高分辨率、大视场的立体显示技术。但是,目前的视网膜显示技术仍具有以下几方面的技术缺陷。

图 2.21 视网膜显示器原理图^[39]

(1) 由于目前暂时没有较便宜的低功率蓝绿激光器,显示器所成的像是单色的(红色)。

(2) 由于图像直接成在视网膜上,不支持视距和焦距调节。

(3) 所成图像没有立体感。

2.1.5 投影显示设备

目前,视频透视式头盔显示器和光学透视式头盔显示器作为增强现实的传统显示输出设备已经使用了很多年。然而,它们也面临着技术上和人类环境改造学上的不足。例如,显示设备的小型化、应用光学的局限性限制了观察视场。此外,视觉的不断调整使观察者感觉不适等。为弥补上述不足,近年来国内外学者相继提出采用投影技术构建增强现实应用系统。

德国包豪斯大学的 Bimber 研究组根据图像的产生方法、图像相对观察者的位置以及图像的类型(如平面或者曲面等)将增强现实中的显示技术按图 2.22 进行了划分。由图 2.22 可以看出,投影技术是增强现实显示技术的重要组成部分。更多关于空间投影增强现实的内容可参考 Bimber 等出版的 *Spatial Augmented Reality Merging Real and Virtual Worlds*^[40]。

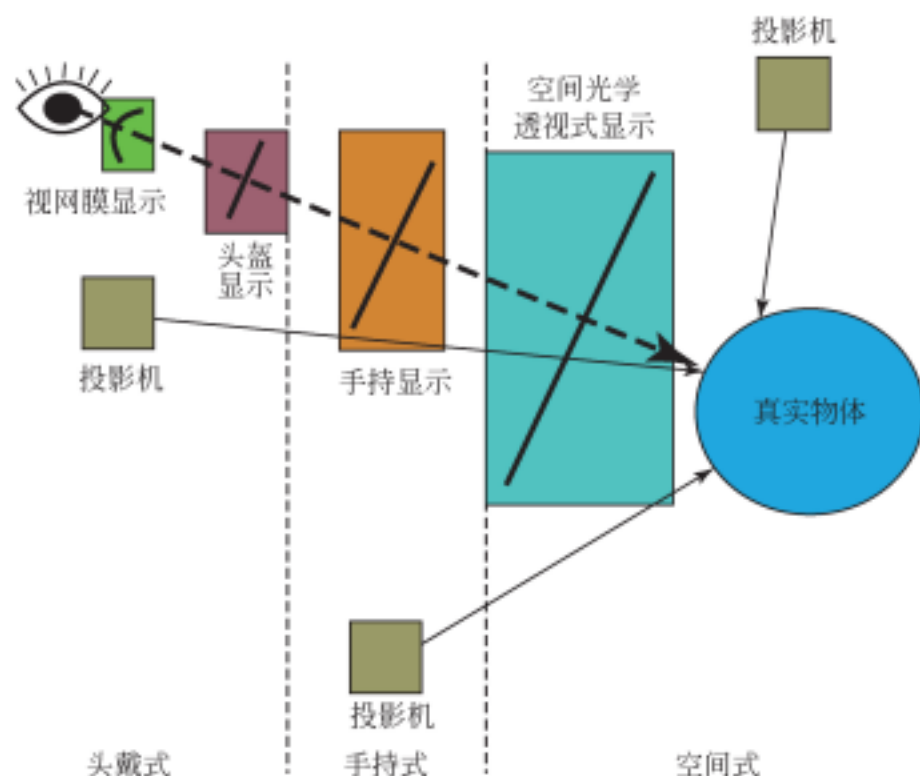


图 2.22 增强现实中的显示技术划分^[40]

1. 固定投影机的投影增强现实系统

投影显示 (projection display) 可以将所需的虚拟信息投影到真实对象上进行增

强。投影显示并不是简单地将虚拟图像投影在用户的视场中,而是将图像与用户所处的环境相融合后实现增强。

2001年,Low等的人类生活环境模型引起了人们的关注,如图2.23所示。他们首先利用相关软件构建了虚拟模型,然后用聚氯乙烯泡沫面板搭建了初级简单的建筑物实体(墙、柱、柜、桌等),并以此为真实物理显示环境,利用基于投影显示的增强现实技术,将其他细节(如墙壁颜色、壁上图画、窗外自然景观等)以图像的方式叠加到真实物体上,使用户在增强现实环境中有身临其境的感觉。图2.24所示为在一个白色漫射花瓶上(也可以是其他物体,如书籍和鲜花等真实对象),利用空间增强现实技术,在花瓶上投影具有阴影、散射效果、纹理、强度修正信息的投影图像,并通过修改模型的反射系数来改变视觉效果。2010年上海世博会上,投影增强显示更是处处可见,如图2.25所示。

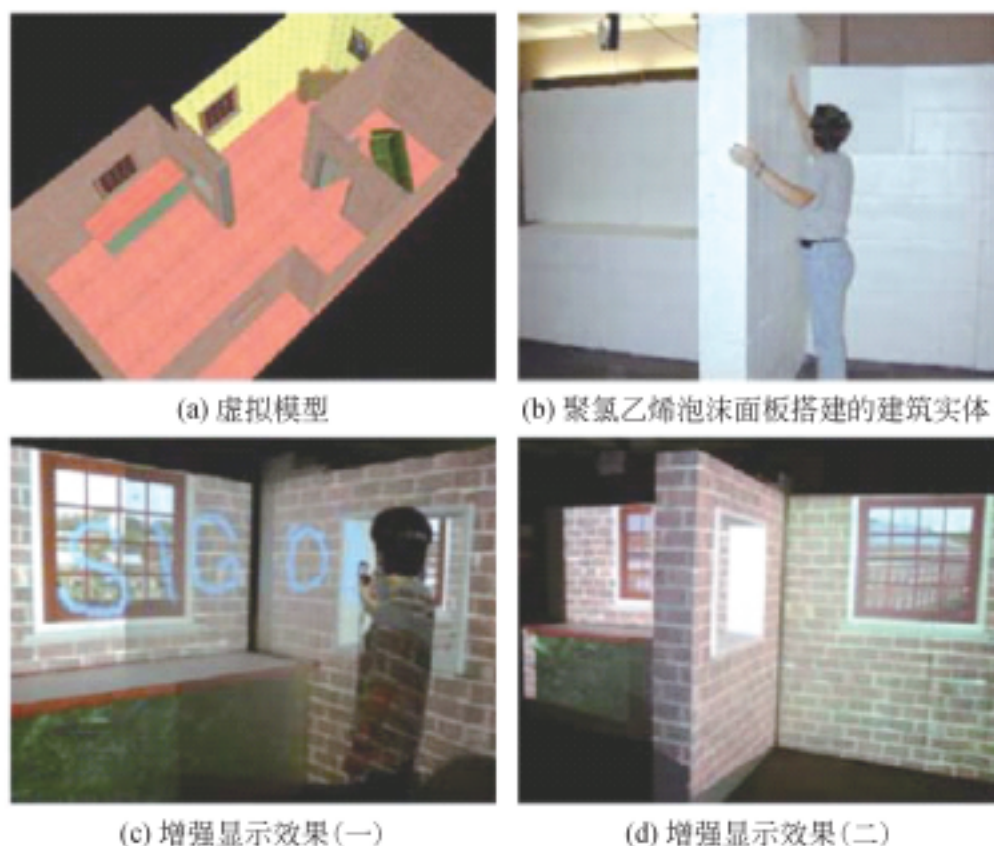


图 2.23 基于投影显示技术的大环境增强现实^[4]

2. 手持投影设备的投影增强现实系统

目前比较常见的基于投影显示技术的增强现实是采用固定的投影机形成一个固定的投影显示应用。近年来,随着科技的进步,数字投影机的外形越来越小巧、重量越来越轻,轻巧性和便携性的发展趋势为固定设备的可移动使用提供了可能。

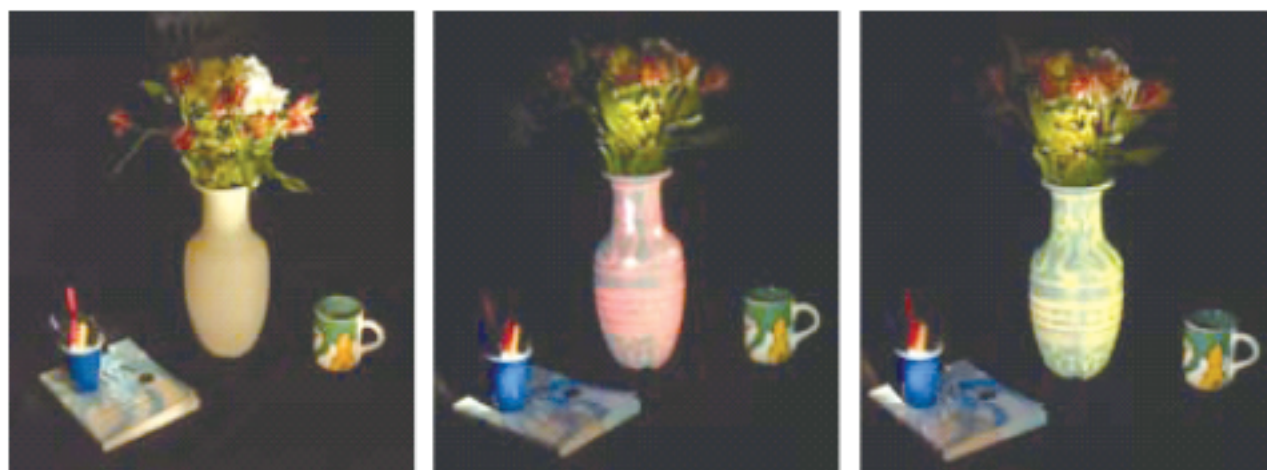


图 2.24 花瓶的着色(左),基础对象为白色花瓶(中、右),着色处理后的视觉效果^[42]



图 2.25 上海世博会投影增强显示部分展项^[43]

同时随着移动过程中自校正技术的突破,手持式投影已成为投影显示一个新的发展方向。

手持式投影设备一般由投影仪和照相机组合而成,2005 年三菱电气研究实验室的 Beardsley 等研制的手持式投影样机如图 2.26 所示^[44]。在投影机底部有手柄,手柄上食指放置的位置有按钮,总的组成部分包括以下几项。

- (1) 1024×768 像素、帧频为 60Hz 的 V-1080 投影机。
- (2) 640×480 像素、帧频为 100Hz 的 Basler A602F 相机。
- (3) 四个硬性“激光笔”。
- (4) 连接计算机的数据线。

InterSense 的 Foxlin 和 Naimark 研制了图 2.27 所示的手持式投影设备,提出应用手持视频投影机作为一个“手电筒”,以交互方式在实际表面上产生虚拟物体的阴影效应(图 2.27)。他们还将手持式视频投影仪和照相机结合起来,用于实验其光学跟踪系统的性能,这一观念可能被用于增强现实中的建筑和维修领域^[45]。

如图 2.28 所示,现实生活中的显示平面多种多样,包括平面、规则曲面和不规则曲面。近年来,曲面投影已越来越多地被运用于视觉环境中。2004 年三菱电气

研究实验室的 Raskar 等提出利用多台显示设备在二次曲面上实现图像无缝重叠投影的方法(图 2.29)^[46]。



图 2.26 手持式投影样机^[44]

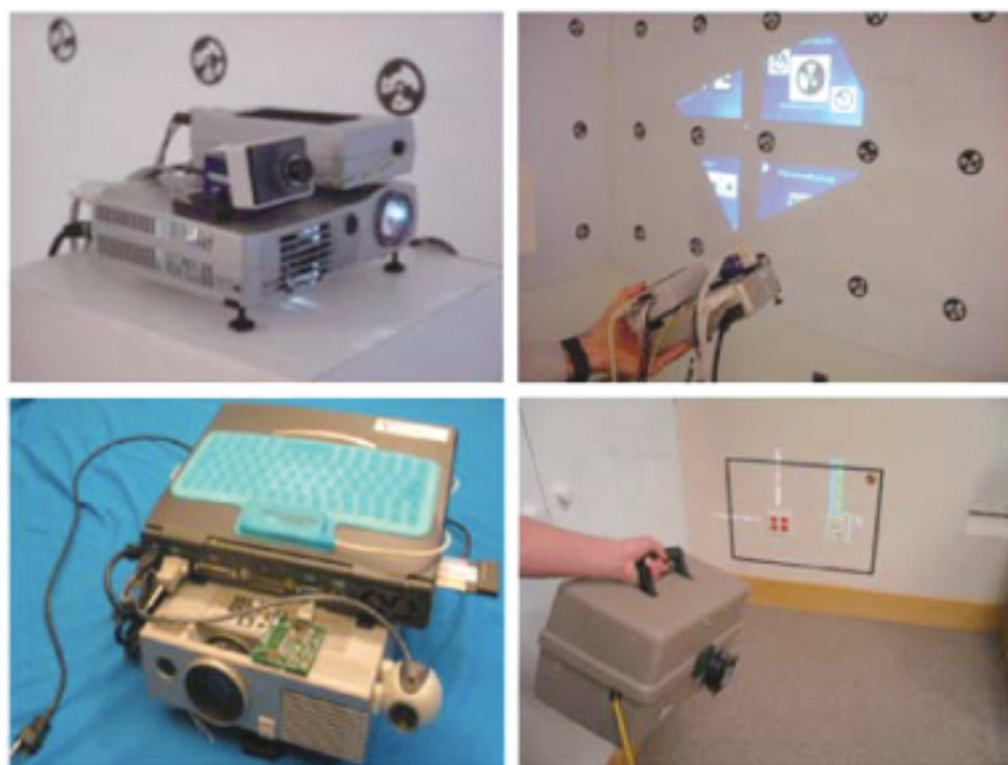


图 2.27 手持式投影设备^[45]

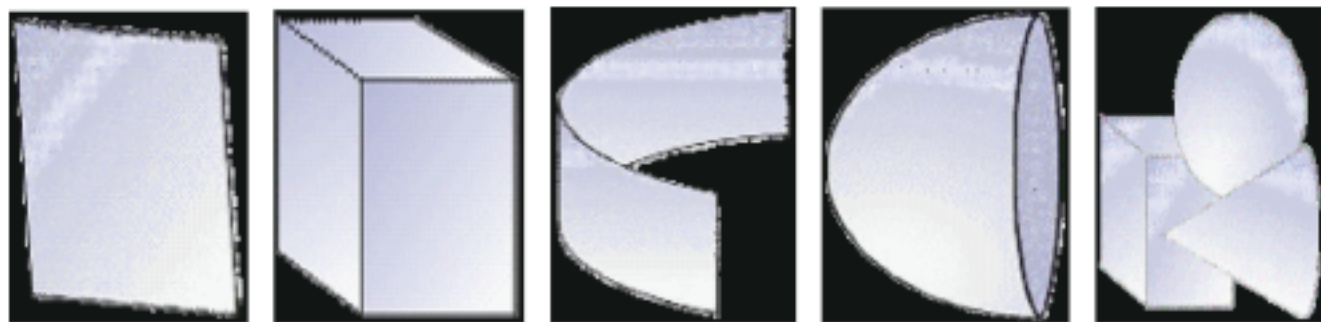


图 2.28 各种类型的投影面(平面、曲面)

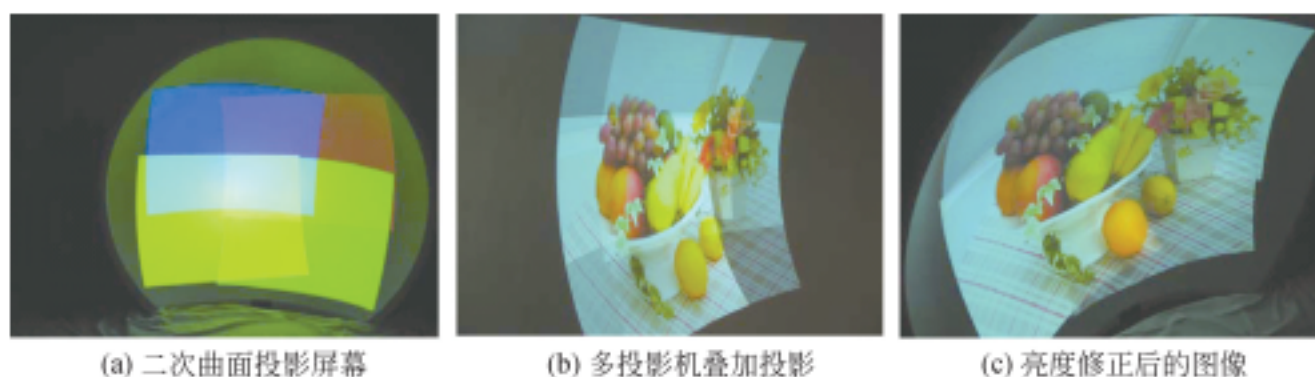


图 2.29 曲面投影^[46]

2.1.6 智能可穿戴显示设备

2012 年 6 月 28 日,谷歌在 I/O 开发者大会上发布了谷歌智能眼镜,并接受开发者现场预订。该产品一经发布就吸引了全世界的目光,被时代杂志评为 2012 年度最佳发明。最终外形已经公布,并于 2014 年向公众开放销售,售价 1500 美元 [图 2.30(a)]。谷歌智能眼镜集智能手机、GPS、照相机于一身,可以在用户眼前展现实时信息(增强现实),用户只要眨眨眼就能拍照并上传、收发短信、查询天气及路况等。它可以让用户在行走过程中,通过发出语音指令享受智能手机的大部分功能。



图 2.30 可穿戴式智能眼镜

Android 可穿戴设备开发商 Vuzix 在 2013 年初也发布了一款类似谷歌智能眼镜的“眼镜”——Smart Glasses M100,如图 2.30(b)所示。Smart Glasses M100 像一个超大的蓝牙耳机,使用 1GHz OMAP4430 处理器、1GB 内存、4GB 存储空间以及 Android 4.0 Ice Cream Sandwich 系统。它可以挂在耳朵上,然后向前伸出并搭载一

个虚拟显示镜片,同时配备一个 720P 的高清摄像头和三轴头部跟踪传感系统,该系统配有陀螺仪、GPS 以及数字罗盘等技术。

美国 SBG 实验室研制的 iShades 5G 将激光二极管微型投影仪内置于镜腿中,投射出高度会聚的光束到有 RGB 三个通道的全息光学元件上并最终进入人眼,其产品效果图如图 2.30(e)所示。目前 SBG 智能眼镜还处于样机阶段,主要面向军事和航空应用,最后面向民用市场。美国的 Syndiant 公司也发布了 ViewLink 系列最新的 Vizcom WiFi 云终端近眼可视通信系统,其效果如图 2.30(d)所示,含有一个近眼显示器、720P 的视频相机以及 Android 智能控制器,Vizcom 允许显示内容直接通过内置的 WiFi 或者 3G/4G 无线智能电话传输到云终端。

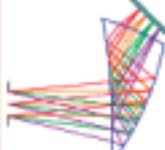
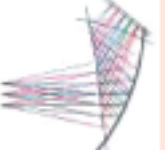
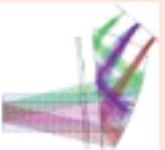
2.2 头盔显示器的分类及其核心技术

头盔显示器按技术分类可分为自由曲面头盔显示器、波导型头盔显示器以及投影式头盔显示器等。

2.2.1 自由曲面头盔显示器

在 20 世纪 90 年代,飞利浦、任天堂、索尼、佳能等一些大公司都曾试图向市场推广个人头盔显示器,但均没能获得成功。失败的原因主要包括视觉遮拦,移动安全性差,设计不够小巧,外形不够美观,价格偏高,分辨率低,像质差等。为了克服这些问题,研究人员不断地把最新的光学技术应用于头盔显示器设计,使其向着高画质、可透视、超轻、超薄的方向发展,使其位于人眼前部分的形状和厚度都尽量接近眼镜。目前,采用的技术包括自由曲面(FFS)光学元件和全息光学元件(HOE)。自由曲面棱镜的应用增加了头盔显示目视光学系统设计的自由度,使系统有效折叠,其长度、体积、重量都大幅下降。表 2.1 给出了部分头盔显示器的设计参数。

表 2.1 部分头盔显示器的参数^[51]

模型名称	Eye-Trek FMD 220	Z800 3Dvisor	i-Visor	ProView SL40	北理工设计结果
模型视图					

续表

模型名称	Eye-Trek FMD 220	Z800 3Dvisor	i-Visor	ProView SL40	北理工设计结果
视场角/(°)	37	39.5	42	40	53.5
出瞳距离/mm	23	27	22	30	18.25
出瞳直径/mm	4	4	3	5	8
焦距/mm	21	22	26.7	20.6	15
像面尺寸/in	0.55	0.61	0.81 *	0.59	0.61
$F/\#$	5.25	5.5	8.9	4.1	1.875

一些公司已推出相关产品,如美国 eMagin 公司的 Z800^[52]、Vuzix 公司的 VR920^[53]、韩国 Daeyang 公司的 i-Visor^[54]、日本 Olympus 公司的 Eye-Trek^[55]、中国深圳富瑞丰公司的 VMP^[56]。图 2.31 给出了基于自由曲面棱镜的部分头盔显示器产品外形图。



(a) Vuzix公司的VR920^[53]



(b) 富瑞丰的VMP^[56]

图 2.31 基于自由曲面技术的眼镜型头盔

2.2.2 全息波导型头盔显示器

波导型头盔显示器主要包括全息波导型和波导型头盔显示技术。全息光学元件有着优良的消色差特性,可进一步实现超轻、超薄、可透视的头盔目视光学系统的设计目标,是目前国际上的另一个研究热点。SBG 实验室^[57]、Vuzix 公司^[58]、日本的 Sony 公司^[59]已先后推出采用全息光学元件的超薄头盔显示器样机,如图 2.32 所示。

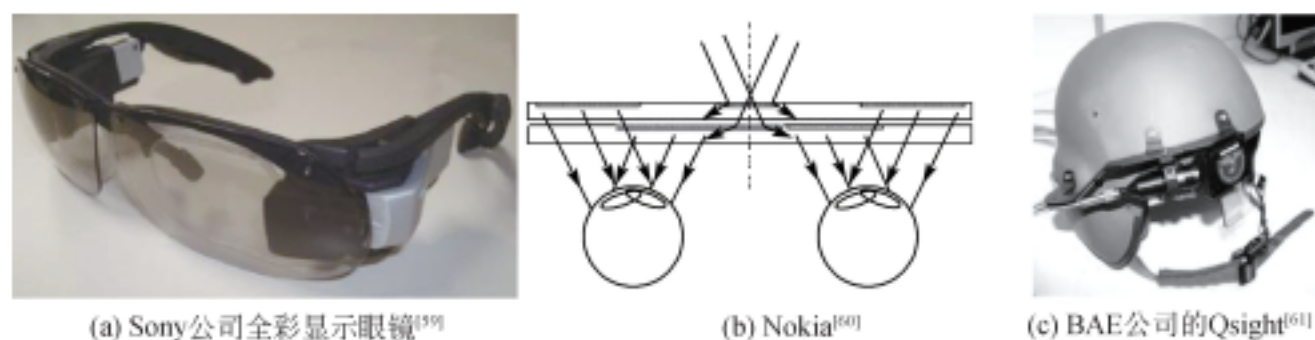
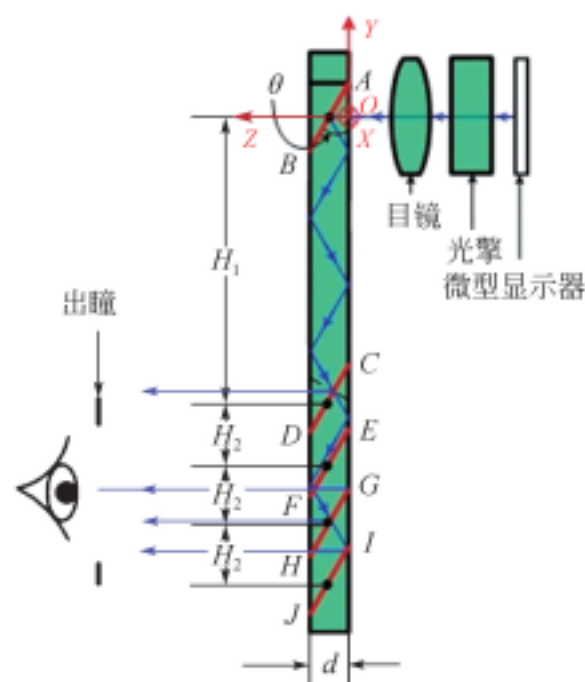
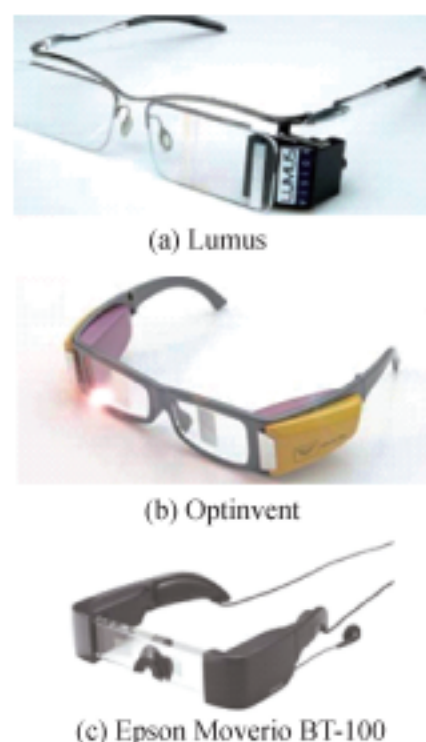


图 2.32 采用全息光学元件的眼镜型头盔

2.2.3 几何波导型头盔显示器

目前,从事几何波导型头盔显示技术研究的公司主要有日本的 Epson、以色列的 Lumus 公司和法国的 Optinvent 公司。图 2.33 给出了几何波导型头盔显示器的工作原理示意图。系统由微型显示器、投影光学系统和平面波导光学元件组成。平面波导光学元件由耦合和耦合出光学部分组成,耦合出光学部分由平面分光镜组成。平面波导光学元件不产生光焦度,放大和准直功能由投影光学系统实现。几何波导型头盔显示器的部分产品外观如图 2.34 所示。

图 2.33 几何波导型头盔显示器工作原理图^[62]图 2.34 几何波导型头盔显示器^[63-65]

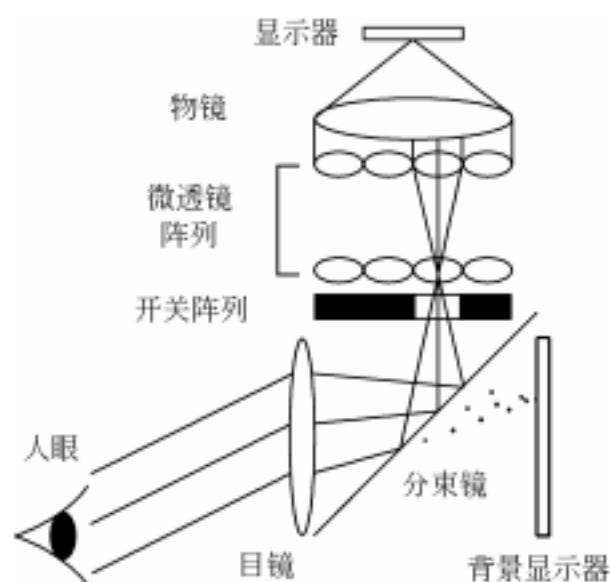
2.2.4 高性能头盔显示技术

为了提高头盔显示的浸没感,必须尽量提高目视光学系统的视场角。同时,头盔佩戴方便性和舒适性的指标又要求其目镜具有大出瞳直径(方便不同瞳距的人使用)和大出瞳距离(方便戴眼镜的人使用)。这种大视场、大相对孔径(出瞳)的目镜设计难度极大,能够达到像质要求的系统往往十分复杂、笨重。北京理工大学在这方面做了大量工作,成功地设计研制出了基于自由曲面棱镜的大视场(水平 45°)、大出瞳(8mm)目镜,每眼光学元件的重量只有 $5\text{g}^{[51]}$ 。

当实际应用要求头盔显示器的视场角进一步增大时,传统的每眼采用单个显示通道的设计方案会出现分辨率下降的问题,影响显示的浸没感效果。因为视场角和分辨率存在 $R=2\theta/N$ 的关系,其中, R 为最小可分辨角度, θ 为半视场角, N 为头盔中使用的微型显示器的像素数。在 N 为一定值时, R 与FOV相互制约,无法同时满足大视场和高分辨率的要求。因此,研制高分辨率大视场头盔显示器必须采用新的解决方法。目前,国际上主要提出了以下四种途径解决上述矛盾^[66,67]。

(1)关注区域高清化(high-resolution area of interest)技术。在大视场范围内显示一幅较低分辨率的背景图像,同时利用眼部跟踪技术获知用户的关注区域,另将小视场的高分辨率图像重叠至该区域,从而使用户在大视场范围看到的一直是高清图像。这种方案的优点是符合人眼视觉特性,即中心区域分辨率高,边缘分辨率较低。缺点是需要快速、低噪声、高精度的眼球跟踪装置,结构比较复杂。此外,背景图像的视场仍然受到前述目镜设计难度的限制,所以这只能是一种部分解决方案。美国中佛罗里达大学提出了一种完全利用固定光电器件实现的插入式头盔显示器模型(opto-electronic high resolution inset HMD, OHRI-HMD),如图2.35所示^[68]。利用微透镜阵列将高分辨率的插入图像复制成 $M \times M$ 的图像阵列,再利用光学开关阵列选通与用户视线位置一致的图像单元,并将其叠加在背景图像的相应位置,从而实现无活动元件的轻小型图像插入机制。这种模型的潜在对准误差小,不存在由扫描元件引起的机械震动,结构紧凑。

(2)双目分视(dichoptic area of interest)技术。将一幅大视场、低分辨率的背景图像显示到用户的一只眼睛,将一幅小视场、高分辨率的图像显示到用户的另一只眼睛。两幅图像通过用户大脑融合,实现大视场、低分辨率的背景图像,中心是用户感兴趣的高分辨率图像。该方案原理简单,成本低廉,但是用户只能在中心区域看到高清晰度的图像。此外,该技术无法实现立体显示。美国Kaiser Electro-Optics (KEO)公司已推出相关产品HiDefTM^[69]。

图 2.35 光电插入式头盔显示器模型^[68]

(3) 双目交叠 (partial binocular overlap) 技术。人的两眼看到虚拟场景中的区域不完全相同,仅中心部分有交叠。由此实现在不降低分辨率和不增加头盔重量的情况下扩大水平视场。该方案原理简单,人的每只眼只需一个显示通道。但是该技术也是只能部分解决视场问题,且要求低畸变目视光学系统,设计装调难度较大。此外,还容易产生融像困难、视疲劳等问题。采用该技术的产品有 KEO 公司的 SIM EYE™ SX100,如图 2.36 所示。

图 2.36 KEO 公司的 SIM EYE™ SX100^[70]

(4) 光学拼接 (optical tiling) 技术。将一系列 (N 个) 小视场、高分辨率的显示单元按特定的方式排列安装在一起,在相接的区域采取部分视场重叠的方式消除缝隙,如图 2.37 所示^[71]。拼接式头盔显示器将视场扩大到近乎原来单个模块的 N 倍,并且在整个视场内的分辨率与单个模块相同。对于视场和分辨率之间的矛盾,这是一个完全解决方案,可以实现真正意义上的高分辨率大视场头盔显示器,非常适用于浸没式虚拟显示环境。但是该技术需要多个显示通道,结构复杂,且拼接装

调相对困难。此外,目前国际上此类研究均采用传统轴对称式光学结构的显示单元,使得整个系统的体积和重量偏大。

图 2.38(a) 中头盔显示系统的视场达到 $82^{\circ} \times 32^{\circ}$, 采用重叠一小部分视场的方法消除过渡区域的渐晕。图 2.38(b) 中, 4 片自由曲面棱镜拼接在一起, 使得视场达到 $82^{\circ} \times 56^{\circ}$ 。图 2.38(c) 中的头盔显示系统采用 6 片棱镜拼接的方式, 视场角可达 $119^{\circ} \times 56^{\circ}$ 。以上几种头盔显示器的分辨率均为 3.2 弧分。

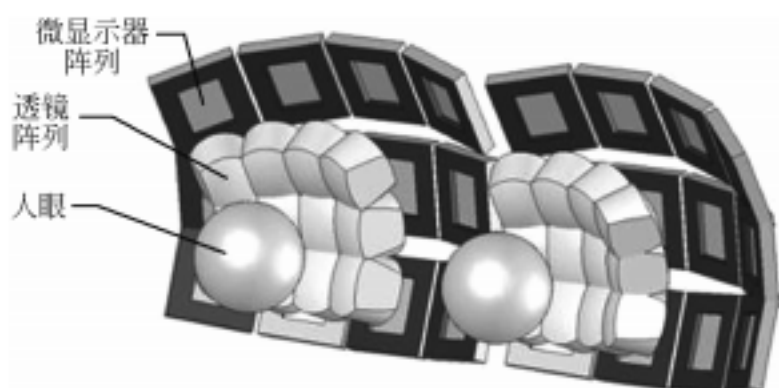


图 2.37 传统光学拼接式头盔显示技术示意图^[71]

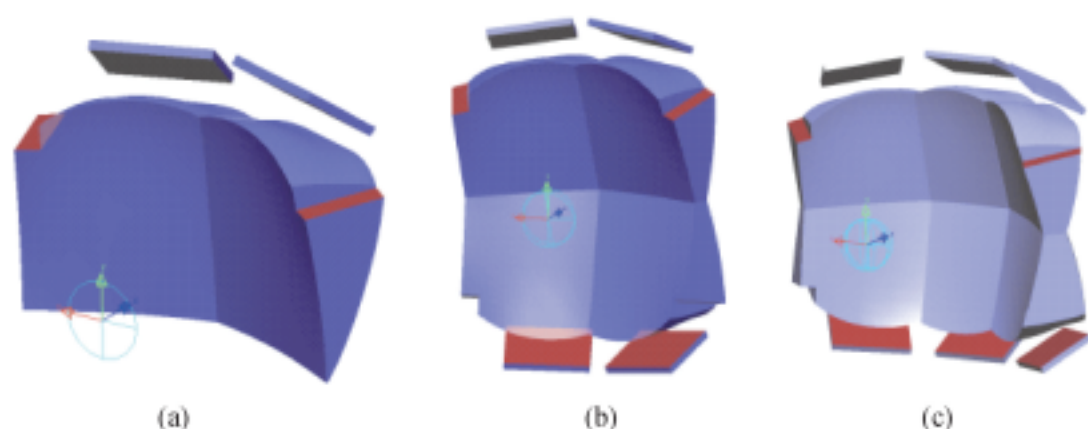


图 2.38 自由曲面拼接式头盔显示系统外形图^[72]

2.2.5 真实立体感头盔显示技术

人眼为了能看清近处和远处的目标,需要调整眼的屈光能力,使远处的平行光和近处的发散光均能在视网膜上清晰地成像,这种屈光能力称为调节 (accommodation)。而当人们要看清近处物体时除了调节,还需要两眼内转,使双眼视轴交于注视点。眼球内转的作用称为集合,也称为辐辏 (convergence)。在眼球内转时,瞳孔还会缩小,从而减少像差,提高视网膜像的质量。调节、集合和瞳孔缩小称为近反射的三联运动。眼的调节与集合在生理方面有着密切的关系,在日常

生活中,调节与集合是相互协同联合运动的。

在自然环境中,人的立体视觉是靠人眼的调节和集合共同产生的。而传统的立体头盔显示器仅有一个固定的焦面,即人眼的调节是固定的,造成其集合与调节不相匹配,与人眼自然视觉特性不符。这会导致视觉失真,人眼很难同时进行聚焦和图像融合,从而造成用户视觉不适和视疲劳。如图 2.39 所示,在真实环境中,人眼的集合和调节是同步的。当人眼聚焦于某一点时,该点前后的物体会发生不同程度模糊[图 2.39(c)],离关注点越远,模糊程度越严重。而在传统立体头盔显示器中,人眼的集合和调节不一致,无论物体离人眼关注点远近,人眼观察到的均是清晰图像[图 2.39(d)]。

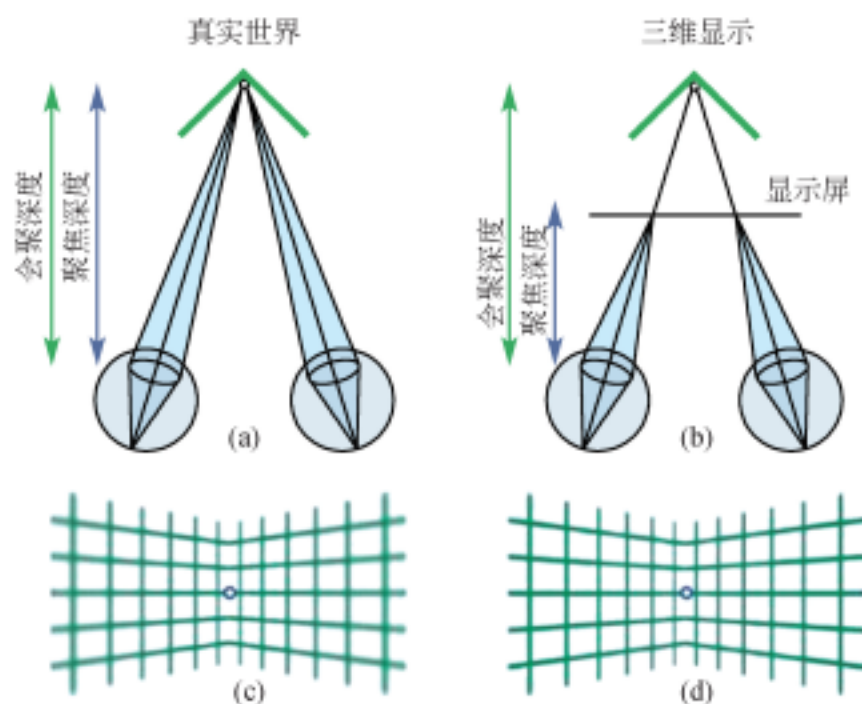


图 2.39 人眼自然视觉和头盔立体视觉^[73]

解决头盔显示器中人眼调节与集合不协调问题的途径之一是发展多焦面显示技术,即通过某种手段生成若干焦面,在不同焦面上渲染不同景深的虚拟物体。不同焦面的场景相互融合形成有立体感的三维图像,有效地改变人眼的调节距离,产生具有真实感的立体视觉。其中的关键技术是要求光学系统能够生成多个焦面。目前,国际上主要提出以下三种方法。

(1) 液体透镜调节焦面^[74]。液体透镜调节焦面技术是在系统中引入液体镜头调节光学系统的光焦度,进而改变系统的焦面位置,并通过分时复用生成若干焦平面(图 2.40)。由于受到人眼最小刷新频率、液体镜头、微型显示器刷新频率的限制,一个微型显示器和液体镜头可以生成 2 个或 3 个焦平面。如果在系统中引入

N 个分光镜、液体镜头和微型显示器,则能产生 2^N 个焦平面。

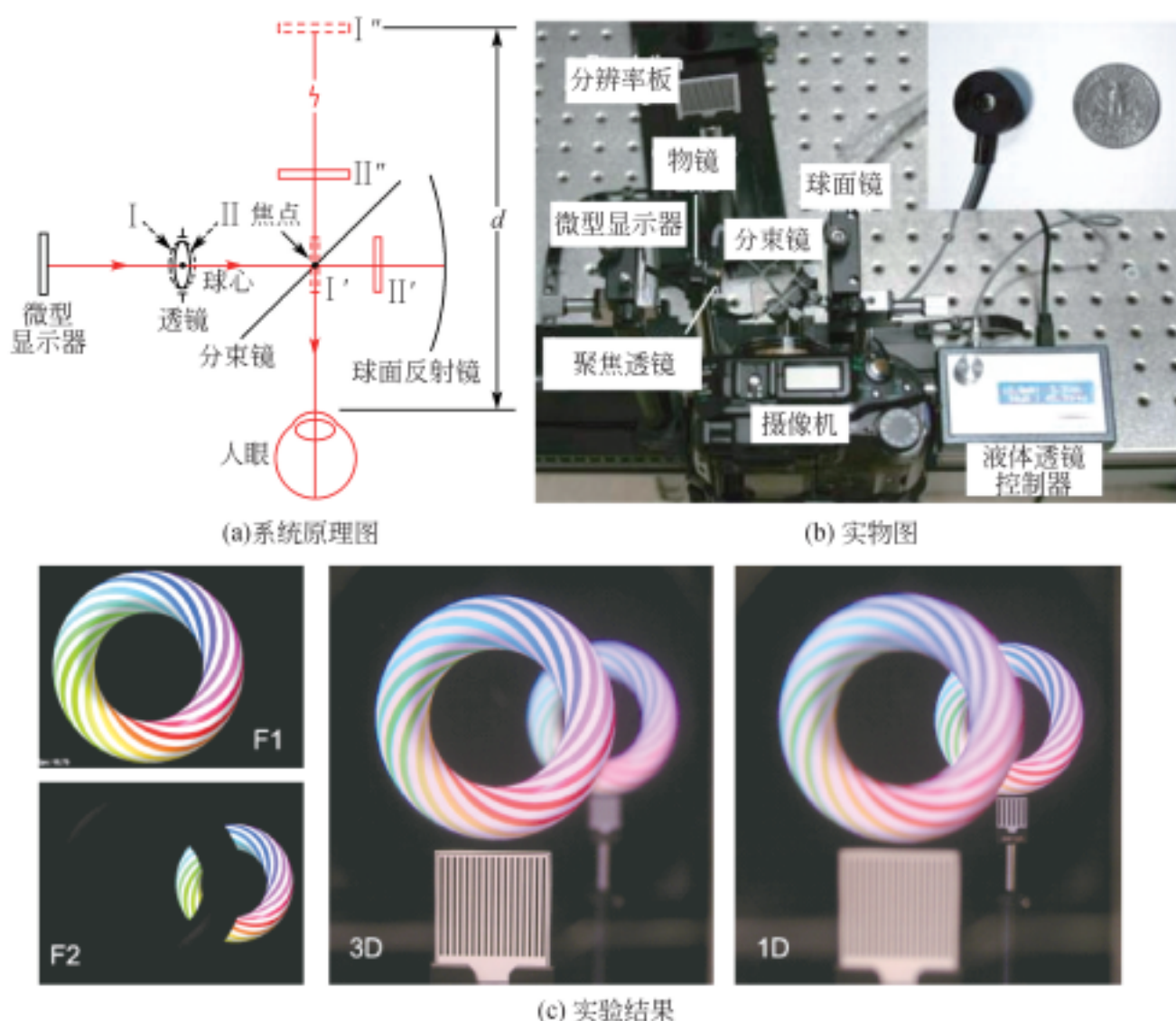
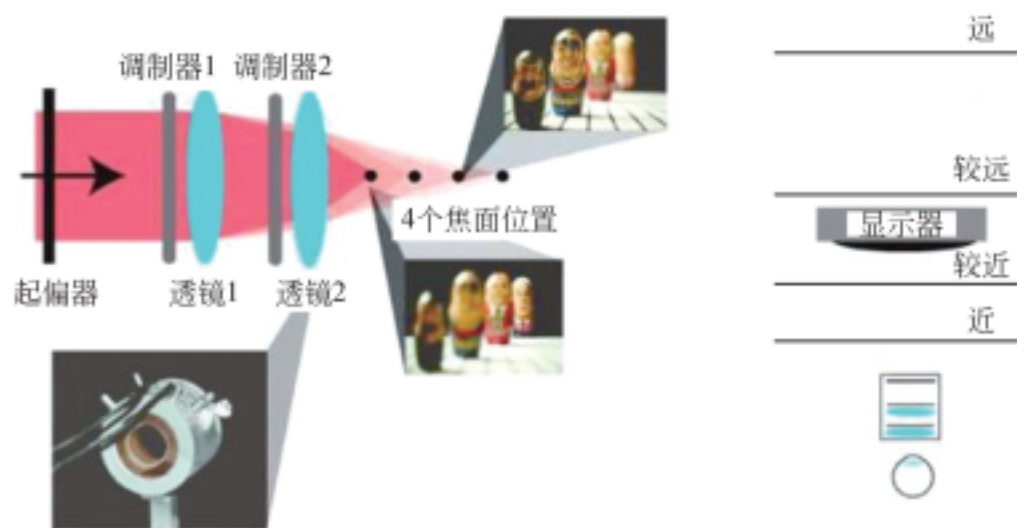


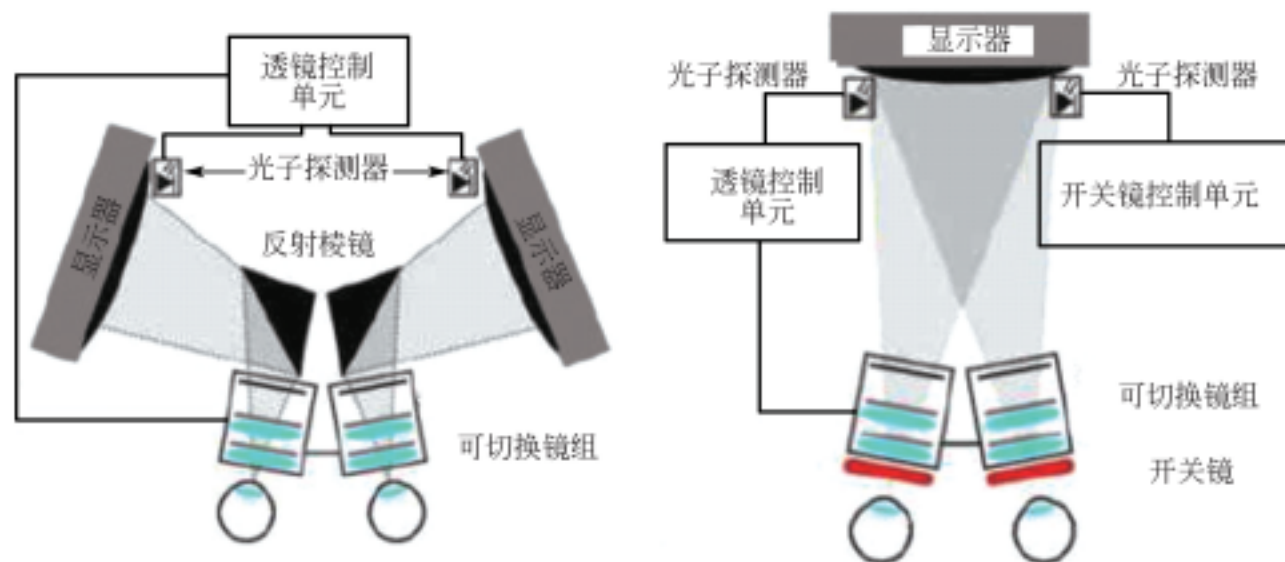
图 2.40 可变焦面光学透射式头盔显示器^[74]

(2) 用双折射玻璃材料和偏振片产生多焦面^[75]。此类系统由一个固定的双折射透镜和偏振片组成。双折射材料在寻常光(o 光)和非常光(e 光)方向有两个不同的折射率值,因此,在双折射材料透镜前加液晶调制器调整入射光的偏振状态就可以得到两个不同的焦平面。如果引入 N 个液晶调制器和双折射透镜组合,系统就能生成 2^N 个焦面。图 2.41(a)所示系统由一个起偏器、两个双折射率透镜和两个液晶调制器组成,一共可生成 4 个焦面。图 2.41(b)所示立体显示系统结构每眼配备一个显示器,该系统刷新频率为 45Hz。图 2.41(c)所示结构双目共享一个显示器,图像以分时复用的方法显示到人眼,刷新频率为 22.5Hz。

(3) 通过分光镜引入多光路^[76]。通过半反半透镜在系统中引入多条光路,每条光路有独立的微型显示器,即对应一个焦面。图 2.42 所示的系统使用了三个分



(a) 由双折射材料透镜和液晶调制器构成的可高速切换焦面成像系统



(b) 由两个显示器组成的立体显示系统

(c) 由一个显示器组成的立体显示系统

图 2.41 基于双折射玻璃材料和偏振片的真实立体感头盔示意图

光镜将系统分成三路,使系统有三个焦平面:近焦面、中焦面和远焦面。通过图形渲染技术生成位于近物面和远物面之间的虚拟物体。图 2.43 所示为具有三个焦平面多光路头盔显示器显示效果图。

1. 新型立体感显示技术

此外,还有许多新型的立体感显示技术,例如,采用变形镜技术改变光学系统的光焦度,进而改变系统的显示深度(图 2.44),或是采用扫描镜阵列的真实立体感显示技术等。

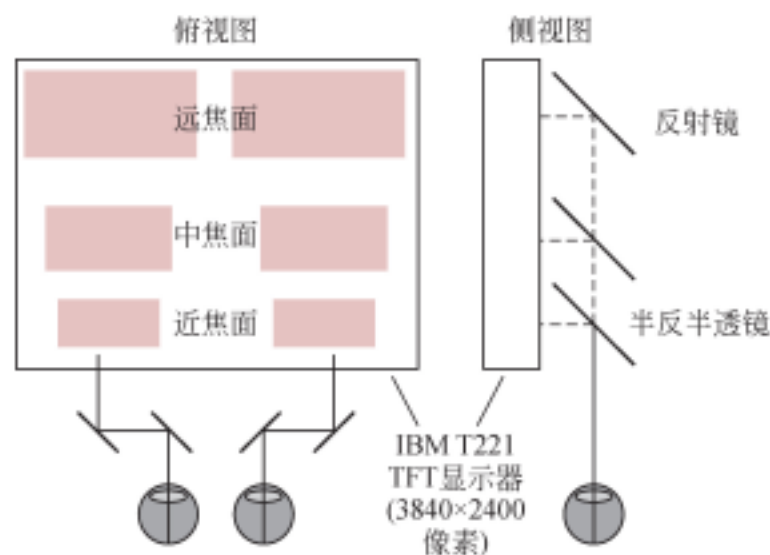


图 2.42 具有三个焦平面多光路头盔显示器结构示意图^[76]



图 2.43 具有三个焦平面多光路头盔显示器显示效果图^[76]

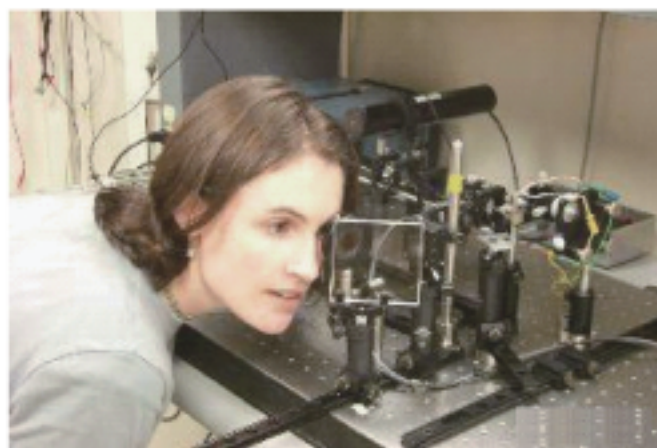
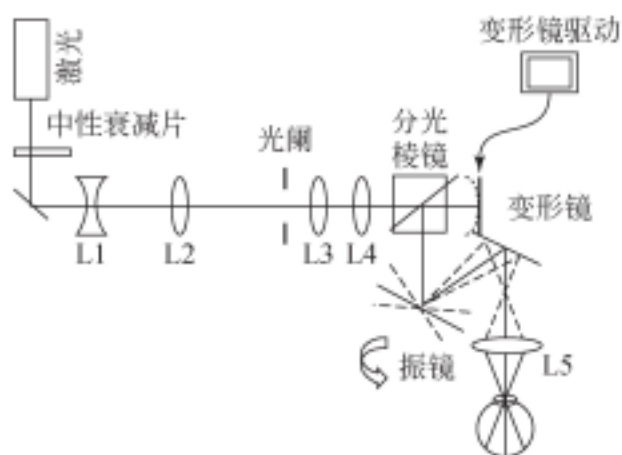


图 2.44 采用变形镜的真实立体感显示技术^[77]

2. 多维感知头盔系统

到目前为止,绝大部分头盔显示器只能输出图像和声音信息。绝大多数虚拟现实系统只能为用户提供视觉和听觉感知,少部分通过力反馈设备、带振动功能的

数据手套、喷水设施、摇动座椅等引发用户的力觉、触觉和运动感知。随着高端应用对虚拟环境沉浸感要求的不断提高,需要在未来的虚拟现实系统中引入更多的感知,如味觉、嗅觉、触觉,以满足用户多方位的感觉享受。在英国工程和物理科学研究委员会的专款支持下,约克大学和华威大学的科学家正在研制一款能够模拟视觉、听觉、嗅觉、味觉、触觉等多种感知的虚拟现实头盔系统——虚拟茧^[78],如图2.45所示。为达到多感同时兼备,虚拟茧较之传统头盔显示器在多方面进行了改进,它的图像由计算机无线传输,采用基于LED和LCD技术的高清晰、高动态显示;听觉由环绕声扬声器提供;嗅觉采用连接到装有化学物质盒子的嗅觉管在用户鼻子下方释放气味;味觉希望通过向用户口内喷入有质感和味道的物质实现;触觉感知由专用装置模拟;温度和湿度则通过风扇和加热器调节控制。目前的实验可以模拟扬帆出海,用户可以看到风帆,听到海浪,感受到海风扑面而来,并被海水打湿脸颊。多维感知头盔显示器在训练培训、虚拟旅游、历史重现、远程会议、虚拟角色扮演游戏等方面具有广阔的应用前景。



图2.45 虚拟茧样机^[78]

3. 无遮拦增强现实头盔显示技术

增强现实头盔存在的问题是叠加的虚拟物体无法遮住真实场景。例如,需要修改图2.46(a)所示真实场景中的一部分,叠加图2.46(b)所示的大楼,图2.46(c)为理想的叠加显示效果,通过普通光学透射式头盔显示器观看的效果如图2.46(d)所示,叠加的场景无法遮住位于它后面的真实景物,出现漂浮的现象^[79]。为此,科学家提出了具有遮挡关系的光学透视式头盔显示方案。

Sony公司于1992年提出了互相遮拦增强现实的光学理念^[80],如图2.47所示。随后,Robinson公布了一个概念系统^[81]。1999年Tatham发表文章,公布了如何在不使用成像光学元件的情况下仅使用传输光线阻挡阵列所获得的结果^[82],他

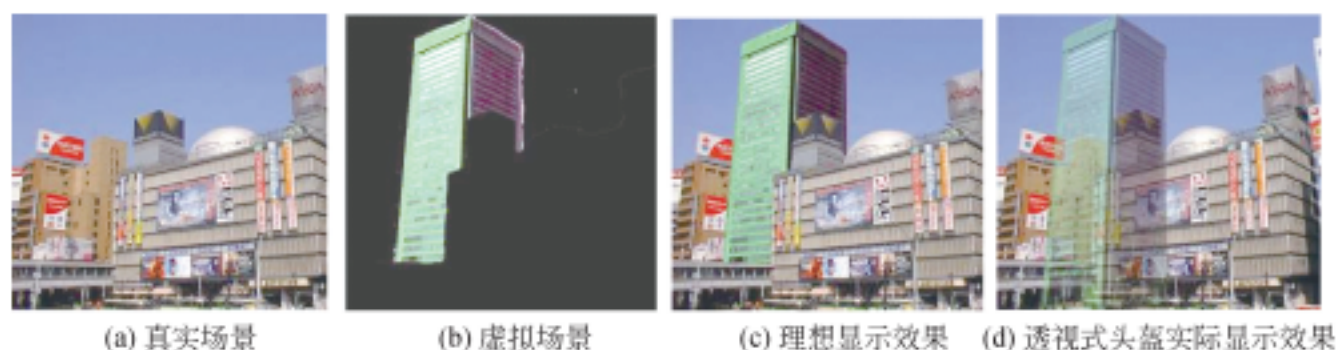


图 2.46 增强现实头盔中的遮拦问题^[79]

提出动态模板调制景物内容并和显示融合在一起。Tatham 更进一步提出在传输模板中使用数字微型镜面装置 (DMD) 的潜在优势,但是没有提出相应的光学设计^[83]。2000 年,Kiyokawa 等在他们的原型机 ELMO 中提出了遮拦的问题^[79],现在该样机已发展到第 4 代。2004 年,Cakmakei 等设计了样机^[83]。

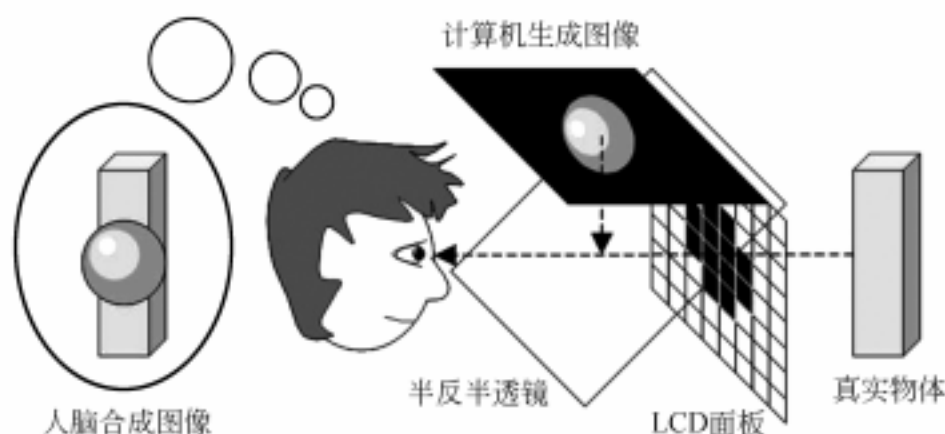


图 2.47 互相遮拦增强现实的光学理念示意图^[79]

遮挡合成图像中的真实景物有四种方法:①遮挡光源与真实景物之间的光线;②通过设置特殊物体来遮挡光线;③遮挡实物与用户眼睛之间的光线;④通过增加合成图像的亮度减少真实场景的可见性。其中,较常用的为第三种方法。

图 2.48 所示为使用上述第三种方法的光学系统设计的一个基本思路。该系统的设计核心就是,在普通透射式显示器之前放置一块 LCD 面板,以阻挡外界光线。但是仅如此还不够,因为 LCD 面板离眼睛太近,当用户看外界景物时,LCD 面板上的图样会出现离焦。可将该 LCD 面板定位在前后两个凸透镜的一倍焦距内,形成一个放大率一倍的光学望远镜设计。最后,使用一个正像棱镜颠倒外界景物,观看者可以同时观察外界景物和 LCD 面板上的图样。通过打开 LCD 面板上真实景物应该出现的像素点,关闭 LCD 面板上对应的虚拟景物应该出现的像素点,真

实景物和虚拟景物的相互遮拦才能真正地做到光学呈现。该方法的三大优势如下。

(1) 实用性: 由于显示器不影响真实环境, 也不需要任何额外的环境设置, 它可以在任何地方任何情况下使用, 包括室外。

(2) 色彩保真度: 由于显示器能够阻挡任何外来景物光线, 在任何情况下, 虚拟图像都保持原来的颜色, 并优于普通光学透视式显示器的保真度。

(3) 相容性: 由于挡光部件与彩色图形显示器分离, 它可以用于大多数现有的光学透视显示器。

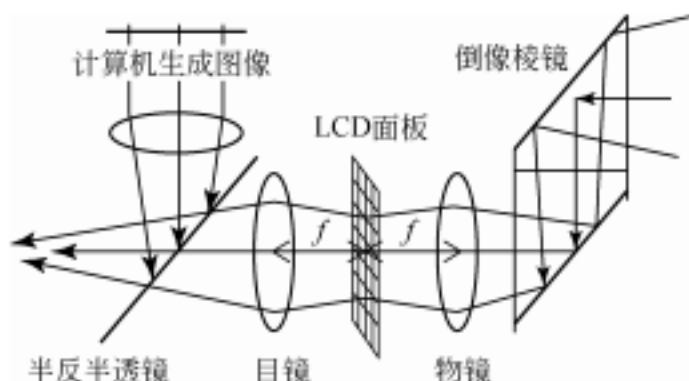


图 2.48 遮挡实物与用户眼睛之间的光线方法的光学系统设计^[79]

2.3 增强现实头盔显示器的设计要点

2.3.1 头盔显示器的组成与原理

头盔显示器一般由以下几部分组成: 图像信息显示源、图像成像的光学系统、定位传感系统、电路控制及连接系统、头盔及配重装置。图像信息显示源是指图像信息显示器件, 一般采用微型高分辨率 CRT 或 LCD、EL、VFD、LED、FED、PDP、OLED 和 LCOS 等平板显示器件。定位传感系统既包括头部的定位, 又包括眼球的定位, 其中, 眼球的定位主要应用在瞄准系统上, 一般采用红外图像的识别跟踪来获得眼球的运动信息; 而头部定位采用的方法比较多, 如超声波、磁、红外、发光二极管等器件的定位系统, 头部定位系统提供用户头部的位置和方向六自由度信息。整体来说, 系统对定位传感系统的要求是灵敏度高、延迟小。电路控制及连接系统一般与头盔显示器分开放置, 以减轻头盔重量。头盔是显示器的固定部件, 在机载或车载应用时可直接把驾驶员头盔进行改进制作。在其他应用中, 可采用吊带将显示器固定在头部。由于显示器的重量在头的前部, 头部的重心发生了变化, 容易

发生疲劳。因此,应在头后部加配重,以保持重心不变。但头盔的重量应尽可能轻,特别是在机载情况下,头盔的重量在飞行员脱离飞机的弹射时将产生很大的冲击力,容易使飞行员颈椎骨折。

2.3.2 头盔显示器的光学系统设计要点

光学系统的设计不仅关系到成像质量的好坏,还影响到头盔显示器的体积和重量。在 HMD 中,中继光学系统将图像源成像在目镜系统的焦平面上,再经过目镜系统转变成平行光,经使用者眼前的光学组合玻璃将其投射到眼中。在头盔显示器的设计过程中,需要对下述像差进行校正。

(1) 色差。色差的产生是因为在头盔显示器系统中,所有透镜元件对不同颜色的光有不同的折射率,通过同一个透镜的不同颜色光的偏折量有所不同。为了补偿色差,减小由透镜元件引入的彩虹效应,设计过程中应该尽量采用具有不同色散值和折射率的正负透镜对(往往是胶合的),即消色差透镜。此外,其他消色差的方法还有使用窄带光源或者磷光剂;使用光谱滤波器来限制显示器的光谱范围;使用二向色性的组合元件或者分光计来反射特定光波。

(2) 球差。为了缩减加工费用,便于对透镜表面的精度和质量进行优化,大多数光学透镜的前后表面都采用球面。为了在不增加透镜重量的情况下,尽可能增大透镜的折射效果,透镜前后表面的形状应该是相似的,而弯曲方向相反。然而,这种两个表面都向外凸或者向内凹的结构形式就导致不同入射角的同一频率的光波在经过透镜后发生不同的偏折,即入射角越大,折射程度越大,这就是球差产生的原因。在对视场中心的分辨率测试图进行观察时,球差的存在导致在目镜焦点确定的情况下,观察者很难在轴向区域和离轴区域两个位置同时观察到清晰的图像。

为了减小球面透镜的球差,在设计的过程中可以使用一组消色差透镜,同时适当地改变它们的曲率。另一种相对简单的方法是使用非球面透镜。与球面透镜不同,非球面透镜的表面是一个由球面演变而来与之有相似形状的曲面,这样的曲面可以降低透镜边缘对同一频率光线的偏折程度,达到减小球差的目的。不过,受到当前精密光学加工工艺的限制,非球面光学元件的造价都比较高,除非需要大规模生产(如 Polaroid Land™ 摄像机)或是有极其特殊需要(如哈勃望远镜)的情况下,一般不会使用到非球面。

(3) 畸变。头盔显示器中的图像源成像系统(包括组合元件)在理想的情况下,能够把显示器的图像毫无变形地如实提供给观察者,但是实际的光学系统总会

存在畸变。一般的光学畸变包括枕形畸变、桶形畸变、梯形畸变等。其中,枕形和桶形畸变在共轴光学系统中比较常见,梯形畸变总存在于离轴系统中。在头盔显示器系统中,既可以对畸变进行光学补偿(针对光学系统),又可以通过电学的方法来进行校正(显示器等元件)。电子校正是针对处理模拟信号的阴极射线管显示器的,并且不会在信号的处理过程中引入延迟,但是会降低显示器的分辨率或者使分辨率在显示器上的分布不均匀。数字校正既可以应用在阴极射线管显示器上,又可以应用在液晶显示器上,但是数字校正方法会导致显示器的图像延迟。

(4)场曲。场曲导致显示器在中心和边缘的折射能力发生变化,这种效果与球差相似,只是显示器的中心和边缘有不同的焦距。对于不同的目镜交点,中心和边缘不能够同时呈现清晰的图像。场曲可以通过添加透镜或者弯曲显示器表面来进行补偿。

以上介绍了对头盔显示器成像质量影响较大的几种像差及其校正方法,在对头盔显示器进行光学设计的过程中,为了达到校正像差的目的,往往不得不在头盔显示器中添加更多元件,这使得光学系统中采用了数量众多的光学元件,不仅增加了HMD的整体重量,而且使头盔内部的空间结构也显得较为狭小,使用者佩戴这种头盔时会感觉笨重而且不舒服。其改进方式有两种:从HMD光学系统的传统结构入手进行改进和从光学系统使用的光学元件入手进行改进。主要实现方法是,利用头盔护目镜作为HMD的光学系统准直组合元件。例如,可利用自由面棱镜改进HMD光学系统的结构和采用全息光学元件、二元光学元件(BOE)。其中全息光学元件由于具有重量轻、致密,且比传统光学元件能更好地满足头盔显示器机械结构设计要求等优点,在新型头盔显示器中得到了广泛应用^[84]。

由于头盔显示器佩戴在用户头部,设计时不仅要满足一般光学显示仪器的要求,还要更多地考虑到人体因素,具体有以下几方面。

(1)视场。一般来说,人的裸眼可见水平视场为 200° 、垂直视场为 100° ,但人眼主要对中心 20° 视场敏感,所以在设计目视光学系统时应保证中心 20° 的像质。需要指出的是,由于图像源大多数为LCD或CRT,其输出图像宽高比为4:3的传统TV图像,所以水平视场与垂直视场的比值都约为4:3(实际比值应为角度的正切的比值)。因此,头盔显示器的光学设计要求保证水平视场与垂直视场的比值。

(2)重量。头盔显示器在使用中将长时间佩戴在用户头部,这就要求头盔的光学与机械结构紧凑、重量轻,这可以通过采用新型材料、用LCD代替CRT、使用非球面等方法实现。同时,若用户头部受力不均衡,可能产生头晕、眩目、恶心等不适症状。因此,要求设计时使整个系统的重心尽量靠近头顶,减小使头部失衡的力

矩。由于系统显示部分大都位于头的前部,为了平衡力矩,一般都将用于固定显示部分的机械结构的重心后移。各种头盔显示系统由于其功能和结构不同,重量也不相同,一般在 0.25 ~ 3.0kg 范围内。

(3)分辨率。人眼最小分辨率为 1",大约为 0.05 毫弧度。而头盔显示器的分辨率取决于图像源的分辨率和光学系统的像差,目前的主要矛盾在于图像源的分辨率在设计视场下产生的理想角分辨率不高。当像面位于无穷远时,头盔显示器的角分辨率可由以下公式决定

$$\theta \approx \frac{2 \tan \frac{\omega}{2}}{m} = \frac{2 \tan \frac{\nu}{2}}{n} (\text{rad})$$

式中, ω 、 ν 分别为垂直和水平方向的全视场角; m 、 n 分别是图像源有效显示面(方形 4:3)上垂直和水平方向的像素数。例如:1-O Display Systems 公司生产的头盔显示器 i-glasses X2 的视场为 $24^\circ \times 18^\circ$,其 LCD 有效显示的分辨率为 280×210 像素。它对应的角分辨率计算公式为

$$\theta \approx \frac{2 \tan \frac{\omega}{2}}{m} = \frac{2 \tan \frac{24}{2}}{700} = 6 \times 10^{-4} (\text{rad}) = 0.6 (\text{mrad})$$

(4)出瞳距。出瞳距是指光学系统的边缘到人眼瞳孔的最小距离。为了保证使用者佩戴方便,出瞳距不应过小,应大于 15mm。如果允许用户在使用时佩戴眼镜,则要求出瞳距大于 25mm。一般军用光学系统要求光学出瞳距大于 25mm。以美军 AH-64A Apache 直升机用头盔显示器 IHADSS 为例,其出瞳距为 30mm,这是为了配合飞行员佩戴护目镜,以便于驾驶。

(5)出瞳大小。人眼瞳孔正常状态下直径为 2mm 左右,在黑暗环境下将适当放大。为了允许人眼眼球有一定范围的移动,一般要求头盔显示器的光学系统出瞳直径大小在 8mm 以上。对于要求较高的军用光学系统,要求出瞳大小为 $10\text{mm} \times 15\text{mm}$ 。需要指出的是,出瞳尺寸越大,光学系统的像差校正越困难。在光学设计时,只能设法找到一个平衡值。

(6)瞳距。在设计双目头盔系统时,还应考虑光学系统的瞳距问题。大多数双目显示头盔由两部分对称的光学系统组成,它们各自出瞳的主光线的距离为此头盔显示系统的瞳距。人眼的出瞳一般在 54 ~ 70mm 范围内,双目头盔的瞳距应能良好地配合使用者的双目瞳距。由于各个用户的双目瞳距不等,在设计双目用头盔时要考虑瞳距的可调节功能(如 Virtual Research 公司生产的 V8 头盔显示器)。也有些双目头盔显示器为了使系统结构简单,而使得瞳距不可调,设计时使用一个平均值,

如 62mm(如 I-O Display Systems 公司生产的 i-glasses X2 头盔显示器)。

(7) 光能利用率。光能利用率为到达人眼的光能与入射光能量的比值,它反映了头盔显示器的光学系统对光的吸收或反射,也就是光在到达人眼前的损耗部分。设计时应考虑尽可能增大光能利用率,以减少光能的损耗。对于传统的 VR 用的头盔显示器,只需要考虑图像源发出的光在通过光学系统时的光能损失。而对于增强现实用的穿透式双通道 HMD,设计时要同时照顾到两个通道的光能利用率,即图像源发出的光在通过光学系统后的光能损失与外界光经过光学系统投射后的光能损失。

以上七点为设计头盔显示器的一些重要指标,与其他光学系统一样,这些因素之间并不孤立,它们之间存在联系并相互制约。例如,增加视场或出瞳,光学元件将增大,整体重量增加;视场增加,角分辨率下降。所以在进行光学设计时,应根据实际运用找到一个较好的平衡点。

2.3.3 部分国内外头盔显示器设计

(1) 图 2.49 所示为美国 Kaiser 公司生产的 Proview60 头盔。表 2.2 所示为 Proview 系列头盔的参数比较。



图 2.49 Proview60 头盔^[85]

表 2.2 Proview 系列头盔参数比较

技术参数\型号	PV30	PV40ST/PV50ST	XL40STM/XL50STM	PV60	XL35/XL50
视场角/(°)	30	40/50	40/50	60	35/50
分辨率/弧分	2.25	3.4	2.2	4.5	1.6/2.3
重叠率/%	100	100/24	100/25	100	100

续表

技术参数\型号	PV30	PV40ST/PV50ST	XL40STM/XL50STm	PV60	XL35/XL50
亮度/ cd	25	>6	20	25	5 ~ 50
颜色	全彩	全彩	单色	全彩	全彩
对比度	25 : 1	20 : 1	25 : 1	25 : 1	40 : 1
通过率	遮蔽	24%	24%	遮蔽	遮蔽
光学性能	成型的	反射的	反射的	成型的	成型的
视觉调节	50mm	30mm	30mm	25mm	25mm
质量/oz	25	28	28	28	35

(2) 美国中佛罗里达大学光电学院 (School of Optics/CREOL) 开发的离轴式头盔显示器的具体参数如下, 离轴光学系统设计图见图 2.50。

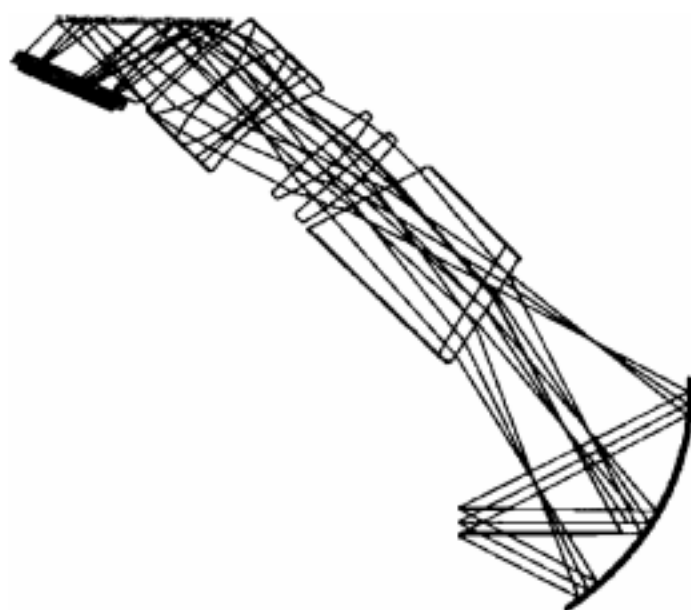


图 2.50 离轴式设计^[86]

有效显示区域: 25.4mm(直径)。

分辨率: 640×480 像素。

颜色: 采用泰克公司的色盘。

长度: 101.6mm。

目镜有关参数如下。

有效焦距: 23.7mm。

出瞳直径: 10mm。

物距: 500mm (距离入瞳)。

眼睛间隙:大于 17mm (已经达到 18mm)。

其他系统参数如下。

波长范围:597 ~ 485nm。

视场:60°。

渐晕:全视场 10%。

像质有关参数如下。

畸变:电子矫正。

像质:全视场点列图均方根直径 25 μ m;20lp/mm 处 MTF 大于 20%。

(3)南开大学现代光学研究所光电信息技术科学教育部重点实验室和长春光学精密机械与物理研究所应用光学国家重点实验室联合开发了折/衍混合增强现实头盔显示器光学系统^[87],该光学系统的具体设计参数如下。

出瞳距离:26mm。

出瞳直径:12mm。

视场角为:20°(H)×15.4°(V)。

内外两个光通道的光能量利用率:1/4 和 1/2。

图像源分辨率:800×600 像素。

系统镜头直径:小于 46mm。

(4)北京理工大学的自由曲面光学透射式头盔显示器。“十一五”期间北京理工大学光电学院成功研制了基于自由曲面的离轴式结构头盔显示器,表 2.3 列出了该头盔显示器的光学特性参数,选用的微型图像源为 eMagin 公司的有机发光二极管(organic light-emitting diode, OLED),显示区域为 0.61in,分辨率为高级视频图形阵列(super video graphics array, SVGA)。与其他类型的微型图像源相比,它具有自发光功能,不需要背光源,因此功耗低,小于 200mW。此外,它还具有厚度薄、对比度高、视角广、刷新频率高等优点,同时它的显示色彩也是最为丰富和逼真的。为了满足一定的观察视场,要求系统的视场角大于 50°,这就要求焦距不能大于 16.6mm,因此,将焦距定为 15mm,对应的视场角为 53.5°,显示分辨率为 3.2 弧分。由于成年人的双目瞳距一般为 54 ~ 72mm,为了让不同用户在不调节瞳距的情况下都能正常使用头盔显示器,要求其目镜光学系统有较大的出瞳直径,同时为了保证系统的结构紧凑和像质优良,将系统的出瞳直径定为 8mm,出瞳距离大于 18mm。

表 2.3 自由曲面头盔显示器的特性参数

特性参数	参数值
OLED	
尺寸/mm	对角 15.5
显示区域/mm ²	12.7×9.0
分辨率/像素	852×600
虚拟成像光路	
结构形式	自由曲面折/反射棱镜
有效焦距/mm	15
出瞳直径/mm	8
出瞳距/mm	>17(18.25)
<i>F</i> /#	1.875
非球面总数	3
透射成像光路	
结构形式	自由曲面透镜
非球面总数	2
其他参数	
波长/nm	656.3 ~ 486.1
视场角	45°(H)×32°V
渐晕系数	0.15 上下边缘

参 考 文 献

- [1] The Cheptics360TM Experience. <http://www.vizoo.com>[2014-9-5].
- [2] 耿征. 真三维高清晰度显示技术——信息显示领域的重要发展方向. 科技导报, 2007, 25(20): 21 ~ 26.
- [3] Favalora G E, Napoli J, Hall D M, et al. 100-million-voxel volumetric display. Proceedings of SPIE 4712, 2002, 4712: 300 ~ 312.
- [4] Langhans K, Bahr D, Bezecny D, et al. FELIX 3D display: An interactive tool for volumetric imaging. Proceedings of SPIE 4660, 2002, 4660: 176 ~ 190.
- [5] Langhans K, Guill C, Rieper E, et al. SOLID FELIX: A static volume 3D-laser display. Proceedings of SPIE, 2003, 5006: 161 ~ 174.
- [6] 林远芳, 刘旭, 刘向东, 等. 基于旋转二维发光二极管阵列的体三维显示系统. 光学学报, 2003, 23(10): 1158 ~ 1162.
- [7] 李莉, 李玉峰, 沈春林, 等. 基于数字微镜的旋转体三维显示装置研究. 仪器仪表学报,

- 2008,29(1):67~72.
- [8]李莉. 体三维显示系统关键技术研究[博士学位论文]. 南京:南京航空航天大学,2008.
- [9]王恒,朱秋东. 反射式多路投影立体显示系统的研究. 光学技术,2008,34(2):178~180.
- [10]朱秋东,王涌天,郝群,等. 投影机旋转屏的体三维显示系统:中国,CN 101038421B. 2010.
- [11]<http://www.vizoo.com>[2013-8-15].
- [12]Nonzen Team. Diesel 全息时装秀. <http://nonzen.com/style/diesel-liquid-space>[2014-9-5].
- [13]Otsuka R, Hoshino T, Horry Y. Transpost:360 deg-viewable three-dimensional displaysystem. Proceedings of the IEEE,2006,94(3):629~635.
- [14]Sutherland I E. The ultimate display//Information Processing 1965,Proceedings of IFIP Congress 65,New York,1965.
- [15]Sutherland I E. A head-mounted three dimensional display//Fall Joint Computer Conference, AFIPS'68 (Fall,part I),San Francisco,1968.
- [16]Buchroeder R, Seeley G, Vukobratovich D. Design of a catadioptric VCASS helmet-mounted display, under contract to U.S. Air Force Armstrong Aerospace Medical Research Laboratory,1981.
- [17]Furness T A. The super cockpit and its human factors challenges. Proceedings of the Human Factors and Ergonomics Society Annual Meeting,1986,30(1):48~52.
- [18]Droessler J G, Rotier D J. Tilted cat helmet-mounted display. Optical Engineering,1990,29(8):849~854.
- [19]Barrette R E. Wide field of view, full-color, high-resolution, helmet-mounted display//SID International Symposium,Boston,1992.
- [20]Kandebo S W. Navy to evaluate Agile Eye helmet-mounted display system. Aviation Week and Space Technology,1988,129(7):94~100.
- [21]ProView™ Head-mounted Display. http://www.digitaltools.de/a_prod_con_sicht_headmounted.htm[2014-9-5].
- [22]Virtualresearch Products. <http://www.virtualresearch.com/products>[2014-9-5].
- [23]Tripathi A. Augmented reality: An application for architecture. http://www.usc.edu/dept-00/dept/architecture/mbs/thesis/anish/thesis_report.htm[2014-9-5].
- [24]Christoph B. HMD/headset/VR-helmet Comparison Chart. <http://www.stereo3d.com/hmd.htm> [2014-9-5].
- [25]5DT HMD 800-26 Manual. http://www.5dt.com/?page_id=36[2010-5-24].
- [26]Cy-Visor. <http://www.5dt.com/products/pcyvisor.html>[2014-9-5].
- [27]Stan M. Sony Concept Video Glasses. <http://www.sonyinsider.com/2009/02/12/sony-concept-video-glasses>[2014-9-5].

- [28] Tech Specs; Designed for Comfort. <http://www.sony.com.au/microsite/hmd/page5.html> [2014-9-5].
- [29] FMD-200 Face-mounted Display Eye-Trek Series. <http://www.olympus-global.com/en/news/2000a/nr000204fmd200e.jsp> [2014-9-5].
- [30] Launch of New Eye-Trek FMD-700. <http://www.olympus-global.com/en/news/2000a/nr000524fmd700e.jsp> [2014-9-5].
- [31] Ferguson J. Optical system for head mounted display using retro-reflector and method of displaying an image. United States Patent 5621572, 1997.
- [32] Rolland J P, Fuchs H. Optical versus video see-through head-mounted displays in medical visualization. Presence: Teleoperators and Virtual Environments, 2000, 9(3): 287 ~ 309.
- [33] Hua H, Girardot A, Gao C, et al. Engineering of head-mounted projective displays. Applied Optics, 2000, 39(22): 3814 ~ 3824.
- [34] Hua H, Gao C, Biocca F, et al. An ultra-light and compact design and implementation of head-mounted projective displays//IEEE Virtual Reality 2001 Conference, Yokohama, 2001.
- [35] Inami M, Kawakami N, Sekiguchi D, et al. Visuo-haptic display using head-mounted projector//IEEE Virtual Reality 2000 Conference, Washington DC, 2000.
- [36] Zhang R, Hua H. Characterizing polarization management in a p-HMPD system. Applied Optics, 2008, 47(4): 512 ~ 522.
- [37] Hua H, Gao C. Design of a bright polarized head-mounted projection display. Applied Optics, 2007, 46(14): 2600 ~ 2610.
- [38] Hua H, Gao C, Brown D, et al. Using a head-mounted projective display in interactive augmented environments//IEEE and ACM International Symposium on Augmented Reality, New York, 2001.
- [39] Bimber O, Raskar R. Modern approaches to augmented reality//ACM SIGGRAPH 2006 Courses, Boston, 2006.
- [40] Bimber O, Raskar R. Spatial Augmented Reality Merging Real and Virtual Worlds. <http://140.78.90.140/medien/ar/SpatialAR/download.php> [2014-9-5].
- [41] Low K L, Welch G, Lastra A, et al. Life-sized projector-based dioramas//ACM Symposium on Virtual Reality Software and Technology, New York, 2001.
- [42] 赢康案例集锦. <http://www.wincomn.com.cn> [2014-9-5].
- [43] Westaway L. Project Glass is Google's Wearable, Augmented Reality Specs. <http://www.cnet.com/news/project-glass-is-googles-wearable-augmented-reality-specs> [2014-9-5].
- [44] Beardsley P, Van Baar J, Raskar R, et al. Interaction using a handheld projector. Computer Graphics and Applications, 2005, 25(1): 39 ~ 43.
- [45] Raskar R, Van Baar J, Beardsley P, et al. iLamps: Geometrically aware and self-configuring projectors//ACM SIGGRAPH 2006 Courses, Boston, 2006.

- [46] <http://web.media.mit.edu/~raskar/CurvedScreen/curvedScreen.pdf> [2012-10-14].
- [47] <http://www.cnet.com/news/project-glass-is-googles-wearable-augmented-reality-specs/> [2013-5-12].
- [48] Cervantes E. Google Glass competitor Vuzix M100 Now Available for Pre-order at \$1,000. <http://phandroid.com/2013/12/05/vuzix-m100-smart-glass-1000-pre-order/> [2014-9-5].
- [49] Orca S. Get Ready for iShades 5g. <http://hplusmagazine.com/2009/05/15/get-ready-ishades-5g/> [2014-9-5].
- [50] CES 2012 Pico Projector Overview. <http://www.kgutttag.com/2012/01/17/ces-2012-pico-projector-overview/> [2014-9-5].
- [51] Cheng D, Wang Y, Hua H, et al. Design of an optical see-through head-mounted display with a low F-number and large field of view using a free-form prism. *Applied Optics*, 2009, 48(14): 2655 ~ 2668.
- [52] EMagin Z800. <http://www.emagin.com/> [2014-9-5].
- [53] Vuzix VR920. <http://www.vuzix.com/iwear/index.html> [2014-9-5].
- [54] Daeying i-Visor. <http://www.personaldisplay.com/> [2014-9-5].
- [55] Olympus Eye-Trek. <http://www.olympus-global.com/en/news/2001h/nr011010fmd2203.cfm> [2014-9-5].
- [56] VMP. <http://www.free-form.cn/> [2014-9-5].
- [57] Popovich M, Sagan S. *Application Specific Integrated Lenses for Displays*. New York: Blackwell Publishing Ltd, 2000.
- [58] Grifantini K. CES: Clear Glasses to Augment Your Reality. <http://www.technologyreview.com/view/422294/ces-clear-glasses-to-augment-your-reality/> [2014-9-5].
- [59] Mukawa H, Akutsu K, Matsumura I, et al. A full-color eyewear display using planar waveguides with reflection volume holograms. *Journal of the Society for Information Display*, 2009, 17(3): 185 ~ 193.
- [60] Saarikko P. Diffractive exit-pupil expander with a large field of view. *Proceedings of SPIE*, 2008, 7001: 700105-1 ~ 700105-6.
- [61] Cameron A. The application of holographic optical waveguide technology to the Q-Sight family of helmet-mounted displays. *Proceedings of SPIE*, 2009, 7326: 73260H-1 ~ 73260H-11.
- [62] Cheng D, Wang Y, Xu C, et al. Design of an ultra-thin near-eye display with geometrical waveguide and freeform optics. *Optics Express*, 2014, 22(17): 20705 ~ 20719.
- [63] Consumer Products. <http://lumus-optical.com/> [2014-9-5].
- [64] Optinvent. <http://optinvent.com/> [2014-9-5].
- [65] Epson Moverio BT-100. <http://www.epson.com/> [2014-9-5].
- [66] Melzer J E. Overcoming the field-of-view/resolution invariant in head-mounted displays.

- Proceeding of SPIE,1998,3362:284 ~ 293.
- [67] Rolland J P, Hua H. Head-mounted display systems. Encyclopedia of Optical Engineering,2005: 1 ~ 13.
- [68] Rolland J P, Yoshida A, Davis L D, et al. High-resolution inset head-mounted display. Applied Optics,1998,37(19):4183 ~ 4193.
- [69] KEO HiDef. <http://www.rockwellcollins.com>[2014-9-5].
- [70] SimEye SX100 Helmet Mounted Display. http://rockwellcollins.com/sitecore/content/Data/Products/Displays/Soldier_Displays/SimEye_SX100_Helmet_Mounted_Display.aspx [2014-9-5].
- [71] Brown L G, Boger Y S. Applications of the sensics panoramic HMD//SID 2008 International Symposium, Los Angeles,2008.
- [72] Cheng D, Wang Y, Hua H, et al. Design of a wide-angle, lightweight head-mounted display using free-form optics tiling. Optics Letters,2011,36(11):2098 ~ 2100.
- [73] Hoffman D M, Girshick A R, Akeley K, et al. Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. Journal of Vision,2008,8(3):33.
- [74] Liu S, Hua H. Time-multiplexed dual-focal plane head-mounted display with a liquid lens. Optics Letters,2009,34(11):1642 ~ 1644.
- [75] Love G D, Hoffman D M, Hands P J W, et al. High-speed switchable lens enables the development of a volumetric stereoscopic display. Optics Express,2009,17(18):15716 ~ 15725.
- [76] Akeley K, Watt S J, Girshick A R, et al. A stereo display prototype with multiple focal distances. ACM Transactions on Graphics,2004,23(3):804 ~ 813.
- [77] McQuaide S C, Seibel E J, Kelly J P, et al. A retinal scanning display system that produces multiple focal planes with a deformable membrane mirror. Displays,2003,24(2):65 ~ 72.
- [78] Derbyshire D. Revealed:The headset that will mimic all five senses and make the virtual world as convincing as real life. Dailymail,2009.
- [79] Kiyokawa K, Kurata Y, Ohno H. An optical see-through display for mutual occlusion with a real-time stereovision system. Computers & Graphics,2001,25(5):765 ~ 779.
- [80] Kawamura A. Optical Device, Japan Unexamined Patent 1992-1717,1992.
- [81] Robinson I N. Method and apparatus for augmented reality using a see-through head-mounted display, US6037914 A,2000.
- [82] Tatham E. Getting the best of both real and virtual worlds. Communications of the ACM,1999,42(9): 96 ~ 98.
- [83] Cakmakci O, Ha Y, Rolland J P. A compact optical see-through head-worn display with occlusion support// The 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality, Arlington,2004.

- [84] Ando T, Yamasaki K, Okamoto M, et al. Head-mounted Display using a holographic optical element//Photonics West Electronic Imaging, San Jose, 1998.
- [85] ProView™ 60 Head-Mounted Display. <http://www.mindflux.com.au/products/keo/pv60.html#overview> [2014-9-5].
- [86] Rolland J P. Wide-angle, off-axis, see-through head-mounted display. Optical Engineering, 2000, 39(7): 1760 ~ 1767.
- [87] 张慧娟, 王肇圻, 赵秋玲, 等. 折/衍混合增强现实头盔显示器光学系统设计. 光学学报, 2004, 24(1): 121 ~ 124.

第3章 增强现实系统的标定

增强现实系统中为实现虚拟与真实场景的完美结合,计算机产生的虚拟添加信息需通过三维跟踪注册算法与真实场景保持精确的对准关系。如图 3.1 所示,虚拟计算机产生的虚拟 3D 模型信息正确地叠加在真实环境中。为确保虚拟与真实场景的全方位对准,要用到系统中的许多先验知识,这些知识或信息可能以多种方式呈现。而大多数增强现实应用系统都需要以下一条或更多系统信息。

- (1) 真实场景中基准点的三维空间坐标。
- (2) 摄像机的内部参数。
- (3) 用户(或摄像机)相对于真实场景的位置和方向,即摄像机的外部参数。

要获取摄像机的内部参数和外部参数,增强现实系统在初始时刻需要进行标定。目前的增强现实系统大体可分为光学透视式增强现实系统和视频透视式增强现实系统。虽然光学透视式增强现实系统与视频透视式增强现实系统在整体结构上存在较大的差异,其标定方式也不尽相同。但是,在进行系统标定时这两种系统都将涉及世界坐标系、摄像机坐标系、图像平面坐标系间的坐标系变换。因此,下面将简单介绍各种坐标系,并建立这些坐标系之间的转换关系。



图 3.1 高通手机平台上的增强现实软件包^[1]

3.1 摄像机的几何模型和坐标变换

摄像机通过成像透镜将三维场景投影到摄像机的二维像平面,这个投影可用成像变换描述,即摄像机成像模型。为了定量描述光学成像过程,首先定义四个坐

标系:图像坐标系、像素坐标系、摄像机坐标系和世界坐标系。

3.1.1 图像坐标系与像素坐标系

摄像机对现实场景进行采集和数字化,形成数字图像。在数字图像上,经常定义两种坐标系:像素坐标系和图像坐标系。如图3.2所示,像素坐标系 ouv 的原点位于图像的左上角, u 轴和 v 轴分别平行于图像的行和列,其坐标单位是像素(pixel)。因此,每一个像素坐标 (u, v) 是该像素在图像中对应的列数与行数。图像坐标系的 x 轴和 y 轴平行于图像的行和列,其坐标单位一般是毫米。图像坐标系的原点定义在摄像机光轴与图像平面的交点,其在像素坐标系中的坐标为 (u_0, v_0) 。每一个像素在 x 轴和 y 轴方向上的物理尺寸为 Δx 、 Δy ,这两个参数的倒数称为像素尺度系数。像素坐标系与图像坐标系间的变换关系为

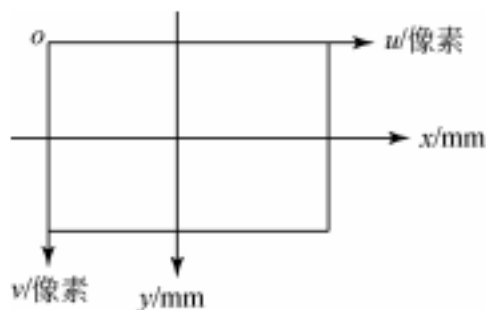


图3.2 像素坐标系与图像坐标系

$$u = \frac{x}{\Delta x} + u_0 \quad (3-1)$$

$$v = \frac{y}{\Delta y} + v_0 \quad (3-2)$$

3.1.2 图像坐标系与摄像机坐标系

摄像机坐标系以图像坐标系中心点为原点,其坐标轴 x_c 和 y_c 分别与图像坐标系的 x 轴和 y 轴平行, z_c 轴为摄像机的光轴。如图3.3所示。由原点 o 与 x_c 、 y_c 、 z_c 轴组成的直角坐标系称为摄像机坐标系。图像坐标系与摄像机坐标系间的转换关系由所采用的摄像机模型决定。若采用针孔摄像机投影模型,则根据透视投影公式,图像坐标系与摄像机坐标系间的转换关系为

$$\mathbf{Z}_c \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} \quad (3-3)$$

事实上,由于实际的镜头并不是理想的透视成像,线性模型不能准确地描述成像几何关系。尤其是在使用广角镜头时,在远离图像中心处会有较大的畸变。描述非线性畸变可用下列公式表示:

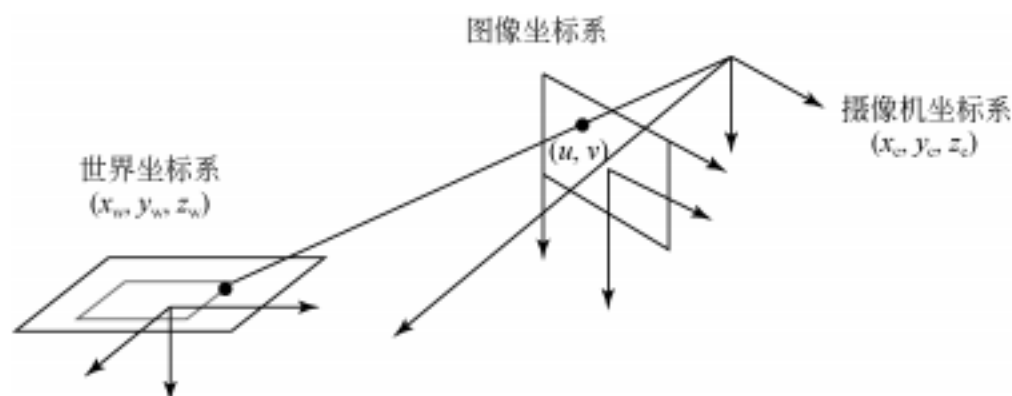


图 3.3 图像坐标系与摄像机坐标系

$$\bar{x} = x + \delta_x(x, y) \quad (3-4)$$

$$\bar{y} = y + \delta_y(x, y) \quad (3-5)$$

式中, $\bar{x}, \bar{y}$ 为由小孔线性模型计算得到的图像点坐标的理想值; (x, y) 是实际的图像点的坐标; δ_x 与 δ_y 是非线性畸变值, 它与图像点在图像中的位置有关, 可用表示为

$$\delta_x(x, y) = k_1 x(x^2 + y^2) + [p_1(3x^2 + y^2) + 2p_2 xy] + s_1(x^2 + y^2) \quad (3-6)$$

$$\delta_y(x, y) = k_2 y(x^2 + y^2) + [p_2(3x^2 + y^2) + 2p_1 xy] + s_2(x^2 + y^2) \quad (3-7)$$

式中, δ_x 和 δ_y 的第一项称为径向畸变, 第二项称为切向畸变, 第三项称为薄棱镜畸变, $k_1, k_2, p_1, p_2, s_1, s_2$ 称为非线性畸变参数。一般情况下, 上述非线性模型的第一项径向畸变已足够描述非线性畸变。

3.1.3 摄像机坐标系与世界坐标系

为确定摄像机在现实空间中的位置, 需要在现实空间中定义一个基准坐标系, 并用它描述空间中任何物体的位置和姿态, 这个坐标系称为世界坐标系。世界坐标系与摄像机坐标系之间是刚体转换关系。设空间中某一点在世界坐标系和摄像机坐标系中的坐标分别是 (x_w, y_w, z_w) 和 (x_c, y_c, z_c) , 其转换关系式下式(3-8)所示:

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = R \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + T \quad (3-8)$$

式中, R 为 3×3 旋转矩阵, T 为三维平移矢量。合并式(3-1) ~ 式(3-3)以及式(3-7), 可得到世界坐标系表示的空间点三维坐标与其对应的投影点的像素坐标 (u, v) 之间的关系

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f/\Delta x & 0 & u_0 & 0 \\ 0 & f/\Delta y & v_0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = M_1 M_2 X_w \quad (3-9)$$

式中, M_1 由 $f, \Delta x, \Delta y, u_0, v_0$ 决定, 由于 $f, \Delta x, \Delta y, u_0, v_0$ 只与摄像机内部参数有关, 故称这些参数为摄像机内部参数; M_2 由摄像机相对于世界坐标系的方位决定, 称为摄像机的外部参数。确定摄像机的内部参数, 称为摄像机标定。

3.2 摄像机标定

3.2.1 线性模型的摄像机标定

在摄像机前放置一个特制的标定参照物, 摄像机获取该物体的图像, 并据此计算摄像机的内外参数, 是较为传统的摄像机标定方法。由于标定参照物上的每个特征点相对于世界坐标系的位置在制作时经过精确测定, 基于参照物上的特征点即可直接求解摄像机线性和非线性模型参数。

摄像机线性模型如下:

$$s_i \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \begin{bmatrix} x_{wi} \\ y_{wi} \\ z_{wi} \\ 1 \end{bmatrix} \quad (3-10)$$

式中, $(x_{wi}, y_{wi}, z_{wi}, 1)$ 为参照物上第 i 个点的齐次世界坐标; $(u_i, v_i, 1)$ 为第 i 点的图像齐次坐标。 m_{ij} 为投影矩阵 M 的第 i 行第 j 列元素。将公式展开包含以下三个方程:

$$s_i u_i = m_{11} x_{wi} + m_{12} y_{wi} + m_{13} z_{wi} + m_{14} \quad (3-11)$$

$$s_i v_i = m_{21} x_{wi} + m_{22} y_{wi} + m_{23} z_{wi} + m_{24} \quad (3-12)$$

$$s_i = m_{31} x_{wi} + m_{32} y_{wi} + m_{33} z_{wi} + m_{34} \quad (3-13)$$

将式(3-11)除以式(3-13), 式(3-12)除以式(3-13)分别消去 s_i 后, 可得关于 m_{ij} 的线性方程

$$x_{wi} m_{11} + y_{wi} m_{12} + z_{wi} m_{13} + m_{14} - u_i x_{wi} m_{31} - u_i y_{wi} m_{32} - u_i z_{wi} m_{33} = u_i m_{34} \quad (3-14)$$

$$x_{wi}m_{21} + y_{wi}m_{22} + z_{wi}m_{23} + m_{24} - v_i x_{wi}m_{31} - v_i y_{wi}m_{32} - v_i z_{wi}m_{33} = v_i m_{34} \quad (3-15)$$

式(3-14)和式(3-15)表示如果参照物上有 n 个已知点,并已知它们的空间坐标 $(x_{wi}, y_{wi}, z_{wi}) (i = 1, 2, \dots, n)$ 与它们的图像点坐标 $(u_i, v_i) (i = 1, 2, \dots, n)$, 则可采用直接线性变换(direct linear transformation, DLT)方式求解 M 矩阵的元素。对于 n 个特征点,有 $2n$ 个关于 M 矩阵元素的线性方程,用矩阵形式表示为

$$\begin{bmatrix} x_{w1} & y_{w1} & z_{w1} & 1 & 0 & 0 & 0 & 0 & -u_1 x_{w1} & -u_1 y_{w1} & -u_1 z_{w1} \\ 0 & 0 & 0 & 0 & x_{w1} & y_{w1} & z_{w1} & 1 & -v_1 x_{w1} & -v_1 y_{w1} & -v_1 z_{w1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{wn} & y_{wn} & z_{wn} & 1 & 0 & 0 & 0 & 0 & -u_n x_{wn} & -u_n y_{wn} & -u_n z_{wn} \\ 0 & 0 & 0 & 0 & x_{wn} & y_{wn} & z_{wn} & 1 & -v_n x_{wn} & -v_n y_{wn} & -v_n z_{wn} \end{bmatrix} \begin{bmatrix} m_{11} \\ m_{12} \\ m_{13} \\ m_{14} \\ m_{21} \\ m_{22} \\ m_{23} \\ m_{24} \\ m_{31} \\ m_{32} \\ m_{33} \end{bmatrix} = \begin{bmatrix} u_1 m_{34} \\ v_1 m_{34} \\ \vdots \\ u_n m_{34} \\ v_n m_{34} \end{bmatrix} \quad (3-16)$$

由式(3-16)可见, M 矩阵乘以任意不为零的常数并不影响 (x_w, y_w, z_w) 与 (u, v) 的关系。因此,在式中可指定 $m_{34} = 1$,从而得到关于 M 矩阵其他元素的 $2n$ 个线性方程。这些未知元素的个数为 11 个,记为 11 维向量 m 。因此,式(3-16)可简写成

$$Km = U \quad (3-17)$$

式中, K 为式(3-16)左边的 $2n \times 11$ 矩阵; m 为未知的 11 维向量; U 为式(3-16)右边的 $2n$ 维向量, K 和 U 为已知向量。当 $2n > 11$ 时,可用最小二乘法求出上述线性方程的解为

$$m = (K^T K)^{-1} K^T U \quad (3-18)$$

m 向量与 $m_{34} = 1$ 构成了所求解的 M 矩阵。由空间 6 个以上已知点与它们的图像点坐标可求解 M 矩阵。在一般的标定过程中,参照物上有数十个已知特征点,使方程的数量大大超出未知数的个数,从而用最小二乘法求解以降低误差造成的影响。

求出 M 矩阵后,还需要计算摄像机的全部内外部参数。将式中 M 矩阵与摄

像机内外部参数的关系写成

$$m_{34} \begin{bmatrix} \mathbf{m}_1^T & m_{14} \\ \mathbf{m}_2^T & m_{24} \\ \mathbf{m}_3^T & 1 \end{bmatrix} = \begin{bmatrix} \alpha_x & 0 & u_0 & 0 \\ 0 & \alpha_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{r}_1^T & t_x \\ \mathbf{r}_2^T & t_y \\ \mathbf{r}_3^T & t_z \\ \mathbf{0}^T & 1 \end{bmatrix} \quad (3-19)$$

式中, $\mathbf{m}_i^T (i=1,2,3)$ 为由式(3-18)求得的矩阵的第 i 行的前三个元素组成的行向量; m_{i4} 为 $\mathbf{M}$ 矩阵第 i 行第 4 列元素; $\mathbf{r}_i^T (i=1,2,3)$ 为旋转矩阵 $\mathbf{R}$ 的第 i 行; t_x, t_y, t_z 分别为平移向量 $\mathbf{t}$ 的三个分量。由式(3-19)可得

$$m_{34} \begin{bmatrix} \mathbf{m}_1^T & m_{14} \\ \mathbf{m}_2^T & m_{24} \\ \mathbf{m}_3^T & 1 \end{bmatrix} = \begin{bmatrix} \alpha_x \mathbf{r}_1^T + u_0 \mathbf{r}_3^T & \alpha_x t_x + u_0 t_z \\ \alpha_y \mathbf{r}_2^T + v_0 \mathbf{r}_3^T & \alpha_y t_y + v_0 t_z \\ \mathbf{r}_3^T & t_z \end{bmatrix} \quad (3-20)$$

比较式(3-20)两边可知, $m_{34} \mathbf{m}_3 = \mathbf{r}_3$, 由于 $\mathbf{r}_3$ 是正交单位矩阵的第三行, $|\mathbf{r}_3| = 1$, 所以可据 $m_{34} |\mathbf{m}_3| = 1$ 求出 $m_{34} = \frac{1}{|\mathbf{m}_3|}$ 。再由以下公式分别求出 $\mathbf{r}_3, u_0, v_0, \alpha_x, \alpha_y$:

$$\mathbf{r}_3 = m_{34} \mathbf{m}_3 \quad (3-21)$$

$$u_0 = (\alpha_x \mathbf{r}_1^T + u_0 \mathbf{r}_3^T) \mathbf{r}_3 = m_{34}^2 \mathbf{m}_1^T \mathbf{m}_3 \quad (3-22)$$

$$v_0 = (\alpha_y \mathbf{r}_2^T + v_0 \mathbf{r}_3^T) \mathbf{r}_3 = m_{34}^2 \mathbf{m}_2^T \mathbf{m}_3 \quad (3-23)$$

$$\alpha_x = m_{34}^2 |\mathbf{m}_1 \times \mathbf{m}_3| \quad (3-24)$$

$$\alpha_y = m_{34}^2 |\mathbf{m}_2 \times \mathbf{m}_3| \quad (3-25)$$

由式(3-21)~式(3-25)求解的参数可进一步解出以下参数:

$$\mathbf{r}_1 = \frac{m_{34}}{\alpha_x} (\mathbf{m}_1 - u_0 \mathbf{m}_3) \quad (3-26)$$

$$\mathbf{r}_2 = \frac{m_{34}}{\alpha_y} (\mathbf{m}_2 - v_0 \mathbf{m}_3) \quad (3-27)$$

$$t_z = m_{34} \quad (3-28)$$

$$t_x = \frac{m_{34}}{\alpha_x} (m_{14} - u_0) \quad (3-29)$$

$$t_y = \frac{m_{34}}{\alpha_y} (m_{24} - v_0) \quad (3-30)$$

因此,由空间 6 个以上已知点以及它们的图像点坐标可求解 $\mathbf{M}$ 矩阵,并按公式的次序求出全部内外参数。以上介绍了摄像机定标的计算过程,在用真实数据

进行实验时,还需要注意以下问题。

(1) M 矩阵确定了空间坐标与其相对应的图像坐标的对应关系,在一些应用场合(如立体视觉),计算出 M 矩阵后,不必要再分解出摄像机内部参数,即 M 矩阵本身也代表了摄像机参数,但这些参数没有具体的物理意义,在有些文献中称为隐参数。然而在许多增强现实应用领域,需要进行精确的摄像机跟踪定位,则需要将 M 矩阵分解,求出摄像机的内外部参数。

(2) M 矩阵由 4 个摄像机内部参数及 R 和 t 确定。由 R 矩阵是正交单位矩阵可知, R 和 t 的独立变量数为 6。因此, M 矩阵由 10 个独立变量确定,但 M 矩阵为 3×4 矩阵,有 12 个参数。由于在求 M 矩阵时, m_{34} 可指定为任意不为零的常数,所以 M 矩阵由 11 个参数决定。可见这 11 个参数并非互相独立,而是存在着变量之间的约束关系。但是在采用线性方法求解这些参数时,并没有考虑这些变量间的约束关系。因此在数据有误差的情况下,计算结果是有误差的,而且误差在各参数间的分配也没有按它们之间的约束关系考虑。

3.2.2 基于平面方格点的摄像机标定

一般来讲,3D 立体参照物的靶标制作成本较高,且加工精度受到一定的限制。Zhang 等基于 2D 平面参照物提出了摄像机标定算法^[2]。在该方法中,要求摄像机在两个以上不同的方位拍摄一个平面靶标,摄像机和 2D 平面靶标都可以自由移动,不需要知道运动参数。在标定过程中,需要假定摄像机内部参数始终保持不变,即无论摄像机从任何角度拍摄靶标,摄像机内部参数都为常数,只有外部参数发生变化。

1. 靶标平面与其图像平面之间的映射关系

靶标平面上的三维点记为 $M = (x, y, z)^T$,其图像平面上的二维点记为 $m = (u, v)^T$,则相应的齐次坐标分别记为 $\hat{M} = (x, y, z, 1)^T$ 与 $\hat{m} = (u, v, 1)^T$ 。摄像机基于针孔成像模型,空间点 M 与图像点 m 之间的映射关系为

$$s\hat{m} = K[R | t]\hat{M} \quad (3-31)$$

式中, s 为任意非零尺度因子;旋转矩阵 R 与平移向量 t 称为摄像机外部参数矩阵; K 为摄像机内部参数矩阵,定义为

$$K = \begin{bmatrix} \alpha_x & \gamma & u_0 \\ 0 & \alpha_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3-32)$$

式中, (u_0, v_0) 为主点坐标; α_u 和 α_v 分别是 u 轴和 v 轴的尺度因子; γ 是 u 轴和 v 轴的不垂直因子。不失一般性, 可以假设靶标平面位于世界坐标系的 xoy 平面, 即 $z = 0$ 。记旋转矩阵 R 的第 i 列为 r_i , 则式(3-31)可重新写为

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \begin{bmatrix} r_1 & r_2 & r_3 & t \end{bmatrix} \begin{bmatrix} x \\ y \\ 0 \\ 1 \end{bmatrix} = K \begin{bmatrix} r_1 & r_2 & t \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (3-33)$$

在这里仍采用 M 表示靶标平面上的点, 不过此时 $M = (x, y)^T$, $\tilde{M} = (x, y, 1)^T$ 。这样靶标平面上的点 M 与对应的图像点 m 之间存在一个矩阵变换 H

$$s\tilde{m} = H\tilde{M} \quad (3-34)$$

式中, $H = \lambda K \begin{bmatrix} r_1 & r_2 & t \end{bmatrix}$ 为一个 3×3 矩阵, λ 为常数因子。记 $H = \begin{bmatrix} h_1 & h_2 & h_3 \end{bmatrix}$, 则

$$\begin{bmatrix} h_1 & h_2 & h_3 \end{bmatrix} = \lambda K \begin{bmatrix} r_1 & r_2 & t \end{bmatrix} \quad (3-35)$$

式中, 平移矢量 t 为从世界坐标系的原点到光心的矢量; r_1 和 r_2 为图像平面两个坐标轴在世界坐标系中的方向矢量。显然 t 不会位于 r_1 和 r_2 构成的平面上, 由于 r_1 与 r_2 正交, 所以 $\det \begin{bmatrix} r_1 & r_2 & t \end{bmatrix} \neq 0$ 。又由于 $\det[K] \neq 0$, 因此 $\det[H] \neq 0$ 。

H 的计算是使实际图像坐标 m_i 与根据式(3-21)计算出的图像坐标 $\tilde{m}_i$ 之间残差最小的过程, 设目标函数为

$$\min \sum_i \|m_i - \tilde{m}_i\|^2 \quad (3-36)$$

2. 求解摄像机参数矩阵

当求解出 H 后, 由式(3-34)和 R 的正交性 ($r_1^T r_2 = 0, r_1^T r_1 = r_2^T r_2$) 可得到两个基本方程

$$h_1^T K^{-T} K^{-1} h_2 = 0 \quad (3-37)$$

$$h_1^T K^{-T} K^{-1} h_1 = h_2^T K^{-T} K^{-1} h_2 \quad (3-38)$$

式(3-37)和式(3-38)是关于摄像机内部参数的两个基本约束, 因为一个转换矩阵 H 有 8 个自由度, 而外部参数有 6 个 (3 个旋转, 3 个平移), 所以从一个转换矩阵 H 只能获得关于摄像机内部参数的两个约束。

空间上的二次曲面可表示为 $\tilde{x}^T B \tilde{x} = 0$, 其中 $\tilde{x} = (x, y, z, 1)^T$, B 是一个 4×4 的对称矩阵。显然, B 乘以任何一个不为零的标量仍描述同一二次曲面, 而平面上的二次曲线可表示为 $\tilde{x}^T B \tilde{x} = 0$, 其中 $\tilde{x} = (x, y, 1)^T$, B 是一个 3×3 的对称矩阵。显然,

B 乘以任何一个不为零的标量描述的仍然为同一二次曲面。因此, $K^{-T}K^{-1}$ 事实上描述了绝对二次曲线在图像平面上的投影。令

$$B = K^{-T}K^{-1} = \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\alpha_x^2} & -\frac{r}{\alpha_x^2\alpha_y} & \frac{v_0r - u_0\alpha_y}{\alpha_x^2\alpha_y} \\ -\frac{r}{\alpha_x^2\alpha_y} & \frac{r^2}{\alpha_x^2\alpha_y^2} + \frac{1}{\alpha_y^2} & -\frac{r(v_0r - u_0\alpha_y)}{\alpha_x^2\alpha_y^2} - \frac{v_0}{\alpha_y^2} \\ \frac{v_0r - u_0\alpha_y}{\alpha_x^2\alpha_y} & -\frac{r(v_0r - u_0\alpha_y)}{\alpha_x^2\alpha_y^2} - \frac{v_0}{\alpha_y^2} & \frac{(v_0r - u_0\alpha_y)^2}{\alpha_x^2\alpha_y^2} + \frac{v_0^2}{\alpha_y^2} + 1 \end{bmatrix} \quad (3-39)$$

注意到 B 是对称矩阵, 可以另表示为

$$\mathbf{b} = (B_{11}, B_{12}, B_{22}, B_{13}, B_{23}, B_{33})^T \quad (3-40)$$

设 H 中的第 i 列向量为 $\mathbf{h}_i = (h_{i1}, h_{i2}, h_{i3})^T$, 因此有

$$\mathbf{h}_i^T B \mathbf{h}_j = \mathbf{v}_{ij}^T \mathbf{b} \quad (3-41)$$

式中, $\mathbf{v}_{ij} = (h_{i1}h_{j1}, h_{i1}h_{j2} + h_{i2}h_{j1}, h_{i2}h_{j2} + h_{i3}h_{j1} + h_{i1}h_{j3}, h_{i3}h_{j2} + h_{i2}h_{j3}, h_{i3}h_{j3})^T$ 。

这样式(3-41)可以写成两个关于 $\mathbf{b}$ 的齐次方程

$$\begin{bmatrix} \mathbf{v}_{12}^T \\ (\mathbf{v}_{11} - \mathbf{v}_{22})^T \end{bmatrix} \mathbf{b} = 0 \quad (3-42)$$

如果对靶标平面拍摄 n 幅图像, 将 n 个这样的方程组叠加起来可得

$$V\mathbf{b} = 0 \quad (3-43)$$

式中, V 为 $2n \times 6$ 的矩阵。如果 $n \geq 3$, 一般来说, $\mathbf{b}$ 可以在相差一个尺度因子的意义下唯一确定。如果 $n = 2$, 可以附加约束 $\gamma = 0$, 即 $B_{12} = 0$ 。因此, 可用 $[0 \ 1 \ 0 \ 0 \ 0 \ 0]^T \mathbf{b} = 0$ 作为式(3-43)的一个附加方程。方程的解是 $V^T V$ 的最小特征值对应的特征向量, 或通过对矩阵 V 进行奇异值分解(singular value decomposition, SVD)求解出 $\mathbf{b}$ 。

当求解出 $\mathbf{b}$ 后, 可以利用 Cholesky 矩阵分解算法求解 K^{-1} , 再求逆得到 K 。一旦 K 求出后, 每幅图像的外部参数就很容易求出。由式(3-35)有

$$\mathbf{r}_1 = \lambda K^{-1} \mathbf{h}_1, \mathbf{r}_2 = \lambda K^{-1} \mathbf{h}_2, \mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2, t = \lambda K^{-1} \mathbf{h}_3 \quad (3-44)$$

式中, $\lambda = 1/\|\mathbf{K}^{-1}\mathbf{h}_1\| = 1/\|\mathbf{K}^{-1}\mathbf{h}_2\|$ 。

通常情况下, 摄像机镜头是有畸变的。因此, 以上获得的参数可作为初值进行

优化搜索,从而计算出所有参数的准确值。

3.2.3 非线性模型的摄像机标定

摄像机非线性模型除包括线性模型中的全部参数外,还包括径向畸变参数 k_1 、 k_2 和切向畸变参数 p_1 、 p_2 。线性模型参数与非线性畸变参数 k_1 、 k_2 、 p_1 、 p_2 一起构成了非线性模型的摄像机内部参数。

Faig^[3]提出了对这些参数标定的非线性优化算法,Tsai^[4]给出了在假定只存在径向畸变条件下的标定算法。这些方法都涉及非线性方程求解,或需假设摄像机部分内部参数可由其他方法获得;或者用线性模型首先计算出线性模型的参数,作为近似值,再用迭代的方法计算精确解。摄像机非线性模型的标定算法以及后续研究的摄像机自标定算法涉及的知识面较广,感兴趣的读者可参考中文版的《计算机视觉中的多视图几何》^[5]、《计算机视觉中的数学方法》^[6]以及《计算机视觉——计算理论与算法基础》^[7]等经典书籍。

3.3 基于透视式头盔显示器的增强现实系统标定

设计增强现实系统时,最基本的问题就是实现虚拟信息和现实世界的融合。一般而言,增强现实系统较常采用的显示主要有头盔显示和投影显示两种。头盔显示器是增强现实系统中的关键显示设备,分为视频透视式头盔显示器和光学透视式头盔显示器两类。前者是利用摄像机对真实世界进行实时拍摄,将视频信号同步送入计算机中与虚拟物体进行注册融合,然后在显示器上输出。用户通过显示器观看到融合的场景,如图3.4所示;而后者是利用光学组合器直接将虚拟物体与真实场景在人眼中融合,如图3.5所示,实现对真实场景的增强。在很多方面,如在医学手术导航应用领域,光学透视式头盔显示器具有更大的优势。但是它的研制难度较大,除传统的半反半透组合镜外,科研人员还将自由曲面^[9]、全息元件^[10]、投影结构^[11]等新技术应用到头盔显示器中,实现大视场的超轻型光学透射式头盔显示器。图3.5所示为北京理工大学采用自由曲面和离轴结构实现的一种拼接式大视场、高分辨率光学透视式头盔显示器。以下分别就采用光学透视式头盔显示器和视频透视式头盔显示器的增强现实系统标定作详细介绍。

3.3.1 视频透视式增强现实系统的标定

在增强现实系统中,当用户观察自然世界的视角改变时,虚拟摄像机参数必须



图 3.4 视频透视式头盔显示器^[8]



图 3.5 光学透视式头盔显示器^[9]

与真实摄像机参数保持一致,同时需要实时跟踪真实物体的位置和姿态参数,通过这些参数更新虚拟物体的位置和姿态。在虚实对准过程中,系统中某些设备(如摄像机)的内部参数以及系统某些设备之间的相对位置方向等参数始终保持不变,因而可以提前对这些参数进行测量或者标定(calibration)。增强现实系统常用的两种显示设备存在不同的标定方法。

就视频透视式增强现实系统而言,用户对其周围真实场景的感知是通过 CCD 摄像机获得的。因而在视频透视式增强现实系统中,摄像机代替用户眼睛的功能来感知周围的世界。对于只是用 CCD 获取图像的系统而言,CCD 的内部参数有出入,对系统影响不大。但是对于增强现实系统来说,跟踪注册算法的数据直接源于 CCD 的图像信息,并且整个注册算法是基于摄像机投影原理的,如果 CCD 摄像机的内部参数不准确,就会使整个注册模型产生严重误差,甚至无法正常使用。基于以上考虑,对于 CCD 厂家标称的内部参数,只能作为参考。要实现准确的跟踪注册,必须对 CCD 的内部参数进行实验室标定。因而,对视频透视式增强现实系统的标定实质上是指对 CCD 摄像机的标定,而摄像机标定在前面已作了详细的介绍,在此不再赘述。

3.3.2 光学透视式头盔显示器的标定技术

与视频透视式 AR 系统相比,光学透视式 AR 系统的标定更复杂、更困难,其主要原因有以下两方面。

(1)在视频透视式 AR 系统中,用户通过由摄像机获取的真实环境图像间接获得真实环境信息。而在光学透视式 AR 系统中,用户使用眼睛透过半透明目镜直接获得自然环境中的物体。因此,光学透视式 AR 系统标定不能像视频透视式一样可以直接、方便地处理真实物体图像中的特征。

(2)光学透视式 AR 系统在使用过程中,用户身份的改变等原因都将导致人眼

眼点位置的潜在变化,这种人为因素不可避免地增大了光学透视式 AR 系统的标定难度。在视频透视式 AR 系统中,虚拟摄像机参数应该与摄取真实场景的摄像机参数相等,因此,视频透视式标定主要是对摄像机的内部参数进行标定。而在光学透视式 AR 系统中,摄取真实场景的光学系统是人眼和光学透视式显示器,因而光学透视式标定主要是对由人眼和光学透视式显示器构成的虚拟摄像机进行标定。

针对该问题,已有很多研究学者提出各种不同的光学透视式标定方法和步骤。Janin 等采用手工测量和优化方法确定光学透视式参数^[12]。Azuma 等提出预测跟踪技术来提高光学透视式的静态和动态注册^[13]。Oishi 等提出了一种新校准方法用于最小化光学透视式射影变换参数的系统误差^[14]。Kato 等提出用网格点方法手工标定光学透视式系统^[15]。McGarrrity 等提出了基于 through-the-lens 摄像机控制的光学透视式标定方法^[16]。Tuceryan 等提出了一个用户友好的单点主动对准法(single point active alignment method, SPAAM),用于标定光学透视式 AR 系统^[17]。之后,在 SPAAM 的基础上,Genc 等进一步提出了立体显示的 SPAAM 方法^[18,19]以及扩展的 SPAAM 双步骤方法^[20]。Hua 等提出用校准摄像机代替人眼,使用传统的基于图像的测量方法标定光学投射头部固定投影显示器(OSTHMPD)^[21,22],该标定方法经过修正也适用于光学透视式标定。Tang 等在 Hua 等提出的标定方法的基础上,提出了一个新的双步骤方法标定光学透视式 AR 系统^[23]。

本节主要对近十多年来光学透视式 AR 系统的标定技术进行概括性综述,在简要介绍光学透视式标定系统和标定要求后,阐述光学透视式标定的计算模型和标定算法,其中重点介绍当前比较实用的基于 SPAAM 和基于图像的标定算法。

1. 问题描述

当前使用的光学透视式标定系统的组成结构大同小异。相同之处在于都是由标定标靶、头部跟踪器、人眼以及光学透视式头盔构成。不同之处在于所使用的标靶模式、头部跟踪器类型和光学透视式头盔型号。如图 3.6 所示为由视频跟踪器构成的典型光学透视式标定系统,该标定系统有虚、实图像对应的两个成像过程。光学透视式显示器屏幕、光学系统以及半透明观察像平面构成一个虚拟图像成像过程,即计算机生成的虚拟图像经过视频通道显示在光学透视式的显示器屏幕上后,再经过光学透视式成像系统成像到用户观察的像平面上。人眼观察自然世界中的物体,将世界坐标系中的物体射影到光学透视式显示器的像平面上的过程是

真实图像的成像过程。如果人眼能够精确地位于光学透视式光学系统的出瞳位置,则这两个成像过程的像就能够在光学透视式像平面上达到精确对准。在这两个成像过程中,人眼与光学透视式显示器构成了一个综合成像系统,或称为虚拟摄像机系统。人眼或光学透视式光学系统的出瞳位置是虚拟摄像机系统的坐标原点。

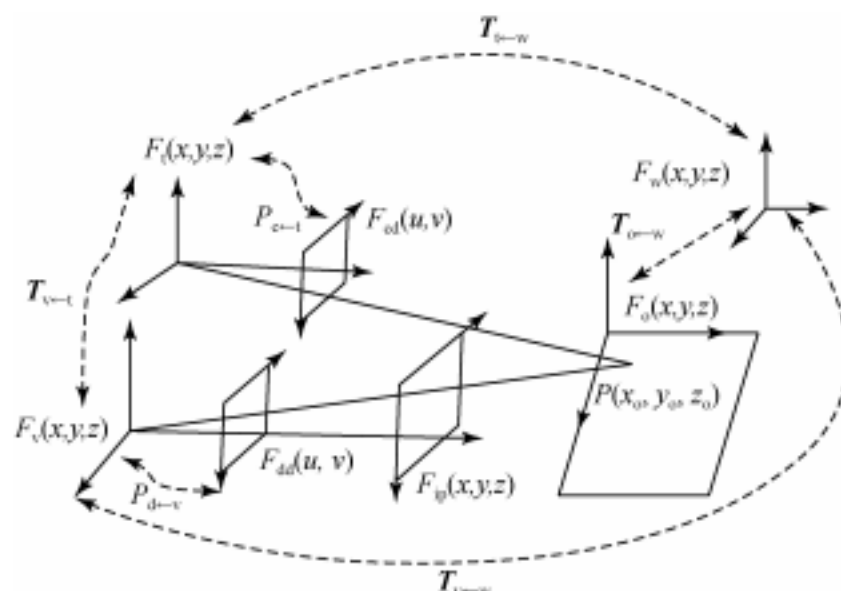


图 3.6 标定系统的坐标系统及其变换

如图 3.6 所示的光学透视式标定系统的每一个组成部分对应一个参考坐标系。标定标靶对应一个标靶参考坐标系 $F_e(x, y, z)$, 光学头部跟踪器对应一个摄像机坐标系 $F_v(x, y, z)$ 和一个摄像机屏幕坐标系 $F_{cd}(u, v)$ 。光学透视式显示器的显示屏幕对应一个显示屏幕坐标系 $F_{ad}(u, v)$, 光学透视式半透明像平面显示器对应一个观察像平面坐标系 $F_{ip}(x, y, z)$, 人眼则对应人眼坐标系 $F_e(x, y, z)$ 。虚拟摄像机坐标系 $F_v(x, y, z)$ 等同于人眼坐标系 $F_e(x, y, z)$ 。世界坐标系 $F_w(x, y, z)$ 可以设定在真实场景的某一个位置。

标定系统有刚体变换和透视变换两类坐标变换,前者描述 3D 空间中坐标系之间的变换关系,后者描述一个光学成像系统(摄像机等)的 3D-2D 成像关系。两者可按照如下规则统一定义,定义 $T_{b \leftarrow a}$ 表示从坐标系 a 到坐标系 b 之间的刚体变换,设定 3D 世界坐标系 a 中任意一点的非齐次坐标为 $p_a = (x_a, y_a, z_a, 1)^T$,其在坐标系 b 中对应的齐次坐标为 $p_b = (x_b, y_b, z_b, 1)^T$,则刚体变换 $T_{b \leftarrow a}$ 可写为

$$p_b = T_{b \leftarrow a} p_a = \begin{bmatrix} r_{b \leftarrow a} & t_{b \leftarrow a} \\ 0 & 1 \end{bmatrix} p_a \quad (3-45)$$

$$\mathbf{r}_{b \leftarrow a} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}_{b \leftarrow a}, \quad \mathbf{t}_{b \leftarrow a} = (t_x, t_y, t_z)_{b \leftarrow a}^T$$

式中, $\mathbf{r}_{b \leftarrow a}$ 表示坐标系 a 到坐标系 b 变换的旋转部分, 后者表示坐标系 a 到坐标系 b 变换的平移部分。上述定义下的刚体变换同时也满足

$$\mathbf{T}_{b \leftarrow a} = \mathbf{T}_{a \leftarrow b}^{-1}, \quad \mathbf{r}_{b \leftarrow a} = \mathbf{r}_{a \leftarrow b}^{-1}, \quad \mathbf{t}_{b \leftarrow a} = \mathbf{t}_{a \leftarrow b}^{-1} \quad (3-46)$$

以符号 $\mathbf{A}_{b \leftarrow a} = \mathbf{A}_{a \leftarrow b}^{-1}$ 表示从坐标系 a 到坐标系 b 的变换矩阵, 它等于坐标系 a 到坐标系 b 的变换矩阵 $\mathbf{A}_{a \leftarrow b}$ 的逆。假设坐标系 b 是一个摄像机坐标系, c 是摄像机的成像平面坐标系, 则定义 $\mathbf{P}_{c \leftarrow b}$ 表示从坐标系 b 到坐标系 c 的透射变换。而 p 定义的是点的坐标。设定摄像机坐标系 b 中的点 p_b 在 2D 图像坐标系 c 中的对应成像点齐次坐标为 $p_c = [u_c \quad v_c \quad 1]^T$, 则在针孔摄像机模型下的透视变换 $\mathbf{P}_{c \leftarrow b}$ 为

$$\rho p_c = \mathbf{P}_{c \leftarrow b} p_b = [\mathbf{P}_{c \leftarrow b} \quad 0] p_b \quad (3-47)$$

式中, $\mathbf{P}_{c \leftarrow b} = \begin{bmatrix} f_u & \gamma & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}_{c \leftarrow b}$, f_u, f_v 为焦距, (u_0, v_0) 为光轴与图像平面的交点坐标, γ 为图像畸变。因此, 世界坐标系 a 与摄像机像平面坐标系 c 之间的 3D-2D 成像关系可描述为

$$\rho p_c = \mathbf{P}_{c \leftarrow b} p_b = \mathbf{P}_{c \leftarrow b} \mathbf{T}_{b \leftarrow a} p_a = \mathbf{M}_{c \leftarrow a} p_a \quad (3-48)$$

依据上述坐标变换定义, 光学透视式标定系统的坐标变换如表 3.1 所示。根据这些变换的属性特征, 将光学透视式标定分为四类变换标定: 标靶标定、跟踪器标定、人眼标定、显示器标定。

表 3.1 标定系统的变换描述

变换	描述	属性
$\mathbf{T}_{w \leftarrow a}$	标靶参考系 $\rightarrow$ 世界坐标系	刚性, 固定
$\mathbf{T}_{t \leftarrow w}$	世界坐标系 $\rightarrow$ 跟踪摄像机坐标系	刚性, 变化
$\mathbf{P}_{c \leftarrow t}$	跟踪器坐标系 $\rightarrow$ 成像平面参考系	透视, 固定
$\mathbf{T}_{v \leftarrow t}$	跟踪器坐标系 $\rightarrow$ 虚拟摄像机坐标系	刚性, 固定
$\mathbf{T}_{v \leftarrow w}$	世界坐标系 $\rightarrow$ 虚拟摄像机坐标系	刚性, 变化
$\mathbf{P}_{d \leftarrow v}$	虚拟摄像机坐标系 $\rightarrow$ 显示屏幕坐标系	透视, 固定

标靶标定是测量标靶相对于世界参考系的位置和方向, 确保虚拟标靶能够时

刻与世界参考点对准。可使用刻度尺等度量工具进行标靶标定,将多次直接测量得到的姿态参数数据的平均值作为有效值。本节讨论的标定算法均将标靶坐标转换为世界参考坐标系中的坐标使用。

跟踪器标定是测量跟踪器相对于世界参考系的位置和姿态。若使用电磁跟踪器,则通过测量电磁跟踪器的传感器相对于发射器的姿态以及发射器相对于世界参考点的姿态数据来计算跟踪器标定参数。由于容易受环境影响,必须对电磁跟踪器的测量数据进行多次测量取平均值处理。若使用视频跟踪器,则利用标靶模式的图像计算跟踪器位置姿态参数。使用此方法必须提前对摄像机的内部参数进行标定,可参考 Zhang^[2] 和 Tsai^[4] 提出的方法。人眼标定是测定人眼相对于跟踪器的位置和方向,显示器标定是测量虚拟摄像机的内部参数。两者是光学透视式标定需要解决的主要问题。

2. 基本计算模型

依据前述坐标定义,光学透视式标定系统中虚拟摄像机的成像关系为

$$p_d = M_{d \rightarrow w} p_w = P_{d \rightarrow v} T_{v \rightarrow w} p_w \quad (3-49)$$

设定光学透视式标定系统的人眼始终位于虚拟摄像机的位置,则人眼位置与世界坐标系中的标靶之间可以通过两条变换路径联系起来,一条变换路径直接经过光学透视式显示器屏幕,另外一条变换路径经过光学透视式的跟踪器。在理论上,这两条路径上的刚体变换应该相等,即

$$T_{v \rightarrow w} = T_{v \rightarrow t} T_{t \rightarrow w} \quad (3-50)$$

将式(3-50)代入式(3-49)得到

$$p_d = P_{d \rightarrow v} T_{v \rightarrow t} T_{t \rightarrow w} p_w = M_{d \rightarrow t} p_t \quad (3-51)$$

式中, $p_t = T_{t \rightarrow w} p_w$ 代表世界坐标系中的校准点在跟踪器坐标中的位置和方向; $M_{d \rightarrow t} = P_{d \rightarrow v} T_{v \rightarrow t}$ 是跟踪器坐标系中的校准点在虚拟摄像机屏幕上成像的有效射影矩阵。在式(3-51)中, $P_{d \rightarrow v}$ 包含光学透视式显示器标定的参数 $(f_u^d, f_v^d, u_0^d, v_0^d, \gamma^d)$, $T_{v \rightarrow t}$ 包含人眼标定的参数 $(r_{v \rightarrow t}, t_{v \rightarrow t})$, 因此, $M_{d \rightarrow t}$ 隐性地描述了光学透视式标定参数。 $T_{t \rightarrow w}$ 可通过跟踪器标定确定,则容易计算出 p_t 。在 p_t, p_v 可测的条件下,可通过测量 p_t 与 p_v 多点对应,将光学透视式标定转化为求解 $M_{d \rightarrow t}$ 的数学问题。更进一步,若跟踪器为单个视频摄像机,则跟踪器的成像关系为

$$T_{c \rightarrow w} = P_{c \rightarrow t} T_{t \rightarrow w} = P_{c \rightarrow t} [r_{t \rightarrow w} \quad t_{t \rightarrow w}] \quad (3-52)$$

对式(3-51)进行矩阵变换,并把式(3-52)代入,可得到

$$M_{d \rightarrow w} = M^1 \lambda M_{c \rightarrow w} + M^2 \quad (3-53)$$

式中, $M^1 = P_{d-v} r_{v-1} P_{c-1}^{-1}$, $M^2 = [0_{3 \times 3} \quad P_{d-v} t_{v-1}]$ 。在式(3-53)中, M^1 和 M^2 包含了光学透视式标定参数, 通过测量 p_d 与 p_w 多点对应, 可计算出 M^1 与 M^2 。

式(3-51)和式(3-53)描述了光学透视式标定的计算模型。基于此计算模型, 可采用不同方法对光学透视式进行标定。下面分别介绍比较实用的基于 SPAAM 和基于图像的标定方法。

3. 标定方法

早期的光学透视式标定方法是静态标定法, 即在保持头部静止不动条件下标定。原理是人眼透过光学透视式成像屏幕观察真实场景中的网格校准标靶, 然后用鼠标手工控制光学透视式显示器屏幕上虚拟的瞄准器, 使之与标靶上的多个网格点对准, 依次记录相应的虚拟瞄准器的坐标。根据瞄准器坐标与对应的标靶网格点世界坐标计算光学透视式的标定参数。此方法的缺陷是过程枯燥烦琐、交互性差, 而且必须保持头部静止不动。为克服这种方法的局限性, 研究学者提出了基于 SPAAM^[17~20] 和基于图像的标定方法^[21,22]。

1) 基于 SPAAM 的标定方法

基于 SPAAM 的标定方法的原理是在保持头部运动的条件下, 利用真实场景中的一个校准标靶点进行光学透视式标定。用户的眼睛透过光学透视式屏幕观察该校准点, 然后用鼠标手工控制光学透视式显示屏幕上虚拟的瞄准器, 使之与观察屏幕上的校准点对准。根据式(3-51), 当头部运动到不同方向时, p_i 和 p_d 是变量值, M_{d-1} 是固定值, 因此可以通过在多个头部姿态条件下测量该单个校准点的 p_i 和 p_d , 建立方程组求解 M_{d-1} 。由于 M_{d-1} 是 3×4 矩阵, 有 11 个自由度, 需要至少 6 个不同方向的头部跟踪器姿态位置测量 p_i 和 p_d , 然后用最小二乘法或者奇异值分解方法计算出 M_{d-1} 。

在光学透视式系统中, 实际的人眼位置并不精确地位于虚拟摄像机坐标系的原点, 而是有一定的偏离, 如图 3.7 所示。在此情况下, 可采用改进的 SPAAM 双步骤法进行标定。第一步是使用 SPAAM 计算虚拟摄像机的位置, 第二步是计算人眼相对于虚拟摄像机坐标系的偏离参数。

实际人眼坐标系 $F_e(x, y, z)$ 相对于虚拟摄像机的偏离分为以下几种情况。

(1) 当人眼坐标系的观察方向与虚拟摄像机的观察方向一致时, 人眼-虚拟摄像机偏离将导致观察到在显示屏幕上的像产生缩放或者平移。这使得在虚拟摄像机的点 $p_v = (x_v, y_v, z_v, 1)^T$ 与实际上人眼观察到的点 $p_e = (x_e, y_e, z_e, 1)^T$ 之间的关系可以用式(3-54)模拟, 即

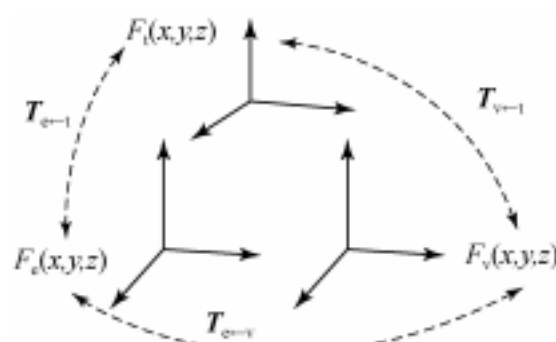


图 3.7 标定系统的坐标系及其变换

$$p_e = T_{e-v} p_v = \begin{bmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} p_v \quad (3-54)$$

(2) 当人眼坐标系的观察方向与虚拟摄像机的观察方向不一致时,人眼-虚拟摄像机偏离将导致观察到在显示屏幕上的像产生畸变。这可以用如下公式模拟:

$$p_e = T_{e-v} p_v = \begin{bmatrix} \alpha_u & s & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} p_v \quad (3-55)$$

在上述两种情况下,光学透视式标定系统的新的投影矩阵为

$$M_{e-w} = T_{e-v} M_{d-w} \quad (3-56)$$

式中, M_{d-w} 的值由式(3-51)或式(3-53)确定;求解 T_{e-v} 的方法与 SPAAM 相同,采集至少 2 组或者 3 组以上的单个校准点的测量数据,用最小二乘法或者 DLT 方法计算 T_{e-v} 。

2) 基于图像的标定方法

基于图像的标定方法是 将一个标定过的摄像机放置在光学透视式系统的人眼位置,用该摄像机模拟人眼对光学透视式进行标定的方法。如图 3.8 所示为修正的标定坐标系示意图,标定摄像机坐标系的原点位于虚拟摄像机坐标系的原点,其光轴与虚拟摄像机的光轴一致,其他两个坐标轴与虚拟摄像机坐标系对应的两个坐标轴有偏差。

基于图像的标定法分为三个步骤:①测量标定摄像机自身的内参数 P_{ced-cc} ;②放置网格状标靶在场景中,用标定摄像机摄取网格状标靶图像,计算 T_{cc-w} ,然后通过 T_{cc-w} 计算 $T_{ce-1} = T_{cc-w} T_{w-1}$;③在光学透视式显示器上显示一个黑白网格模式,利用光学透视式显示器与摄像机像平面之间的对应匹配关系计算虚拟摄像机

的内部参数 $P_{d \rightarrow v}$ 以及虚拟摄像机与标定摄像机之间的 $T_{cc \rightarrow v}$ 。与 SPAAM 方法相似,当标定摄像机与人眼的位置不一致时,需要在上述标定基础上根据对标定精度的不同应用需求,采用瞳距测量法、单点对应或者多点对应的方法对相关标定参数进行修正。

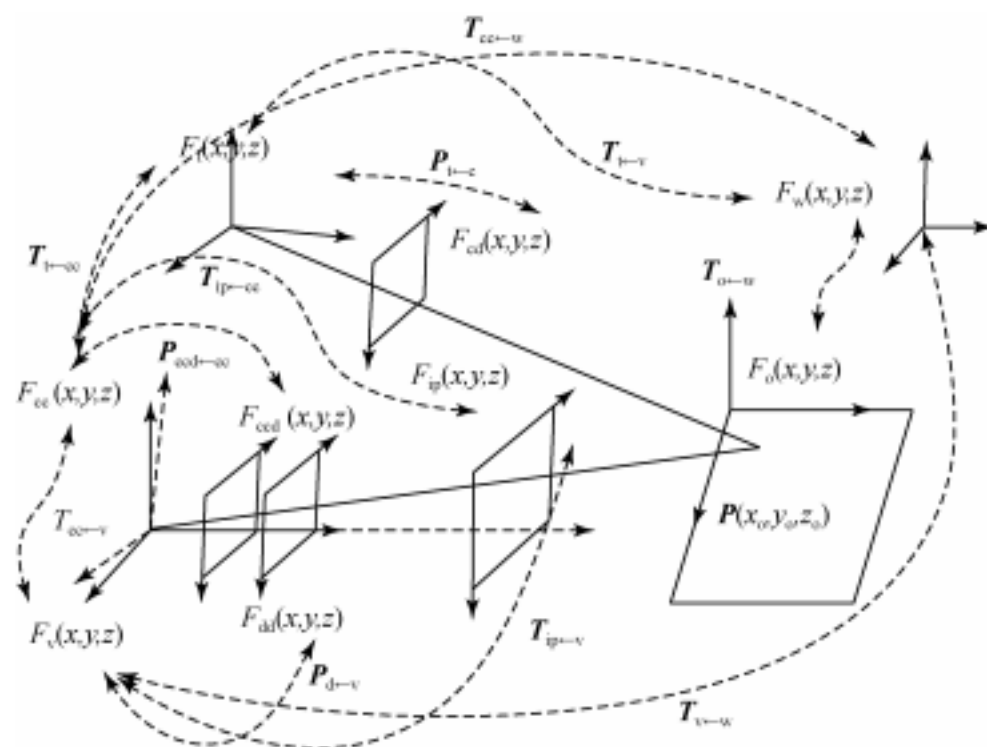


图 3.8 修正的标定系统

4. 标定评估

如何评估光学透视式标定是光学透视式标定方法的一个重要难题。由于无法获取人眼视网膜上的增强图像,无法使用基于图像的方法来确定光学透视式系统的精度。采用用户定性评估和使用摄像机代替人眼进行评估的方法都有其局限性。前者方法简单容易,但得不到具体的精度数量值。后者必须离线处理,过程烦琐,至少需要匹配上千个点对才能够获得较好的精度。一个好的光学透视式标定评估系统应该充分考虑到用户、任务复杂度以及具体的精度测量。例如,要尽量保持用户头部稳定,确保用户在进行虚实物体对准时具有快速、友好的交互性;能够通过评估得到多维的误差数据等。McGarrity^[24,25]和 Navab^[26]等提出一种在线评估光学透视式标定的方法,能够在系统运行时动态实时地获得光学透视式标定的误差参数。

在光学透视式增强现实系统中,由人为因素引入的标定误差在整个系统的总体误差中占据主体地位。因此,如何减小人为涉入因素的影响,提高标定的精度和

鲁棒性是光学透视式标定努力的目标。所采用的光学透视式标定方法也需要从观察主体、点对应的数量、点对应的匹配方式以及计算量等方面考虑。在人眼作为观察主体、多点对应的条件下,用户必须依次在不同的时间点上使用鼠标移动虚拟瞄准器对准多个世界点,这不可避免地增大了人为涉入的可能性。SPAAM 方法虽然属于手工匹配方式,但是它巧妙地利用头部运动,将世界场景中单个点的世界坐标转换为跟踪器坐标,从而建立跟踪器坐标系的单点与其对应的虚拟瞄准点之间的坐标对应。SPAAM 方法尽力减少人为涉入因素的影响,是一种用户友好的、动态的交互标定方法。基于图像的光学透视式标定方法使用现有成熟的基于图像的视频摄像机标定技术,从而可以获得光学透视式标定参数。此方法由于用标定摄像机代替了人眼,避免了人为因素导致的不确定性。虽然需要离线标定,过程也烦琐,标定精度却得到了很大提高。

参 考 文 献

- [1] <https://developer.vuforia.com/resources/sample-apps/features>[2012-6-19].
- [2] Zhang Z Y. Flexible camera calibration by viewing a plane from unknown orientations// International Conference on Computer Vision, Corfu, 1999.
- [3] Faig W. Calibration of close range photometric system; Mathematical formulation. Photogrammetric Eng Remote Sensing, 1975, 42: 1479 ~ 1486.
- [4] Tsai R Y. An efficient and accurate camera calibration technique for 3D machine vision// Proceedings of IEEE Conference of Computer Vision and Pattern Recognition, Miami, 1986.
- [5] 韦穗, 杨尚俊, 章权兵, 等. 计算机视觉中的多视图几何. 合肥: 安徽大学出版社, 2002.
- [6] 吴福朝. 计算机视觉中的数学方法. 北京: 科学出版社, 2008.
- [7] 马颂德, 张正友. 计算机视觉——计算理论与算法基础. 北京: 科学出版社, 1998.
- [8] <http://www.vuzix.com>[2011-10-8].
- [9] Cheng D, Wang Y, Hua H, et al. Design of a wide-angle, lightweight head-mounted display using free-form optics tiling. Optics Letters, 2011, 36(11): 2098 ~ 2100.
- [10] Grifantini K. CES: Clear Glasses to Augment Your Reality. <http://www.technologyreview.com/view/422294/ces-clear-glasses-to-augment-your-reality>[2014-9-5].
- [11] Hua H, Gao C, Ahuja N. Calibration of an HMPD based augmented reality system. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, 2007, 37(3): 416 ~ 430.
- [12] Janin A, Mizell D, Caudell T. Calibration of head-mounted displays for augmented reality applications// Proceedings of VRAIS, Seattle, 1993.
- [13] Azuma R, Bishop G. Improving static and dynamic registration in an optical see-through display//

- Proceedings of ACM SIGGRAPH, Orlando, 1994.
- [14] Oishi T, Tachi S. Methods to calibrate projection transformation parameters for see-through head-mounted displays. Presence, 1996, 5(1): 122 ~ 135.
 - [15] Kato H, Billinghurst M. Marker tracking and HMD calibration for a video-based augmented reality conferencing system//Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality, San Francisco, 1999.
 - [16] McGarrity E, Tuceryan M. A method for calibrating see-through head-mounted displays for AR//Proceedings of 2nd IEEE and ACM Int Workshop Augmented Reality, San Francisco, 1999.
 - [17] Tuceryan M, Navab N. Single point active alignment method (SPAAM) for optical see-through HMD calibration for AR//Proceedings of International Symposium Augmented Reality, Munich, 2000.
 - [18] Genc Y, Sauer F, Wenzel F, et al. Optical see-through HMD calibration: A stereo method validated with a video see-through system//Proceedings of International Symposium for Augmented Reality, Munich, 2000.
 - [19] Genc Y, Tuceryan M, Navab N. Practical solutions for calibration of optical see-through devices//Proceedings of International Symposium on Mixed and Augmented Reality, Darmstadt, 2002.
 - [20] Genc Y, Tuceryan M, Khameney A, et al. Optical see-through calibration with vision-based trackers: Propagation of projection matrices//Proceedings of the IEEE and ACM International Symposium on Augmented Reality, New York, 2001.
 - [21] Hua H, Gao C, Ahuja N. Calibration of a head-mounted projective display for augmented reality systems//Proceedings of IEEE International Symposium Mixed and Augmented Reality, Darmstadt, 2002.
 - [22] Gao C, Hua H, Ahuja N. Easy calibration of a head-mounted projective display for augmented reality systems//Proceedings of IEEE Virtual Reality, Los Angeles, 2003.
 - [23] Tang A, Zhou J, Owen C. Evaluation of calibration procedures for optical see-through head-mounted displays//Proceedings 2nd IEEE and ACM ISMAR, Tokyo, 2003.
 - [24] McGarrity E, Genc Y, Tuceryan M, et al. A new system for online quantitative evaluation of optical see-through augmentation//IEEE and ACM International Symposium on Augmented Reality, New York, 2001.
 - [25] Genc Y, Navab N, Tuceryan M, et al. Evaluation of optical see-through system//Proceedings of Euroimage International Conference on Augmented, Virtual Environments and Three-Dimensional Imaging, Mykonos, 2001.
 - [26] Navab N, Zokai S, Genc Y, et al. An on-line evaluation system for optical see-through augmented reality//IEEE Proceedings of Virtual Reality, Washington DC, 2004.

第4章 基于标识的增强现实跟踪注册技术

目前大多数增强现实系统采用计算机视觉的方法进行场景跟踪注册。视觉注册方法大致可分为基于标识的跟踪注册与无标识跟踪注册两大类。从另外一个角度来讲,视觉跟踪注册可以分为 inside-looking-out (IO) 和 outside-looking-in (OI) 两种不同的系统结构^[1]。其中 IO 系统是指光学感应器(如摄像机)安装在使用者身上,而标识点是固定不动的。与之相反,OI 系统是指光学感应器固定不动,而标识点安装在使用者身上,如图 4.1 所示。

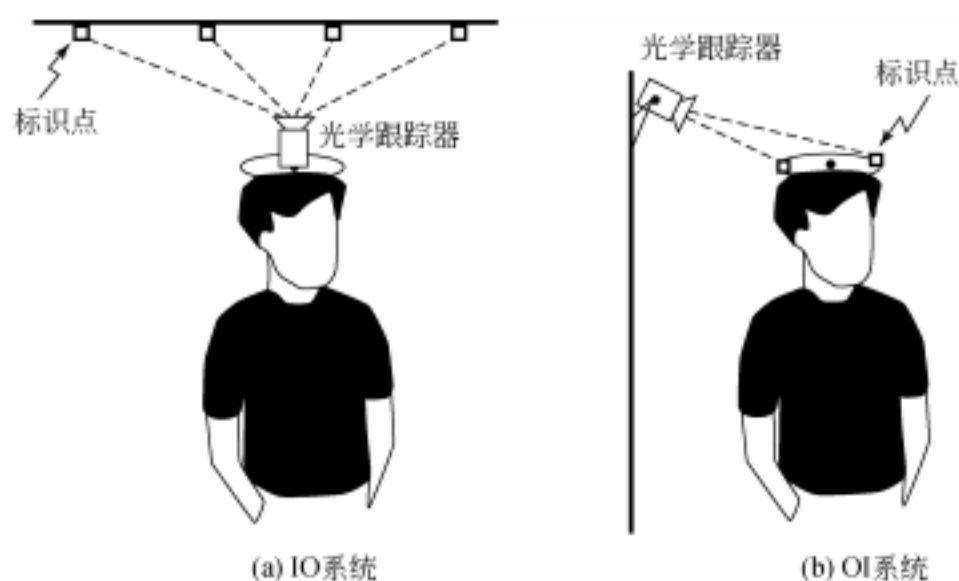


图 4.1 IO 系统和 OI 系统示意图^[1]

这两种结构各有其优点和缺点。对于 IO 光学跟踪器,摄像机被固定在用户头部,摄像机从内向外对准固定在环境中的跟踪标识目标。摄像机随头部运动而相对于环境在运动,跟踪目标相对于环境固定不动。由于光学感应器在用户身上,对光学感应器的操作比较方便。可以在任意大的范围内使用,同时可以使用密集的主动标识点或少量的、低信噪比被动标识点,甚至用自然环境的天然特征作为标识点。IO 光学跟踪器的特点是方向计算精度高,适合室内环境下的大范围跟踪。

美国 UNC 跟踪器项目组在 IO 光学跟踪系统研究方面做出了巨大贡献。1984 年,Bishop 在其博士论文中实现了一个光学自跟踪器^[2]。1990 年,Wang 在 Bishop 研究的基础上实现了一个头盔实时跟踪系统^[3,4]。1992 年,该项目组将此跟踪方

法扩展为大范围室内头盔跟踪样机系统^[5]。之后, Azuma 采用预测跟踪技术^[6]以及 Welch 提出的 SCAAT 跟踪方法^[7,8]逐步减小了该样机系统的跟踪误差,使得该系统不断完善,成为商业上可用的 Hi-Ball 大范围头部跟踪系统^[1,9]。图 4.2 为 Hi-Ball 系统的示意图。其他 IO 光学跟踪器包括 1991 年 Honeywell 公司采用旋转红外光束技术的飞行头盔跟踪器^[10], 华盛顿大学的 Kim 等开发的使用小型 Quad-cell 探测器的 IO 跟踪器^[11]。



图 4.2 UNC Hi-Ball 跟踪器^[9]

而对于 OI 结构来讲,一个或者多个摄像机被固定在环境中,让摄像机从外向内对准固定在用户 HMD 上的跟踪标识目标。跟踪标识目标随头部运动,但摄像机相对环境是固定的。OI 光学跟踪器的显著特点是位置计算精度高,但是其方向计算精度依赖于位置计算精度,容易受到位置计算误差的影响。OI 跟踪器较适用于室内环境下的小范围跟踪,不适合大范围跟踪。

早期的 OI 光学跟踪系统主要包括美国犹他大学的 Twinkle Box 位置跟踪器^[12,13]、瑞典 Selective Electric 公司开发的 SELSPOT^[14]、加拿大的 OP-EYE^[15]以及 CoSTEL^[16]等。1988 年,美国 NASA 武器研究中心为 CSRDF 飞行模拟器开发了一个典型的 OI 跟踪系统^[17],该系统使用四个摄像机扫描飞行员头盔上的一组 LED,通过测量摄像机图像焦平面上的 2D 位置来计算头部位置和方向。1991 年, Hollywell 公司采用红外 LED 灯以及视觉技术开发了飞行头盔 OI 光学跟踪系统^[10]。1994 年,波音公司的 Janin 等在飞机配件维修增强现实系统中使用了基于视觉的 OI 光学跟踪器^[18]。使用两个摄像机的 OI 光学跟踪器采用绝对朝向方法或双目立体三角化方法来计算头部位置。若使用三个及以上数量的标识点,还可估计头部方向。奥地利的 Ribo 等使用三个反射 LED 被动标识点,采用双视图对极几何约束以及三维重建方法,研制了一个六自由度跟踪系统^[19]。另外一个使用双

摄像机的 OI 光学跟踪系统是 Dorfmueller 等开发的三自由度跟踪系统^[20~22]。目前,商业上可用的 OI 光学跟踪系统有英国 Vicon 公司开发的使用多个摄像机传感器的 Vicon 系列运动跟踪系统^[23],德国 A. R. T 公司的 AR Track 红外跟踪器系列^[24],Motion Analysis 公司的 Hawk 系列^[25]。

4.1 基于标识的跟踪注册技术概述

基于标识的跟踪注册技术是当前增强现实系统中最成熟和最接近实际应用的注册技术,该注册技术一般采用 IO 系统,即将一些已知空间相对位置的人工标识点放置在需要注册的真实场景中,利用摄像机跟踪识别标识点。在已知标识点三维空间位置的基础上,采用计算机视觉的方法计算摄像机相对于真实场景的六自由度姿态。此外,为了提高标识点的跟踪精度和扩大其适用范围,可以在场景中放置多个标识点,并对每个标识点进行唯一编码。编码图形的设计和识别技术是实现基于标识点跟踪定位的一个重要环节。

大体来说,基于标识点的注册技术由两部分构成:①在摄像机采集到的图像中识别和跟踪人工标识图形上的特征点;②根据跟踪到的特征点的图像坐标信息,在摄像机透视投影模型基础上,计算摄像机与真实场景间的六自由度姿态。由此可见,跟踪识别标识点是基于标识点注册技术中最关键的部分。由于目前利用 N 个特征点进行姿态计算的各种算法已经非常成熟,如 P-N-P 算法^[26],因此标识点识别、跟踪的精度将直接决定基于标识的增强现实系统的整体注册精度。

基于标识的跟踪注册技术与增强现实系统现有的其他注册技术,如基于模型的姿态计算、基于关键帧的姿态计算相比,具有以下几方面技术优势。

(1) 系统所需的计算量小,执行速度快,跟踪定位精度高。

(2) 应用方便,对系统的整体配置要求不高,只需利用打印机打印简单的标识点图形,即可实现系统的姿态定位。

(3) 对于人为参与的要求不高,系统甚至不需要用户参与即可自动实现姿态计算。

经过多年的研究开发,国内外的研究机构和知名企业设计开发出不同类型的标识码(图 4.3)。而基于标识点注册定位技术的增强现实系统也已经在诸多领域得以广泛应用,其中,影响最为广泛的标识性成果为华盛顿大学的人机交互实验室开发的 ARToolkit 开放软件包^[27]。利用该软件包使用者可轻松构建基于标识的增强现实系统。图 4.4 所示为华盛顿大学利用 ARToolkit 标识点开发的魔法书系

统^[28]。要了解更多更新的基于 ARToolkit 软件包开发的增强现实系统的应用,可登录 www.hitl.washington.edu/ARToolkit 获取。

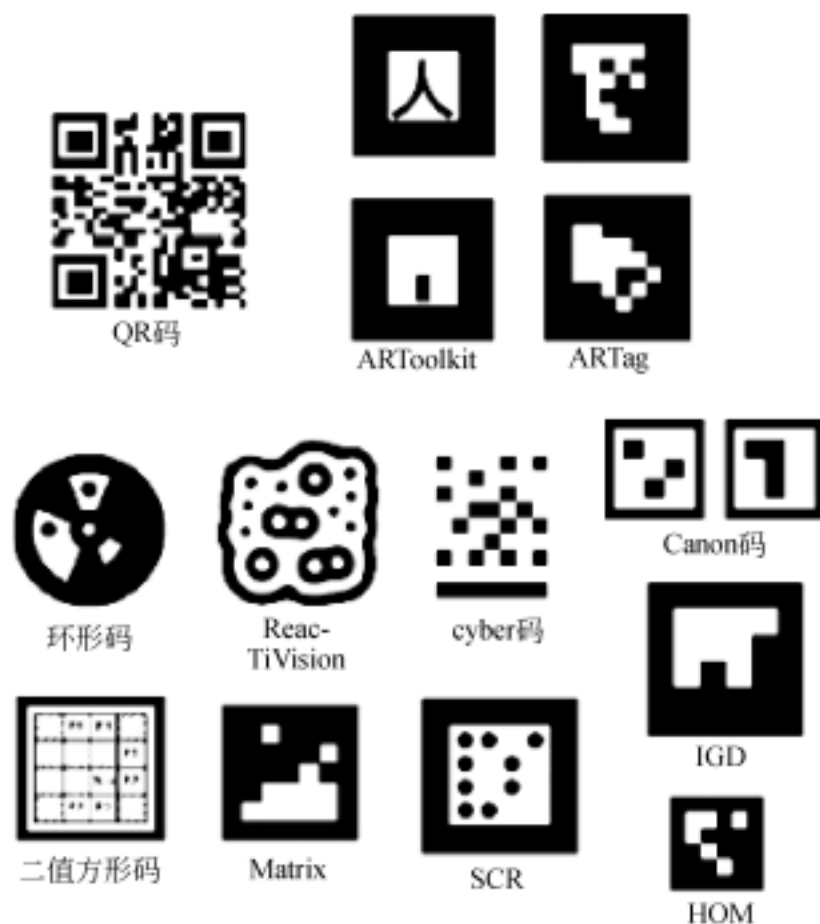


图 4.3 不同类型的标识码

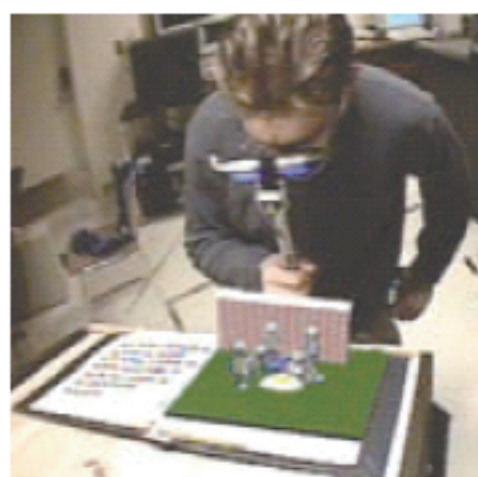


图 4.4 魔法书^[28]

放置了标识点图形的真实场景往往是非常复杂多变的。为使计算机实现标识点图形与复杂背景的分离,必须实现标识点的实时精确跟踪定位,该过程涉及以下

重要的概念。

(1) 识别。从复杂场景图像中检测出标识点图像区域,并识别标识点的编码信息。该过程是通过标识点的识别算法实现的。标识点的识别算法涉及诸如图像二值化、区域分割、边缘检测,甚至图像理解等图像处理和计算机视觉算法。

(2) 编码。单一标识点图形的可视范围是有限的,如果摄像机视野运动到标识点可视范围以外,则失去对场景的跟踪能力。此外,单一标识点也难以解决物体间的相互遮挡问题。因此,往往需要在场景中放置多个标识点扩展可视范围。为了对多个标识点进行区分,标识点上需要添加一些特殊的能代表其唯一性的编码图形,以便识别这些事先约定的编码图形,就能够实现不同标识点间的区分。

(3) 跟踪定位。准确识别出标识点区域后,需要检测标识点图形上的特征点图像坐标,用于后续的姿态计算。例如,方形标识点需要检测其四个边角的坐标,圆形标识点需要检测其中心的坐标。全局定位算法或跟踪算法常被用于此过程。

基于以上对标识点的认识,可对标识点系统进行如下分类,如图 4.5 所示。

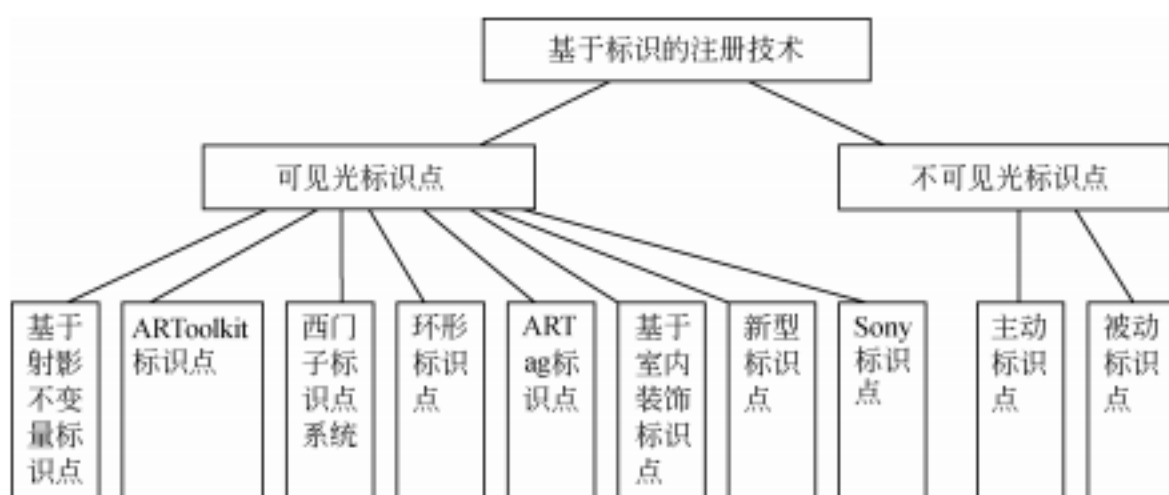


图 4.5 标识点系统分类

识别和定位算法是基于标识的增强现实跟踪注册的重要技术。识别、定位的准确性和鲁棒性直接决定了增强现实系统的整体注册精度和运行效率,是进行后续姿态计算的前提和基础。需要指出的是,标识点的识别定位与标识点图形的设计形式密不可分。不同样式的标识点需要不同的识别定位算法。如何设计标识点,使其识别更为容易,定位更加准确,实现更多数量的编码成为该技术的一个重要研究课题。下面结合各种常见的标识点样式,介绍相应的识别定位算法。

4.2 常用的标识点

4.2.1 ARToolkit 方形标识点

ARToolkit^[27]是由美国华盛顿大学的 HTL 实验室设计开发的,用于实现基于标识的增强现实跟踪注册软件包,该软件包采用的方形标识点是当前增强现实系统中最常被采用的标识点样式。

1. 标识点的设计

ARToolkit 采用如图 4.6 所示的标识点形式。ARToolkit 标识点可通过打印机直接打印获得,因此它的制作成本低廉,并且可以应用于各种真实环境中的物体。它使用可见的标记和视频摄像机来确定真实环境中的物体及其位置和方向。



图 4.6 ARToolkit 标识点^[27]

图 4.6 所示的标识点由三部分组成。

(1)内部的标识。内部的标识为不同的标识点提供了可以相互区别的编码特征。不同的内部标识在外观上必须有足够的差异,以满足计算机对图形分辨能力的要求。过于相似的内部标识会导致标识点识别时产生错误匹配。例如,在使用黑体字的时候,“A”和“4”,“B”和“8”就会经常发生将“A”识别为“4”,“B”识别为“8”或将“4”识别为“A”,“8”识别为“B”的现象。因此,应尽量避免同时使用这些容易产生识别混淆的字符标识。

(2)外部的黑色方框。外部的黑色方框用于对标识点和环境进行区分。在识别过程中,识别算法首先在图像上搜索所有的黑色方框并确认为标识点,之后根据方框内的字符标识识别标识点的编码。

(3)最外层的白色区域。为了能够稳定地识别黑框,在黑框外侧应该有足够

面积的白色区域,这个区域的外边界形状对识别没有影响。

2. 标识点的特征提取与编码识别

ARToolkit 采用以下步骤进行标识点的特征提取与编码识别。

1) 二值化

首先将拍摄到的视频图像二值化。由于标识点图形是黑白颜色构成的,因此二值化不会影响标识点图形的特点,但可以大大简化图像,使识别更容易。

2) 连通区域标识

二值化后的图像是只包含 0 和 1 的黑白图像。采用“连通域标记”算法获得图像中的联通域(图 4.7)。经过连通区域标记算法处理后,会得到一个候选连通区域集合。面积过小的连通区域将被放弃。

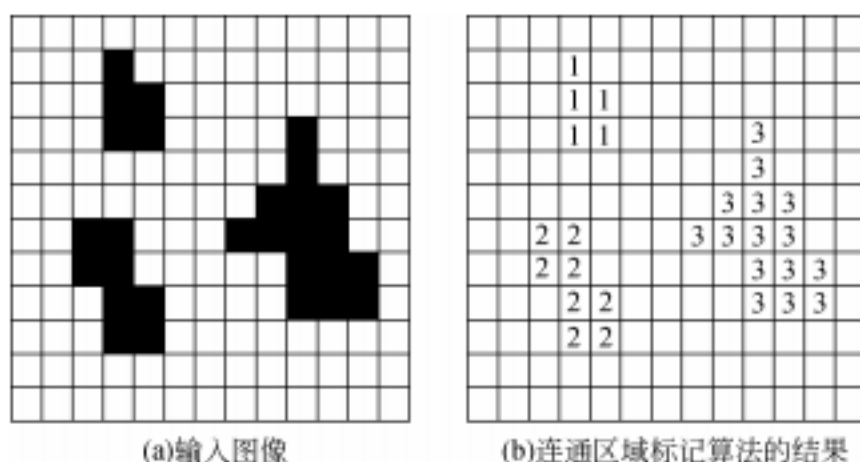


图 4.7 输入图像和连通区域标记算法的结果

3) 连通区域边界跟踪

使用区域边界跟踪算法得到每个候选区域的区域边界。算法步骤简述如下(图 4.8)。

(1)从左到右、从上到下扫描图像,求区域 S 的起始点 $s(k) = (x(k), y(k))$, $k = 0$ 。

(2)用 c 表示当前边界上被跟踪的像素点,置 $c = s(k)$,记 c 左邻点为 b , $b \in \tilde{S}$ 。

(3)按逆时针方向从 b 开始将 c 的 8 个邻点分别记为 $n_1, n_2, \dots, n_8$, $k = k + 1$ 。

(4)从 b 开始沿逆时针方向找到第一个 $n_i \in S$ 。

(5)置 $c = s(k) = n_i$, $b = n_{i-1}$ 。

(6)重复步骤(3)~步骤(5),直到 $s(k) = s(0)$ 。

4) 寻找候选四边形区域

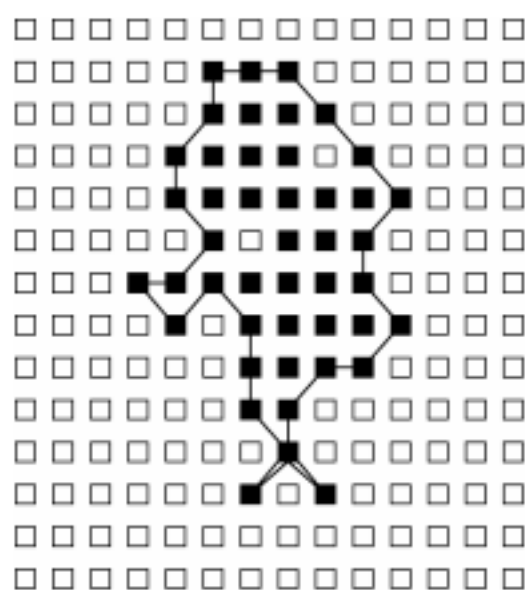


图 4.8 连通区域边界探索

- (1) 选取区域边缘点中最左上角的点作为四边形的顶点 A 。
- (2) 选取区域边缘点中距离点 A 最远的点作为顶点 B 。
- (3) 点 A 和点 B 将边缘闭合曲线分割为两部分,对于按照顺时针方向从 A 到 B 的部分,实施以下操作。
- (4) 首先找到各点到直线 AB 距离最远的点 M 。
- (5) 若此距离小于某一阈值(关于此阈值的选取下面会详述),则可认为 M 离直线 AB 过近,可认为是直线,如图 4.9(a) 所示;若大于此阈值,则存在三种可能, MA 和 MB 均为直线, MA 或 MB 之间仍有顶点,如图 4.9(b) 和图 4.9(c) 所示。对于 MA 段, MB 段递归实施步骤(4)和步骤(5)。
- (6) 对于按顺时针从 B 到 A 的部分,实施步骤(4)和步骤(5)。
- (7) 若按顺时针从 B 到 A 的部分和从 A 到 B 之间都有一个顶点,分别记为 C 、 D ,那么 $ADBC$ 可认为是一个标识点四边形。若 A 到 B (或 B 到 A) 段为直线,而 B 到 A (或 A 到 B) 段之间有两个顶点,依次记为 E 、 F ,则可认为 $ABEF$ 为候选标识点四边形。若两段都为直线或两段都有两个以上的顶点或两段中,一段为直线另一段内只有一个顶点或大于三个顶点,则可排除此区域。

此时,方形标识点的识别过程基本完成,接着提取标识点中的特征点。ARToolkit 采用四边形的四个顶点作为特征点,对四边形的四条边采用“线性回归”算法进行直线拟合,然后取四条直线的交点就可获得四边形的四个顶点。

5) 标识点编码、解码算法

标识点的解码过程实际上就是将当前标识点的内部图案与该系统存储的所

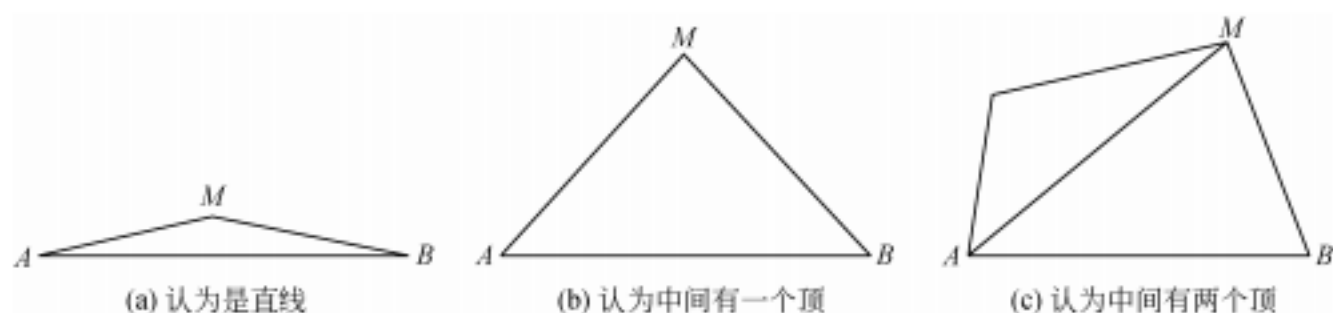


图 4.9 四边形检测算法

有标识点图案逐一比较,找到一致匹配的过程。由于 ARToolkit 的形状是对称的正方形,内部图案的匹配实际上需要在四个方向上进行四次,其具体过程如下。

首先,对于前边得到的四边形区域,将内部图案区域进行单应性变换,将其拉伸到正方形形状。单应性变换是射影变换的一种特殊形式,平面图形在射影变换中的变化就是单应性变换。如图 4.10 所示,单应性变换可以将左边图像中的每一像素点变换到右边的正方形形状对应点。

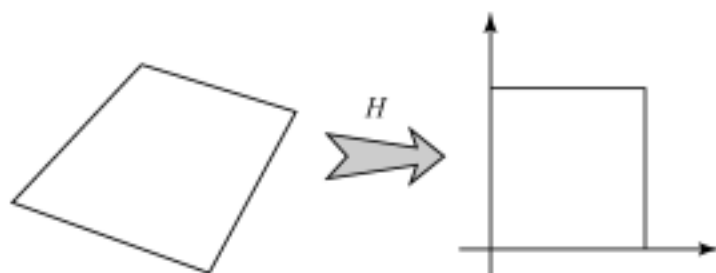


图 4.10 单应性图像变换矫正图像

单应性变换可表示为式(4-1)所示形式

$$m' = \lambda \cdot H \cdot m \quad (4-1)$$

式中, m 为采集图像中的一个 2D 像素点的齐次坐标表示, m' 是目标正方形图像坐标系中的对应 2D 像素点的齐次坐标, H 为此单应性变换矩阵。为获得 H ,可以利用四边形四个顶点的对应关系。设 $m_i (i=1, \dots, 4)$ 为采集图像中四边形四个顶点, $m'_i (i=1, \dots, 4)$ 为目标正方形图像空间中四边形四个顶点,这些点的坐标都是已知的。代入 $m = \lambda \cdot H \cdot m'$ 可得到一个齐次线性方程组,使用 SVD 分解算法可获得 H 的近似解。

接下来就可以利用正方形图像与已知标识点图案进行比较。ARToolkit 采用对图像取“互相关”的方法,减小算法对光照的敏感性。若求采集图像 I 和模板图像 P 的互相关,首先需要计算图像的均值和标准差

$$\mu_I = \frac{1}{xy} \sum_x \sum_y I(x, y), \quad \mu_P = \frac{1}{xy} \sum_x \sum_y P(x, y) \quad (4-2)$$

$$\sigma_I = \left(\sum_x \sum_y (I(x, y) - \mu_I)^2 \right)^{\frac{1}{2}}, \quad \sigma_P = \left(\sum_x \sum_y (P(x, y) - \mu_P)^2 \right)^{\frac{1}{2}} \quad (4-3)$$

式中, $I(x, y)$ 和 $P(x, y)$ 分别表示采集图像和匹配模板图像, 两幅图像的互相关计算公式为

$$\rho = \frac{\sum_x \sum_y (I(x, y) - \mu_I)(P(x, y) - \mu_P)}{\sigma_I \sigma_P} \quad (4-4)$$

对于图像中的每个候选标识, 找到与它互相关最大的模板图像, 若互相关的值大于指定阈值(如 0.5), 则认为图像中的候选标识为对应的模板标识, 解码过程完成。

3. 姿态估计

ARToolkit 标识物上 4 个角点的空间三维点齐次坐标 $(x_{wi}, y_{wi}, z_{wi}, 1)$ 与其在图像上的投影齐次坐标 $(u_i, v_i, 1)$ 满足

$$\lambda \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = K [r_1 \quad r_2 \quad r_3 | T] \begin{bmatrix} x_{wi} \\ y_{wi} \\ z_{wi} \\ 1 \end{bmatrix} \quad (4-5)$$

由于 ARToolkit 标识板是一个平面模型, 则其上的所有标识点位于同一平面上, 因此 $z_{wi} = 0$, 则式(4-5)可表示为

$$\lambda \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = K [r_1 \quad r_2 \quad r_3 | T] \begin{bmatrix} x_{wi} \\ y_{wi} \\ 0 \\ 1 \end{bmatrix} = K [r_1 \quad r_2 | T] \begin{bmatrix} x_{wi} \\ y_{wi} \\ 1 \end{bmatrix}$$

令

$$H = K [r_1 \quad r_2 | T] \quad (4-6)$$

式中, 3×3 矩阵 H 为将真实世界中 $z = 0$ 平面上的点映射到其投影图像上的单应性矩阵。 K 为摄像机内部参数, $[r_1, r_2]$ 为旋转矩阵 R 的前两列, T 为平移矢量。在摄像机内部参数 K 已标定的情况下, 由于标识点的四个顶点三维世界坐标系下的齐次坐标 $(x_{wi}, y_{wi}, 0, 1)$ 与其对应的图像坐标 $(u_i, v_i, 1)$ 已获得, 可以计算出单应性矩阵 H 。根据旋转矩阵 R 的正交归一化约束条件, 有 $r_3 = r_1 \times r_2$ 。由式(4-6)得到

$$K^{-1}H = [r_1, r_2 | T] \quad (4-7)$$

提取 $K^{-1}H$ 矩阵的前两列,进行叉乘运算可得到旋转矩阵的第三列 r_3 。至此,算法可以获得当前帧相对于真实环境的旋转和平移矩阵 $[r_1, r_2, r_1 \times r_2 | T]$ 。

ARToolkit 方形标识点的优点是简单易用,识别效率高,执行速度快,因此被广泛采用。但是由于 ARToolkit 处理过程采用二值图像,注册精度有限。同时在编码匹配方面采用了图形相关性匹配,匹配效率和编码数量都受到制约。

ARTag 标识是在 ARToolkit 的基础上进行的改进^[29],它有一个四方形的边缘,内部有 6×6 个黑白方块。它能够解码多达 2002 个标识点,其中的 1001 个标识点为白背景下的黑色边缘,另外 1001 个则相反,为黑色背景下的白色边缘。整个标识为 10×10 单位长度,边缘厚度为 2 单位长度,图 4.11 为该标识的例子。

ARTag 使用一种边缘启发式搜索算法,使其对光照和遮挡具有一定的鲁棒性。该编码系统具有较低的姿态估计误差和内部编码混淆率。

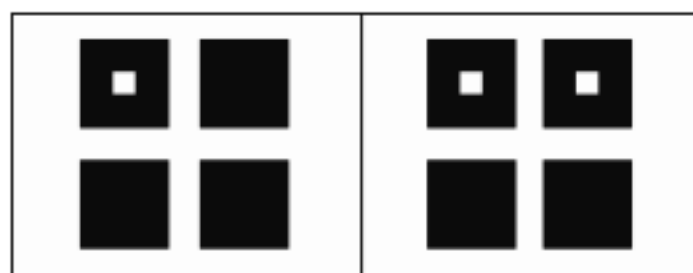
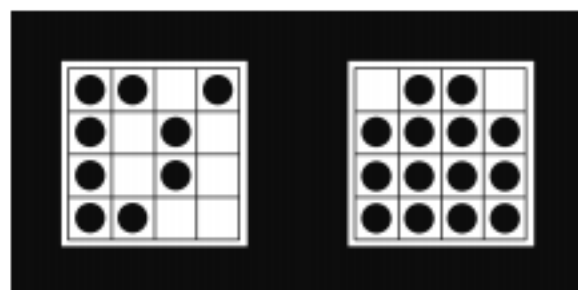


图 4.11 ARTag 标识的设计^[29]

4.2.2 SCR 标识点系统

西门子公司所属研究院的 Zhang 等对 ARToolkit 样式的标识点系统进行了重要改进,设计出一种方形的标识点系统(简称 SCR)^[30],并提出其相应的识别定位算法。在后续的工作中^[31,32],他们又改进了该标识点系统。该算法无论在精度、运行效率和准确性等方面均表现优异,对基于标识的 AR 注册技术贡献很大。

该算法中的标识点较容易被检测识别,从而可以实现精确定位。如图 4.12 所示,该标识板图像被设计为方形,底板中央设置有四个黑色正方形,其大小是预先确定的。从图 4.13 可以看到,一些黑色正方形的中央被设计为白色区域,这是为了便于计算机确定标识点的方向并区分这些标识而设计的。

图 4.12 标定和姿态估计所用的标识点^[30]图 4.13 中央被设计为白色区域的标识点^[30]

在良好的照明条件下,一般可使用阈值算法获得白色背景中的黑色正方形。而在较差的照明条件下,例如,当标识板平面将光线直接反射向摄像机,或者当平面处于较弱的灯光下时,常用的阈值算法将不能产生稳定的计算结果。为了解决这一问题,该系统利用 WAT(watershed adaptive thresholding)算法来跟踪标识点的运动。图像处理中使用的标识点变换始于 1979 年^[33]。在具体的执行过程中,可以使用自适应阈值,该阈值随工作区域内图像亮度的分配情况而发生变化。因此,可以消除由照明变化引起的不稳定性。

SCR 标识点识别定位算法中采用如下规则提取图像中的标识,即根据所选区域像素的亮度所决定的阈值对所选区域像素进行逐个比较。具体实现步骤如下。

- (1) 当一个像素的亮度高于阈值时,该像素将被标为“高”(覆以白色)。
- (2) 当一个像素的亮度低于阈值且该像素是一个边界像素时,该像素将被标为“湮没的”(灰色)。
- (3) 当一个像素的亮度低于阈值并且它周围至少有一个像素是“湮没的”时,该像素同样被标为“湮没的”(灰色)。
- (4) 当一个像素的亮度低于阈值且其周围没有一个像素是“湮没的”时,该像素将被标为“低”(黑色)。
- (5) 采用 WAT 算法输出的图像是三色图像(白,灰,黑),其中的四个黑色区域组成方形标识点。
- (6) 为了检测下一帧视频中的标识点,需要根据当前帧图像标识点的外接矩形进行实时更新。

应用 WAT 算法在每一帧图像中识别每个标识正方形的大致位置之后,SCR 算法采用以下步骤实现高精度的识别定位。

- (1) 估计单应性矩阵,完成标识板空间坐标系到图像坐标系的转换。
- (2) 作出 8 条黑色的直线,它们将形成每一个标识正方形的边缘,应用单应性

矩阵作出每一个标识正方形的边缘线。

(3) 采用 WAT 算法得到的边缘像素进行直线拟合。

(4) 使用一维 Canny 边缘检测算子,找出每一条直线垂直方向上的最大梯度。

(5) 由检测到的边缘像素重新拟合直线。

(6) 根据直线的交点估计特征点(标识块上的拐角或中心)的精确位置,从而进行精确的位姿恢复。

SCR 算法具有识别精度和准确率高、跟踪定位精度高、适用范围广的特点,是一种非常重要的识别定位算法。

4.2.3 环形标识点

较常采用的人工标识系统中,从最常见的 ARToolkit 标识点到高跟踪精度的 SCR 标识点都采用方形标识点的设计方式。Naimark 和 Foxlin 提出了独特的环形标识点系统及其相应的识别、编码与定位算法^[34],以下就该标识点系统进行详细介绍。

1. 标识点图形设计

图 4.14 所示为常用的三种环形标识,这种特别设计的环形标识具有 $2^{15} = 32768$ 种可能的编码量。



图 4.14 基于 101、1967、32767 的条形码^[34]

假设环形标识的直径是 $D=8u$,其中 u 为一个长度单位。若环形标识的直径为 6in , $1u=0.75\text{in}$ 。每一个标识点圆环均由三部分组成,最外层是宽度为 $1u$ 的黑色环,中部是宽度为 $2u$ 的数据环,内部是宽度为 $1u$ 的内环。该环形标识的设计遵循以下 3 个原则。

(1) 外部环的构造:该环总是黑色的。

(2) 内部环的构造:内部环的直径为 $1u$,且在环中央有一个直径为 $3/8u$ 白色圆。

(3)数据环的构造:数据环被均分为8部分,每一部分含一个 45° 角围成的扇形区域。

外部环、内部环、数据环这三部分均有其特殊的构造形式,其中,数据环中包含两个确定标识坐标系的黑色圆,位于距离中央 $17/18u$ 处,它们与内部环中的白色圆构成一个三角形。

2. 识别跟踪算法

图4.15举例说明了确定候选环形标识的具体步骤,该步骤中包括检测候选标识物中是否具有可读编码的过程。

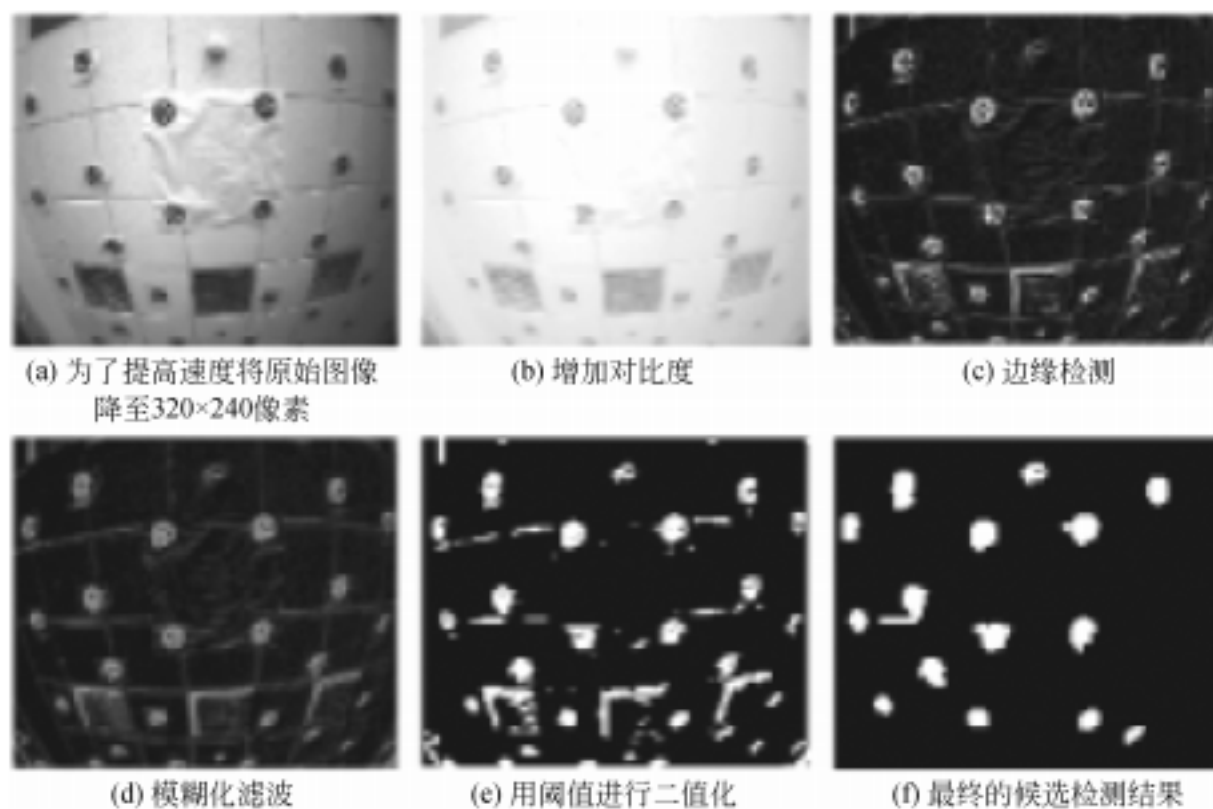


图4.15 搜索候选目标物体的步骤^[34]

在最终候选区域检测过程中,包括以下关键步骤。

(1)图像侵蚀:由于检测到的图像中可能存在许多面积较小的噪声区域。利用图像形态学的腐蚀和膨胀算法,可去除面积较小的噪声区域。

(2)特征提取:保留图像中 $10 \sim 80$ 像素大小的连通区域,并计算该区域的颜色、像素区域形状、标识点图像在水平垂直方向的最大和最小长度等。

(3)大小、颜色和区域检测。

最大区域:排除任一方向上大于80像素的连通区域。

最小区域:排除任一方向上小于 16 像素的连通区域,以确保图形编码被正确识别。

颜色:仅考虑白色物体。

区域比例测试:由于标识物是环状的,所以它们的映射图像将近似为椭圆形,这意味着区域面积 A 应近似满足

$$A = \frac{\pi}{4}(u_{\max} - u_{\min})(v_{\max} - v_{\min}) \quad (4-8)$$

式中, u 和 v 分别是连通区域在 x 轴、 y 轴方向上的长度。利用该约束关系可以忽略那些边缘窄的和 L 形的区域。经处理后获得的图像如图 4.15(f) 所示。经过以上步骤,标识区域的检测正确率可达到 90%。

3. 编码识别算法

完成筛选候选标识物后,下一步将对每个候选标识物的编码进行识别。若二值化的图像为一个黑色的圆,且圆内部包含两个黑色的圆点和一个白色的圆点,那么该物体将被确认为标识物体。其中位于圆形中部的白色圆点被当成图形坐标系的原点,而分别指向黑色圆点的两个向量被当成 x 轴和 y 轴。针对标识物的设置,共有 8 个类梯形区域,每个区域包含 3 个测试点(图 4.16)。可以通过读取测试点的二元值进行标识物解码。

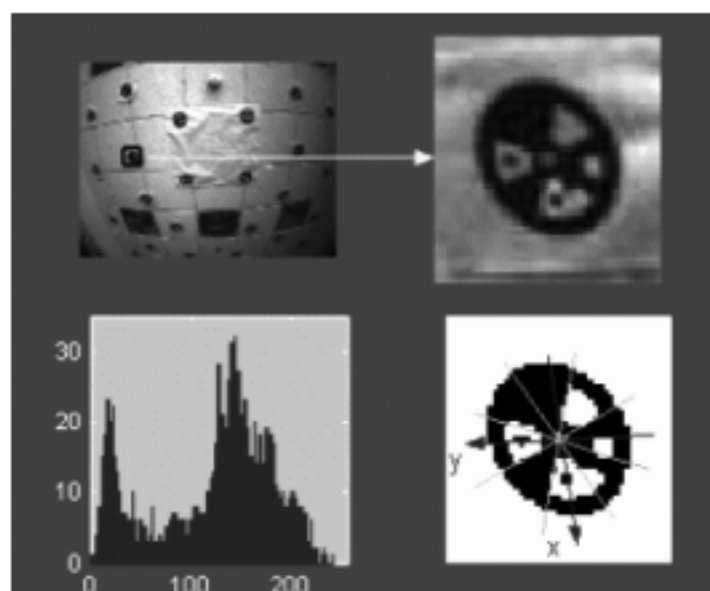


图 4.16 条形码解读步骤^[34]

完成以上过程大约需耗时 250ms,对于实时跟踪而言却是不够的。因为高刷新率和低延时是跟踪算法中的重要技术指标。该算法的完整跟踪过程如图 4.17

所示。算法首先将提取的灰度图像降至 320×240 像素,然后增加图像的对比度,并进行边缘检测选取适当的阈值对图像进行二值化处理。在二值化过程中,也许存在大量与检测标识类似的黑色像素区域,以及一些小噪声点。算法利用该区域的面积、颜色以及区域相似检测选择候选区域。完成以上步骤后,标识物将被准确定位,同时计算出它的重心。

这类标识内部结构复杂,当其面积过小时不容易被识别。因此,在实际应用中,通常采用大量张贴该类标识点的方法,在跟踪过程中进行多次迭代比较,从而增加跟踪精度。其缺点在于响应速度较低。

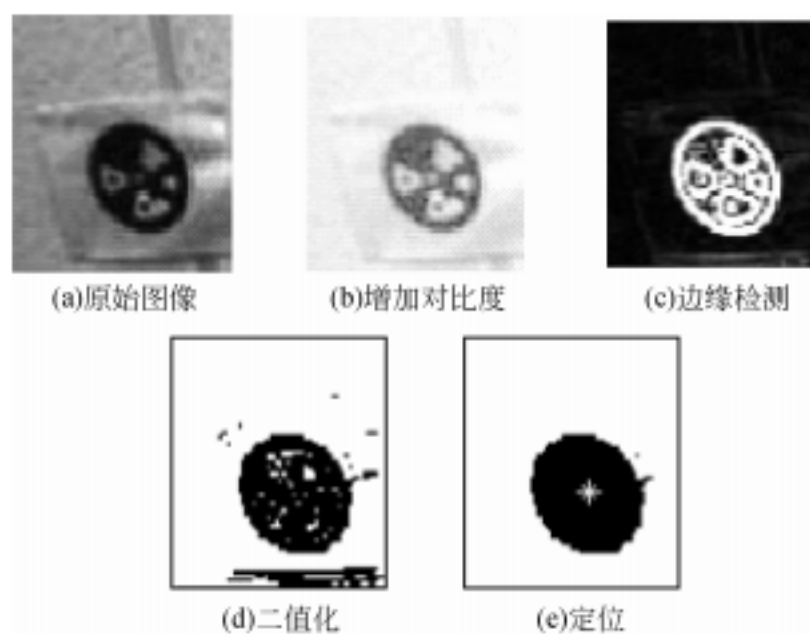


图 4.17 跟踪的步骤^[34]

4.2.4 Sony 计算机实验室标识点

Sony 计算机实验室 1998 年发表了一种与 ARToolkit 比较类似的标识点系统^[35],此系统主要对编码算法进行了改进,可以更方便地设计编码。标识物采用图 4.18 所示的图形,外部为黑色边框,内部与 ARToolkit 系统不同,采用黑白方块构成的二维编码图形。该编码方式一般称为二维条形码(2D barcode)。编码图形中的每一个方格可以看作一个二进制位,经过对每个方格颜色的二值化可以得到表示此标识物编码的二进制数。由于此标识系统的编码加入了 CRC 校验,从而可在一定程度上避免误差带来的影响。

需要注意的是,此编码中需要特定的标记来保证标识图形的不对称性,从而确保可正确识别角点特征的顺序。该标识系统的优点是可以根据实际应用时的图像



图 4.18 Sony 标识物图形设计^[35]

分辨率调整编码的数量,从而得到最合适的编码量。图 4.18 中的标识点可以达到的编码数量为 $2^{25-4} = 2^{21}$ 个。该标识系统的识别定位算法与 ARToolkit 大致相似(图 4.19)。其解码算法可以很容易地确定标识物的朝向,因此效率大大高于 ARToolkit 的解码算法。

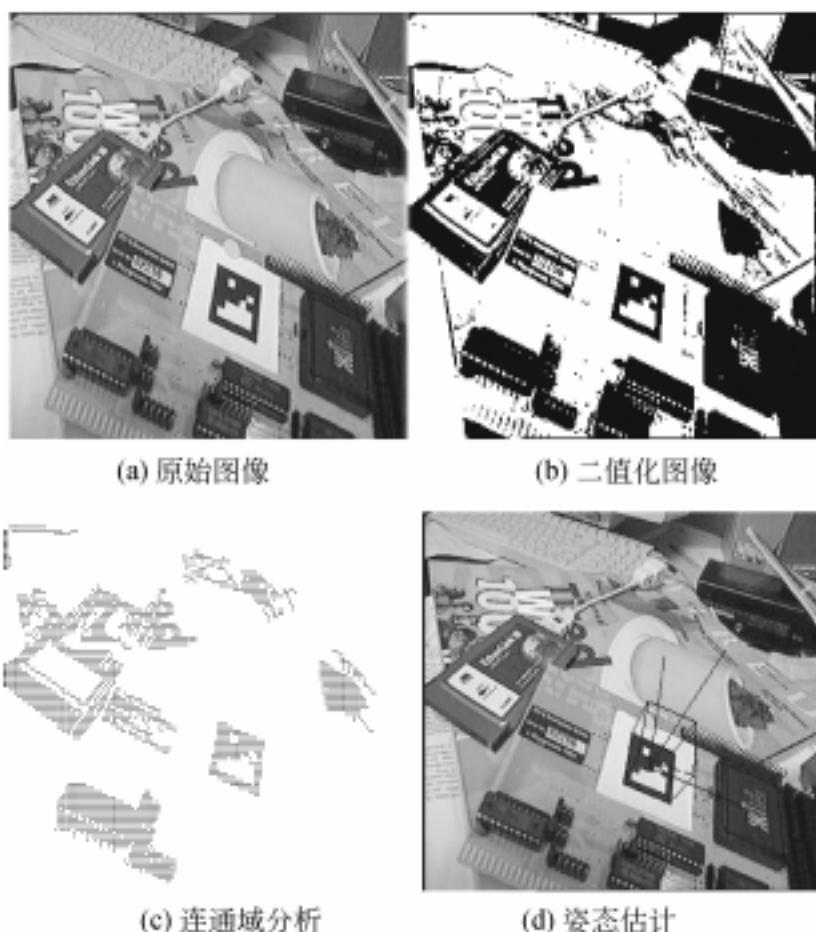


图 4.19 识别定位算法^[35]

4.2.5 室内装饰性人工标识

前面提到的可见标识一定程度上破坏了环境的美观性,同时可能分散使用者

的注意力。而不可见标识,如红外标识则需要大量的硬件设备支持。Saito 等^[36]提出增强现实系统中的跟踪技术应满足四个要求:①使用方便;②价格低廉;③不破坏环境的美观性;④在任意大小环境下有效。基于以上四点,他们设计了一种具有室内装饰效果的人工标识系统。与以往不同的是,他们设计的人工标识隐藏在室内装饰的墙纸、地板或天花板上。

1. 标识点图形设计

图 4.20 展示了一种人工标识图案的样式,像这样的图案可以被当成装饰壁纸贴在墙上、地板上或天花板上,而不易被用户察觉。图 4.20(a)包含四种不同的图形,它们按照不同的倾斜角排列在系统图案中,如图 4.20(b)所示。并且每四个不同的图形组成一组,构成一个标识,每个标识中的四个图形的排列方位信息包含编码信息,并且这四个图形各自的质心为跟踪注册提供了位置信息。

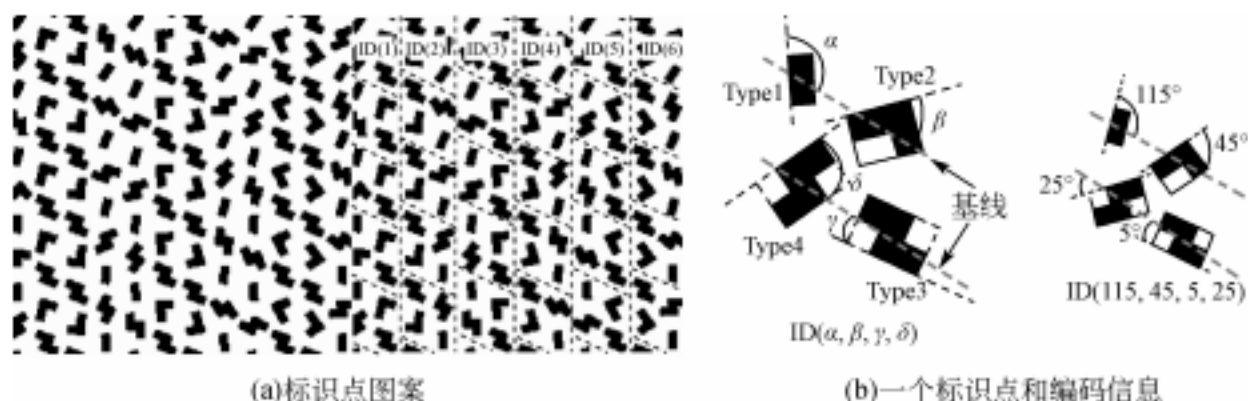


图 4.20 人工标识图案^[36]

将图 4.20(a)中出现的四种不同的图案分别称为 Type1、Type2、Type3 和 Type4。图案的旋转角(rotational angle)用基线(base line)到能够覆盖图形最小的长方形的长边所在直线的夹角确定。其中基线为连接两个图形质心的直线[如图 4.20(b)中灰色虚点线所示]。因为某些图案(如菱形)可能有不止一个最小的长方形能够覆盖它,所以这样的图案对标识是没有意义的。为了使这样的图案有意义,它们的旋转角可以用基线到指定的一条边或者主轴(如长轴)的角来确定。旋转角按顺时针顺序构成的矢量($\alpha, \beta, \gamma, \delta$)将作为标识的 ID 值。

2. 标识的编码量

因为该标识利用图案的旋转角向量作为编码信息,所以该标识系统可以由很少的图形构造出大量不同的人工标识。理论上由于角度可以连续取值,由少量不

同的图案就可以提供无数个标识。但是受噪声和计算精度的影响,摄像机只能准确分辨出相差一定度数的不同角度。因此,需要准确知道摄像机能够达到的角度分辨率。实验数据如表 4.1 所示。

表 4.1 标识点角度不同时的错误识别率,0°指摄像机主轴与标识点垂直,45°指摄像机主轴与标识点平面成 45°角,“运动”指摄像机做不规则运动,测试距离为 1000mm

角度差别/(°)	识别错误率		
	0°	45°	运动
5	0.9%	14.0%	44.9%
10	0.5%	1.1%	7.2%
15	0%	0.1%	1.9%
20	0%	0.1%	1.2%

从表 4.1 可以看出,当摄像机运动时识别错误率较高。当标识旋转角度间的差别大于等于 10°时,识别错误率大大降低。中心对称的长方形图案有 18 种不同的摆放方式,因此,图 4.20 中标识物的个数可达 $18^4 = 104976$ 。为进一步降低识别错误率,该标识系统引入了纠错算法。在图 4.21 中,可将 Type4 作为纠错图形,事先约定其在每一个标识中的朝向都是一样。只有当识别的旋转角与事先确定的旋转角度一致时才认为识别准确。这样,图 4.20 中标识的个数降为 $18^3 = 5832$ 。假如每个标识的大小是 100mm×100mm,那么此类图案可以覆盖 58m² 的地面。如果加大图形之间的距离或增加其他的图形,很容易就可以形成更大面积的图案。考虑到美学上的要求,标识图案可以设计为如图 4.21 所示的类型。

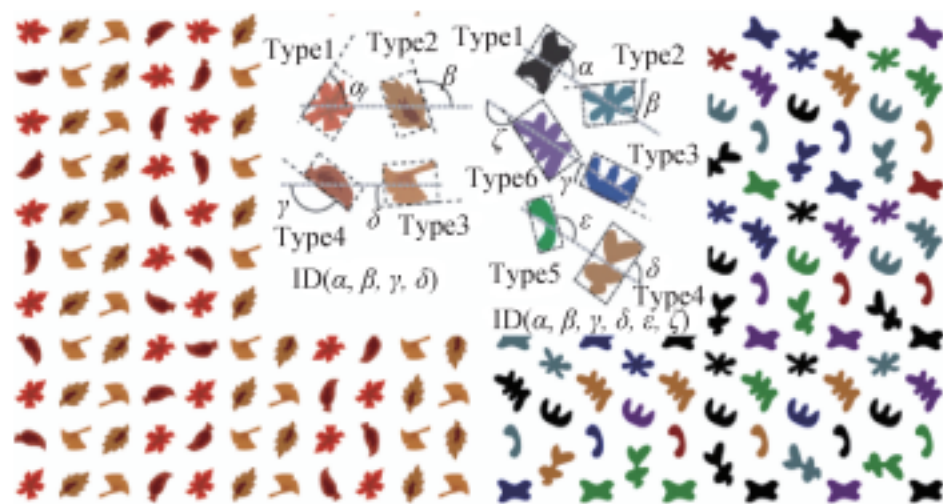


图 4.21 两种更美观的标志类型图案^[36]

3. 基于室内装饰标识的特点

(1) 隐蔽性。标识点隐藏在图案中。利用图形的旋转角向量作为编码信息,利用图形的位置信息计算摄像机的六自由度姿态。这样可以使很多装饰图案可以用来做标识。

(2) 易于设计。利用图形排列算法可以很容易地设计图案,避免了人工标识必须有特殊设计图案的局限性,能够很简便地设计出大量的标识图案。

(3) 跟踪范围大。利用图案中图形的旋转角作为 ID,一个图形就可以提供十个以上的 ID。只要增加图形的数量就可以指数级地增加 ID 的个数。因此,可以提供足够大的标识图案装饰任意大小的室内空间。

(4) 价格低。由于该算法是基于人工标识的跟踪注册算法,因此其具有其他人工标识算法的优点。同时价格低,能够供多人同时使用该系统。

4.2.6 红外标识系统

1999 年北卡罗来纳大学的 Welch 等提出了一种基于红外光学跟踪的大范围跟踪系统——Hi-Ball 跟踪系统^[1,9]。该系统由 Hi-Ball 光学传感器、红外 LED 阵列以及 CIB(ceiling hi-Ball interface board)组成,如图 4.22 所示。Hi-Ball 光学传感器由 6 个镜头和光电二极管排列组成,光电二极管用于接收安装在天花板的红外 LED 发出的红外信号。红外 LED 阵列上的每个 LED 按照一定的时间规律发射红外线实现其自身的编码。该系统覆盖范围非常大,几乎可以不受限制。在整个跟踪空间内均可保持高精度跟踪,且不受金属、磁场、噪声等因素的影响。也可提供非常高的刷新率和低延时,即使在高速度运动下也可进行实时跟踪定位。但是,该系统结构复杂、价格昂贵。

为减少对环境的侵入性,Nakazato 等提出了一种半透明回归反射标识系统(invisible retro-reflective markers)^[37]。反射器反射红外 LED 发出的红外光,可被红外摄像机获取。

1. 标识图案设计

标识的样式可以采用前面提出的各种可见光标识样式,这里采用 Sony 公司提出的标识点样式,如图 4.23 所示,其中黑色部分表示回归反射器。为了捕获到标识点的图像,需要使用红外 LED 和红外摄像机(图 4.24)。

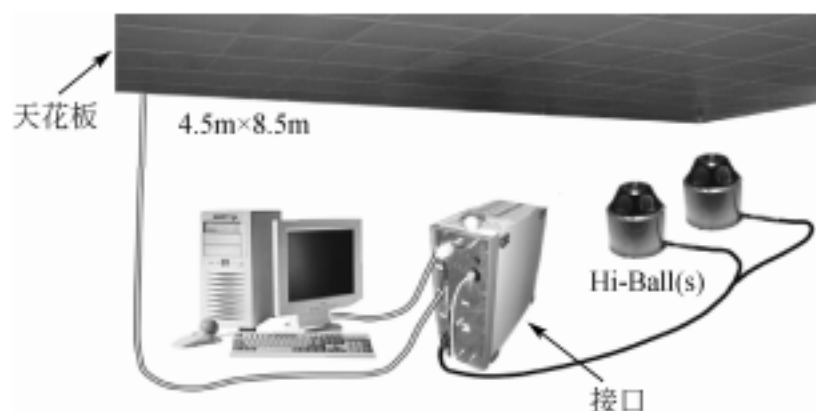


图 4.22 Hi-Ball 光学跟踪系统^[1,9]

2. 标识的识别

因为系统用的摄像机不仅能捕获红外 LED 发出的光线,同时能捕获可见光。为了更好地利用图像信息,在进行标识识别时,由计算机控制红外 LED 的开关。摄像机与红外 LED 的开关保持同步,捕获可见和红外两种环境下的图像,如图 4.25 所示。由于 LED 开关的时间间隔极短,因此可以合理地认为场景没有发生变化,且摄像机的姿态也未发生变化。

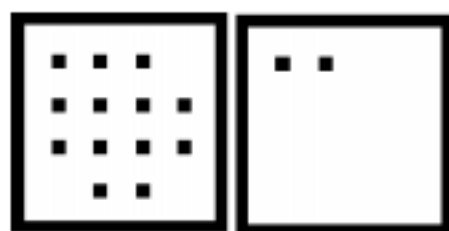


图 4.23 标识点示例^[37]

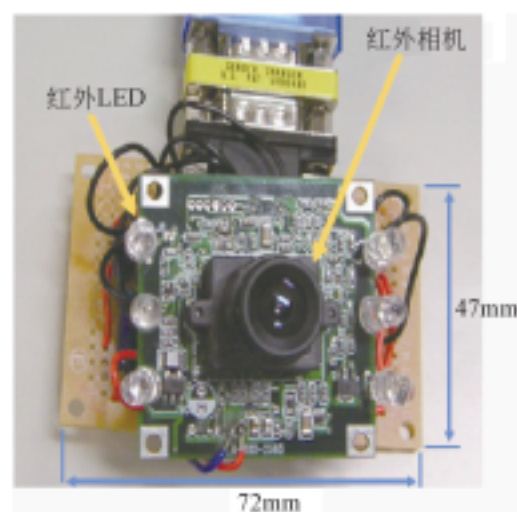
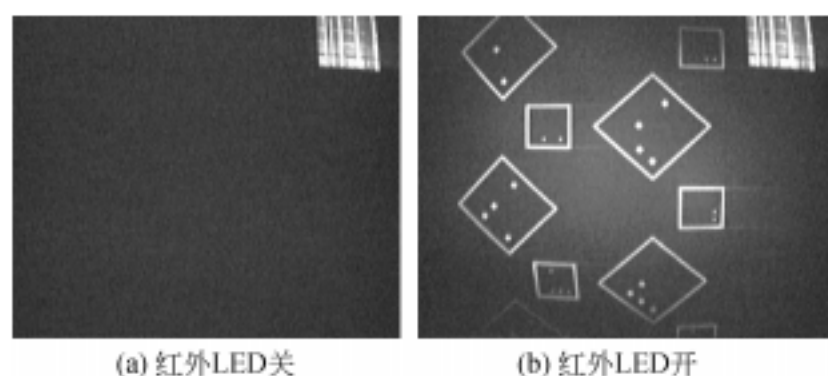
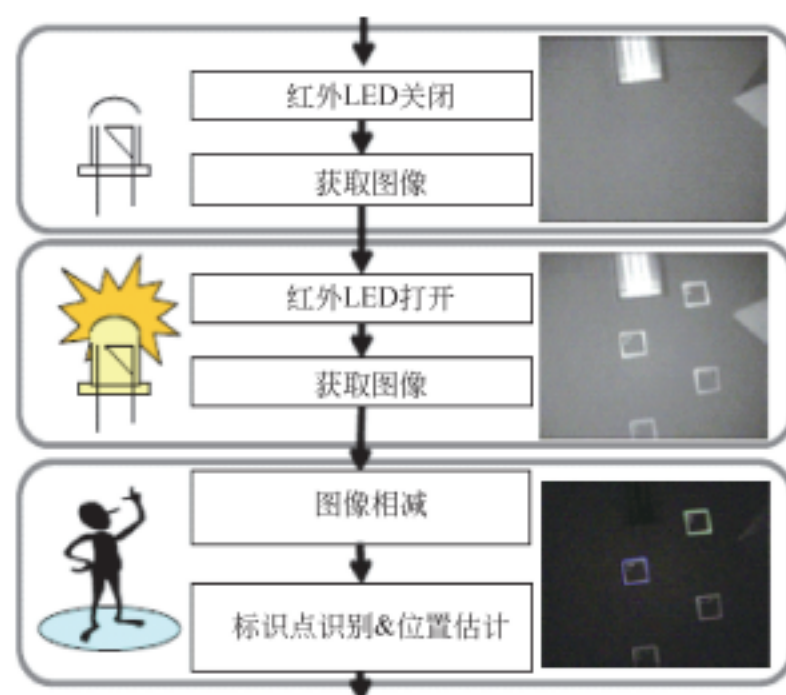


图 4.24 红外摄像机和红外 LED^[37]

当红外 LED 关闭时,摄像机捕获的图像中看不到标识点。当红外 LED 打开时,摄像机捕获的图像中可以很清楚地看到标识点,如图 4.25(b)所示。用图 4.25(b)减去图 4.25(a)可以消除非红外 LED 发出的光线对回归反射标识点以及环境的影响。此后可以利用 ARToolkit 的编/解码技术对标识点进行提取和识别,进而对摄像机六自由度姿态进行估计。其算法流程图如 4.26 所示,算法步骤如下。

图 4.25 红外摄像机捕获的图像^[37]

- (1) 关闭红外 LED, 捕获图像 1。
- (2) 打开红外 LED, 捕获图像 2。
- (3) 用图像 2 减去图像 1。
- (4) 利用 ARToolkit 标识点系统中提到的标识点识别算法识别标识点。
- (5) 摄像机姿态估计。

图 4.26 算法流程^[37]

4.3 各种人工标识系统中的识别定位算法比较

前面介绍了目前常见的各种人工标识系统及其相应的识别定位算法。各种算

法的跟踪定位效果与标识物的图形设计是密切相关的,好的人工标识首先应具有优秀的图形样式设计方案。本节将从标识的形状、颜色、定位、识别、内部图形设计等方面进行详细比较和讨论,并对各种标识的设计方案进行综合比较分析。

4.3.1 人工标识的形状

在场景中放置人工标识的目的是能够快速获得标识上特征点的三维空间坐标与其投影至二维图像上的图像坐标间的对应关系。确切地说,任何一个视觉特征在其位置已知(或能被计算出来)的情况下,都能作为一个标识被自动识别。为确定摄像机相对于标识物的空间位置和方向,要求计算机至少能够识别摄像机所拍摄图像中的三个非共线特征点。事实上,若只利用三个特征对应点计算摄像机的六自由度位姿,其计算结果存在二义性。通常情况下,方程求解可得到三个或四个解。若能够在摄像机拍摄的图像中识别出四个共面点,且已知它们在世界坐标系中的坐标,则可唯一确定摄像机的六自由度位姿。由此可见,一个理想的标识物图形应该至少具有四个可识别的特征。这并不意味着标识点图形必须是方形的。标识的图像形式越简单,其特征点识别提取的计算量越小,就越具有计算优势。方形标识是能够从其上提取四个特征点的最简单的图形形式,而且可以采用最佳匹配算法拟合图形的四条直线边缘,利用直线相交的交点作为特征点。该识别算法能够达到亚像素级精度,如 ARToolkit 的标准标识图形就是一个正方形。

在此需要指出的是,如果能够从圆形标识上提取足够的特征点,圆形标记也可以用来确定摄像机的六自由度姿态。POSE_FROM_CIRCLE 算法就提供了一种鲁棒性的解决方案,该算法能够提供圆形边界上的特征点^[38,39]。

4.3.2 标识的颜色

彩色标识相较于单色标识可增加标识的信息量,但是事实上,许多技术方面的因素决定了采用单色标识更加适宜。

- (1) 不同的摄像机系统具有的色度分辨率不同。
- (2) 单色可简化图像的表征。
- (3) 单色图形便于实现高效定位。

一般而言,人类视觉系统对亮度分量的空间频率灵敏度高于色度分量的灵敏度。不幸的是,许多计算机为设计的图像系统模拟这种特性,在低带宽通道中传送色度信息或用更低的分辨率重现色度信息,这样做的结果势必会减弱彩色标识的分辨率。即使完全捕获由 RGB 颜色表示的图像信息,所得到的彩色图像也将会占

用更多的存储器空间。从而,使得提取识别特征点的运行时间增加3倍(或4倍)。

另外一个需要考虑的问题是识别定位算法的准确性。采用彩色标识作为标识物的优势是,能够利用它的颜色信息区分特定的标识。但是颜色的判断在很大程度上取决于环境光照的条件,某些方向的反射光线能够改变映射颜色的色度。因此,利用颜色信息进行跟踪定位存在精度问题。

4.3.3 定位标识区域

标识的形状和颜色的设计与标识的定位算法直接相关。例如,ARToolkit 标识是一个具有黑色边框的方形单色图案,方形边框内的特殊图案可用于计算机对标识进行识别。在将标识的灰度图像转换为二值图像时,要求标识区域与背景环境存在差异,以便进行标识点区域分割。显然,该标识设计方案具有以下几个非常明显的优势。方形的形状设计为提取、跟踪正方形的四个角点提供了便利,而黑色的边框与背景形成最大的反差。一旦角点被确定,边界直线内部的标识图案就可以通过编码或图像匹配等方法来进行唯一性识别。

4.3.4 标识的可识别范围

标识的可识别范围依赖于摄像机的分辨率和内部参数。一些人工标识系统的设计允许在不同尺度上进行标识的定位识别,这样就使得摄像机可以在更大范围内移动。在一定的空间区域布置多个不同大小的标识点,也可以达到相同的效果。但是,一个标识图像在多小时仍然可以被正确识别呢?这里的大小主要指摄像机采集图像中的标识物的相对大小。选择 16×16 像素大小的图像用于实验,实验结果显示任意轴向上最小的可识别图像元必须为 16 像素。在给定这个标准之后,相应的问题也就出现了:边界要多宽才合适?事实上,为确保可信的轮廓位置,边界必须足够宽。以保证较远距离拍摄时,边界可以被完全拍摄到。如果边界过于狭窄,则会使像素变得过于灰暗,而不能进行正确的边界匹配,所以边界的宽度必须大于任何两个像素之间的距离。这个距离实际上是 2.83 像素。因为两像素间的最短距离存在于对角线上,其大小为 1.41 像素。因此,标识点图像至少应有 $16 + 2.83 + 2.83 = 21.66$ 像素的宽度才能作为可识别的图像。为了设计的目的,边界必须占标识点宽度的 13%,也可适当地选取 15% 的边界宽度。但是,也要注意边界宽度应该被保持在最小,以便增加内部图像的大小,这样才能确保进行大范围的识别。

4.3.5 特征点提取

人工标识最主要的功能是提供特征点的对应点集,因此标识图形还必须要有可以精确提取的特征点。这些特征点要尽可能便于提取,并且提取精度要尽可能高。由于人工标识系统中的特征往往是预先设计的,而且往往是直线交点、图形拐点等射影不变的元素。这样特征的提取就会更具针对性,更加精确。特征点提取常常采用如直线拟合、弧拟合、取重心等算法,设法得到亚像素级的特征位置。

4.3.6 标识的编码信息量

人工标识的编码算法是区别标识系统优劣的一个重要指标。标识图形首先要保证能够正确区分不同的标识。也就是说,要尽量减小将一个标识错误地识别为其他标识的概率。同时,要减小非标识图形被识别为标识的误报率和识别不出标识的漏报率。由于图像采集过程不可避免的误差和图像分辨率的限制,标识图形在采集图像中可能存在一定的误差和变形等,这就需要标识具有容错能力。此外,在保证上述识别能力的情况下,标识的编码空间要尽可能大。也就是说,允许的标识编码数量要尽可能多。当然允许的编码数量是与正确识别标识图形的概率有关的,要共同考虑。

4.3.7 理想特征的概要

综上所述,选择理想标识物的标准应综合考量以下几方面:标识的图像必须在摄像机的大视场范围内有效;标识的图像应该能够提供唯一确定摄像机位置和方向的特征信息,而不存在二义性;标识图像不应该在某些方向占有优势;标识应具有较大的编码信息量,确保大空间范围内的标识均可被唯一地标记识别;标识的设计应能够较为容易地实现图像定位,同时能够使用简单快捷的算法进行识别。

参 考 文 献

- [1] Welch G, Bishop G, Vicci L, et al. High-performance wide-area optical tracking; the hiball tracking system. Presence: Teleoperators and Virtual Environments, 2001, 10(1): 1 ~ 21.
- [2] Bishop T G. Self-tracker; A Smart Optical Sensor on Silicon. Chapel Hill: University of North Carolina, 1984.
- [3] Wang J F. A Real-time Optical 6-D Tracker for Head-mounted Display Systems. Chapel Hill: University of North Carolina, 1990.

- [4] Wang J F, Azuma R T, Bishop G, et al. Tracking a head-mounted display in a room-sized environment with head-mounted cameras//Proceedings of the Helmet-Mounted Displays II, Orlando, 1990.
- [5] Ward M, Azuma R, Bennett R, et al. A demonstrated optical tracker with scalable work area for head-mounted display systems//Proceedings of the 1992 Symposium on Interactive 3D Graphics, Cambridge, 1992.
- [6] Azuma R T. Predictive Tracking for Augmented Reality. Chapel Hill: University of North Carolina, 1995.
- [7] Welch G. SCAAT: Incremental Tracking with Incomplete Information. Chapel Hill: University of North Carolina, 1996.
- [8] Welch G, Bishop G. SCAAT: Incremental tracking with incomplete information//Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, Los Angeles, 1997.
- [9] Welch G, Bishop G, Vicci L, et al. The hiball tracker: High-performance wide-area tracking for virtual and augmented environments//Proceedings of the ACM Symposium on Virtual Reality Software and Technology, London, 1999.
- [10] Ferrin F J. Survey of helmet tracking technologies//Proceedings of SPIE, San Jose, 1991.
- [11] Kim D, Richards S W, Caudell T P. An optical tracker for augmented reality and wearable computers//Proceedings of IEEE Annual International Symposium on Virtual Reality, Albuquerque, 1997.
- [12] Burton R P, Sutherland I E. Twinkle box: Three-dimensional computer-input devices//Proceedings of the National Computer Conference, Chicago, 1974.
- [13] Burton R P. Real-time Measurement of Multiple Three-dimensional Positions. Salt Lake: University of Utah, 1973.
- [14] Woltring H J. New possibilities for human motion studies by real-time light spot position measurement. Biotelemetry, 1974, 1(3): 132 ~ 146.
- [15] United Detector Technology. Trade literature on OP-EYE Optical Position Indicator. Santa Monica: United Detector Technology, 1981.
- [16] Macellari V. CoSTEL: A computer peripheral remote sensing device for 3 dimensional monitoring of human motion. Medical and Biological Engineering and Computing, 1983, 21(3): 311 ~ 318.
- [17] Cook A M. The helmet-mounted visual system in flight simulation//Proceedings of Flight Simulation: Recent Development in Technology and Use, London, 1988.
- [18] Janin A L, Zikan K, Mizell D, et al. Videometric head tracker for augmented reality applications//Proceedings of Telemanipulator and Telepresence Technologies, Boston, 1994.
- [19] Ribo M, Pinz A, Fuhrmann A L. A new optical tracking system for virtual and augmented reality applications//Proceedings of the 18th IEEE Instrumentation and Measurement Technology

- Conference, Hungary, 2001.
- [20] Dorfmüller K. Robust tracking for augmented reality using retroreflective markers. *Computers & Graphics*, 1999, 23(6): 795 ~ 800.
- [21] Dorfmüller K. An optical tracking system for VR/AR applications//*Proceedings of the Virtual Environment '99*, New York, 1999.
- [22] Dorfmüller K, Wirth H. Real-time hand and head tracking for virtual environments using infrared beacons//*Proceedings of the CAPTECH '98*, Geneva, 1998.
- [23] Vicon Inc. [Http://www.vicon.com](http://www.vicon.com)[2012-10-10].
- [24] A. R. T. GmbH. [Http://www.ar-tracking.de](http://www.ar-tracking.de)[2012-11-16].
- [25] Motion Analysis Inc. [Http://www.motionanalysis.com](http://www.motionanalysis.com)[2012-12-10].
- [26] Lepetit V, Moreno-Noguer F, Fua P. Epnp: An accurate solution to the pnp problem. *Journal of Computer Vision*, 2009, 81(2): 155 ~ 166.
- [27] ARToolkit. [Http://www.hitl.washington.edu/ARToolkit](http://www.hitl.washington.edu/ARToolkit)[2012-12-20].
- [28] Billinghurst M, Kato H, Poupyrev I. The magicbook- moving seamlessly between reality and virtuality. *Computer Graphics and Applications*, 2001, 21(3): 6 ~ 8.
- [29] Fiala M. ARTag: A fiducial marker system using digital techniques//*Proceedings of the Conference on Computer Vision and Pattern Recognition*, San Diego, 2005.
- [30] Zhang X, Fronz S, Navab N. Visual marker detection and decoding in AR systems: A comparative study//*Proceedings of the Symp on Mixed and Augmented Reality*, Darmstadt, 2002.
- [31] Zhang X, Navab N. Tracking and pose estimation for computer assisted localization in industrial environments//*Proceedings of the 5th IEEE Work-shop Applications of Computer Vision*, Piscataway, 2000.
- [32] Zhang X, Gene Y, Navab N. Mobile computing and industrial augmented reality for real-time data access//*Proceedings of the 7th IEEE International Conference on Emerging Technologies and Factory Automation*, 2001.
- [33] Beucher S, Lantuejoul C. Use of watersheds in contour detection//*Proceedings of International Workshop on Image Processing, Real-Time Edge and Motion Detection/Estimation*, Rennes, 1979.
- [34] Naimark L, Foxlin E. Circular data matrix fiducial system and robust image processing for a wearable vision-inertial self-tracker mixed and augmented reality//*Proceedings of the ISMAR*, Darmstadt, 2002.
- [35] Rekimoto J. Matrix: A real time object identification and registration method for augmented reality//*Proceedings of the 3rd Asia Pacific Computer Human Interaction*, Hayama-machi, Kanagawa, 1998.
- [36] Saito S, Hiyama A, Tanikawa T, et al. Indoor marker-based localization using coded seamless pattern for interior decoration//*Proceedings of the IEEE Virtual Reality Conference*,

Charlotte, 2007.

- [37] Nakazato Y, Kanbara M, Yokoya N. Localization of wearable users using invisible retro-reflective markers and an IR camera//Proceedings of the IAPR Conference on Machine Vision Applications, Tsukuba Science City, 2005.
- [38] Forsyth D, Mundy J L, Zisserman A, et al. Invariant descriptors for 3D object recognition and pose. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1991, 13(10) :971 ~ 991.
- [39] Trucco E, Verri A. Introductory Techniques for 3-D Computer Vision. Englewood Cliffs; Prentice Hall, 1998.

第5章 无标识增强现实三维跟踪注册

早期的AR跟踪定位技术研究主要针对人为可控的室内环境,采用人工放置标识的方式实现目标跟踪定位,主要采用ARToolkit、ARTag、SCR或环形标识点等人工标识。通过改变人工标识内部的编码图案,可实现多目标的跟踪定位。该方法具有跟踪定位精度高,实现简单的优点。然而在真实环境中放置人工标识物,利用标识物进行跟踪定位无法避免遮挡以及环境光照变化所带来的影响,且影响周围环境的美观与协调。只适用于光照变化可控、场景较为固定的场合。与人为可控环境相比,复杂环境中存在许多随时间变化的因素和噪声,如光照强度随时间的变化、人或运动物体对场景画面带来的改变,以及特征点被遮挡或短暂地移出视野之外的时候,都会在不同程度上影响跟踪定位的精度和稳定性。

随着AR技术的发展以及应用领域的拓宽,要求AR系统实现户外大范围场景的无标识跟踪定位(markerless tracking)。此时,系统必须完全利用自然场景图像确定用户当前的位置和视线方向。无标识AR跟踪定位技术的研究对于AR系统真正用于实际工作具有举足轻重的意义。国内外该领域的知名研究机构也多以无标识的AR跟踪定位作为增强现实核心技术研究的主要内容,并提出了各种跟踪定位算法,以满足实际应用的要求。目前,主要的无标识跟踪定位方法包括基于关键帧匹配的跟踪定位、基于模型的跟踪定位、基于虚拟伺服控制理论的跟踪定位、基于主动重建的跟踪定位以及混合跟踪定位方法等。

基于关键帧匹配的跟踪定位算法是利用二维图像的匹配和分析实现增强现实的跟踪定位算法。此类算法的核心是选择与当前帧视点最接近的关键帧,利用当前帧与关键帧图像间的特征匹配计算出两帧间的姿态变化差异,构建两幅图像间特征的2D/2D匹配是算法的核心。

基于模型的跟踪定位算法一直以来被广泛应用于AR跟踪系统中。该方法要求对目标物体或场景事先建模,根据每帧中得到的目标特征2D投影图像与其在空间中的3D坐标间的对应,求解目标或摄像机的姿态。目前基于模型跟踪定位的方法主要包括基于边缘的方法和基于兴趣点的方法等。总体上,基于模型的跟踪定位技术能够较好地实现复杂环境下的目标跟踪定位,但是也存在建模工作量大而烦琐、目标优化容易陷入局部最小的问题。

基于虚拟伺服控制理论的跟踪定位方法是2002年由法国IRISA实验室的研究人员将机器人控制领域的视觉伺服控制理论引入摄像机跟踪和姿态估计研究领域所开创的^[1]。假设存在一个与实际摄像机相对应的虚拟摄像机,并需要预先获得应用场景的三维几何模型。基于虚拟视觉伺服的方法在一定程度上解决了无人工标识的虚实场景跟踪定位问题,但是该方法存在相当明显的缺陷。即系统需要在人工干预的情况下实现第一帧图像的配准,因而只适用于特定的、较为简单的人工场景,目前无法实现多目标的三维跟踪定位。

基于主动重建的跟踪定位方法的工作原理是,将场景重建和摄像机追踪两个原本相对独立的任务关联起来。摄像机的位姿信息除了用来完成虚实配准之外,还将用于未知场景的在线重建,而场景重建结果也将实时反馈给追踪系统,用于后续的摄像机位姿估计和虚实配准。正是这种相互关联的工作机制带来了预重建跟踪定位方法所不具备的灵活性和实用性。英国牛津大学研制的基于单目视觉的场景在线重建与摄像机追踪系统 MonoSLAM,可成功解决未知场景下的增强现实虚实注册问题^[2]。然而,采用滤波技术来解决场景优化问题,只能对数百个自然特征点进行在线重建和追踪,很大程度上影响了系统的工作范围。Klein 等利用从运动中恢复结构(structure from motion)技术研制了 PTAM 系统^[3],用以解决场景在线重建问题,系统能够有效完成小范围内场景的在线结构恢复与摄像机追踪问题,但是大尺度场景的场景结构恢复与摄像机追踪问题仍有待解决。

本章就几种具有典型代表意义的无标识跟踪定位算法作详细介绍。

5.1 基于场景平面的增强现实跟踪注册

2000年,Simon 等提出一种基于场景中平面结构的无标识点增强现实跟踪注册算法^[4]。虽然,基于平面结构的无标识点跟踪注册算法在一定程度上解决了无人工标识点情况下的虚实场景注册问题,但是该方法也存在相当致命的缺陷。即在注册时,必须首先在人工干预的情况下将第一幅图像配准。同时由于算法自身的原因,在跟踪注册过程中不可避免地会产生累积误差,因此这种算法只能在很短的一段时间内正常工作,经过一段时间后必须使用其他方法去除累积误差。

5.1.1 系统初始化

基于平面结构的实时无标识点增强现实跟踪注册算法不能实现系统参数的自动初始化,需要在初始帧时对场景图像进行摄像机定标、手工圈定平面区域以及初

始 3D/2D 单应性矩阵的求解等一系列初始化工作。

(1) 摄像机标定。该算法使用的是内部参数已知的摄像机,对于大多数应用系统来说这并不难实现。摄像机标定所要完成的主要任务是标定摄像机的内部参数矩阵 K ,可以选用第 3 章所提到比较成熟的摄像机标定方法来事先标定摄像机的内部参数。

(2) 手工圈定平面区域。该算法建立在对场景平面区域的检测和持续追踪的基础之上。因此,需要在算法运行的初始阶段,手工圈定平面区域的大致轮廓。该过程可通过一系列鼠标动作有效地圈定出一个平面范围并提供给系统。

(3) 初始单应性矩阵 H_w 的求解。由于该算法是建立在对摄像机的完全认知以及对场景空间结构的简单理解之上的,因此可定义世界坐标系的 xoy 平面与目标平面重合, z 轴垂直于该平面。此时,空间平面上的三维空间点 $(x,y,0)$ 与其在摄像机投影图像上的对应点 (u,v) 的关系可以用 H_w 描述。单应性矩阵不存在条件数过大的问题,只需要提供四个准确的对应点就可以稳定地求解。因此,算法是根据对场景的简单理解,在首帧图像上手工选取四个已知空间坐标的三维空间点并记录其相应的图像点,按照直接线性解法求解 H_w 。

5.1.2 摄像机位姿估计

初始化过程完成后,即可进行后续帧的摄像机位姿实时计算。首先,考虑算法初始化后的首帧图像,即第二帧图像的情况。若能够建立第二帧图像上的 2D 特征点 m_2 与其所对应的 3D 空间点 M_2 的对应关系,则其满足

$$m_2 = H_w^2 M_2 \quad (5-1)$$

由于已建立了第一帧图像上的特征点 2D/3D 对应关系 (m_1, M_1) ,即存在 H_w^1 且满足

$$m_1 = H_w^1 M_1 \quad (5-2)$$

若 (m_1, m_2) 为 2D/2D 特征匹配点对,则其所对应的空间 3D 点为同一点,即 $M_1 = M_2$,首图像与第二帧图像间的单应性矩阵 H_1^2 满足

$$m_2 = H_1^2 m_1 \quad (5-3)$$

由此可得

$$H_w^2 = H_1^2 H_w^1 \quad (5-4)$$

由式(5-4)可以看出,如果单应性矩阵 H_w^{i-1} 已知,并且可以精确地求解出连续帧图像间的单应性矩阵 H_{i-1}^i ,则当前帧所对应的 3D/2D 单应性矩阵 H_w^i 就可由式(5-4)求解。

继续考察上述第二帧的情况,为了在求解出对应的单应性矩阵 H_w^2 之后求解相机的位姿信息,写出 H_w^2 单应性矩阵的分解形式

$$H_w^2 = K[r_1, r_2, t] \quad (5-5)$$

观察式(5-5)的右端,在摄像机内部参数 K 已知的情况下,通过式(5-5)可直接求解出 r_1 、 r_2 和 t ,此时仍缺少 r_3 。由于 r_1 、 r_2 、 r_3 共同构成了旋转矩阵 R 的三列,并且旋转矩阵 R 满足单位正交矩阵的约束条件,因此可通过式(5-6)求解 r_3

$$r_1 = r_1 / \|r_1\|, r_2 = r_2 / \|r_2\|, r_3 = r_1 \times r_2, r_3 = r_3 / \|r_3\| \quad (5-6)$$

求解出 r_3 之后即完成了当前帧摄像机位姿 R 、 T 的求解,后续帧摄像机位姿的求解可依次按上述方法进行。

基于平面结构的增强现实跟踪注册算法实际上是一个累积计算的过程,所以不可避免地存在误差累加。因此,该算法不适用于漫游式实时增强现实系统。

5.1.3 实验结果

图 5.1 所示为手工选择场景中的平面结构和进行系统初始化的过程,并利用该算法对真实场景进行三维跟踪注册,图 5.2 为采用该算法的不同视点下的虚实融合显示结果。

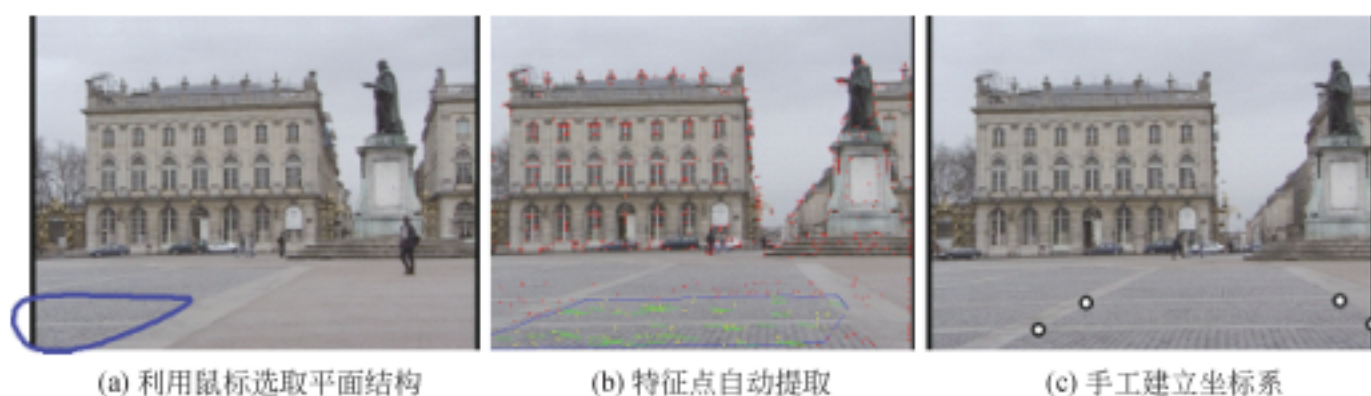


图 5.1 手工选择场景中的平面结构和进行系统初始化的过程^[4]



图 5.2 不同视点下的增强场景显示效果^[4]

基于平面结构的跟踪注册算法是早期研究人员提出的一种有效的实时无标识增强现实跟踪注册算法。该算法的实时性较好,包括给定区域的特征点提取、匹配、单应性矩阵的计算等,在 P4 2.8G、512MB 内存的机器上可在 100ms 内完成。但是由于算法自身的局限性,其不可避免地存在以下缺点。

(1) 对场景的约束较大。由于该算法自身的原因,其所能处理的场景中必须存在一个优势平面。并且由于初始化的需要,至少需要知道位于该平面上的四个空间点坐标,而该约束条件在实际应用中并不总能得到满足,极大地限制了该算法的应用范围。

(2) 鲁棒性较差。鲁棒性较差主要体现在两方面:一方面算法对于摄像机的突然抖动不能正确处理,容易导致算法终止运行;另一方面算法不可避免地存在误差累积,导致算法不能长时间稳定运行。

5.2 基于模型和关键帧的注册方法

本节介绍的增强现实实时跟踪注册算法是由瑞士联邦理工学院计算机视觉实验室的 Lepetit 等在 2003 年提出的^[5~7]。该算法利用场景的离线和在线信息完成增强现实的三维实时跟踪注册,对摄像机的视点变化、环境光照以及物体的遮挡等都保持了较好的鲁棒性,在当时是一种较为优秀的增强现实跟踪注册算法。

该算法分为离线和在线两个阶段。离线阶段,系统利用商业化软件 Boujou^[8]等恢复场景的三维空间结构和摄像机的运动轨迹,并创建场景的关键帧信息库。在线阶段,系统首先选择与当前帧视点最接近的关键帧,建立当前帧与关键帧间的特征点匹配;在已知场景模型的基础上,进一步构建当前帧图像上的特征点与三维模型上相对应的空间点的 2D/3D 匹配,从而可以利用优化算法计算摄像机的位姿。虽然在应用该算法时需要预先获取部分场景的 3D 模型以及关键帧,但是在获取这些先验知识的基础上,系统能够克服传统算法的诸多不足。因此,可以应用于实时性和鲁棒性要求较高的增强现实系统中。

5.2.1 3D 场景模型与关键帧的创建

在离线阶段,系统需要预先获得场景的 3D 模型,建立来自不同视点的真实场景参考图像数据库,并对其进行标定,这一过程可以通过使用商业三维运动结构重建软件 Boujou 实现。在场景 3D 模型已知的情况下,标定一些关键帧一般只需要几分钟。场景关键帧的获取过程如图 5.3 所示,首先绕场景拍摄一段图像序列,利

用 Boujou 估计场景的三维结构以及每一帧图像中摄像机的投影矩阵,选择序列中的 8~10 帧作为关键帧。由于 Boujou 软件是在全局优化的基础上计算获得的场景和摄像机的参数信息,因此具有非常高的计算精度,其计算结果可被用作基准数据。

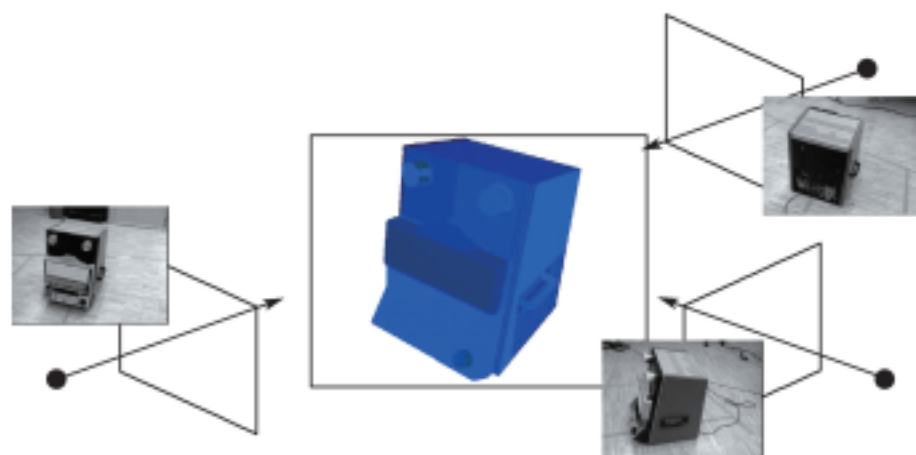


图 5.3 场景关键帧的获取^[5,6]

在已知物体三维结构信息的情况下,将关键帧图像上提取的特征点反投影至物体表面后即可建立特征点的 2D/3D 匹配。用 $m(u_i, v_i)$ 表示特征点的图像坐标,用 $M(x_i, y_i, z_i)$ 表示其三维空间坐标,则可建立如下方程:

$$m = \lambda PM, P = K[R | T] \quad (5-7)$$

式中, P 为投影矩阵,可分解为摄像机内部参数 K 以及外部参数 R 和 T , λ 为比例因子。在 R 和 T 已知的情况下,可对每一关键帧图像创建物体特征点的 2D/3D 匹配、摄像机的投影矩阵、待注册物体每个表面的法向矢量 n 等数据结构信息。

5.2.2 选择参考图像帧

完成系统的离线处理后,即可在在线阶段实现摄像机位姿的实时跟踪。由于摄像机在场景中的位置随用户的视点变化而改变,此时必须确定与用户当前视点最接近的关键帧。关键帧的选择判据如图 5.4 所示,其中 K_1 、 K_2 为关键帧的位置, C 为前一帧摄像机的位置。

在进行关键帧的选择时,只是简单地计算前一帧摄像机与关键帧摄像机的欧式距离显然是不够准确的。例如,对于

图 5.5 所示的情况,如果只是按照距离判据,则应选择 K_1 作为关键帧。但从图 5.5

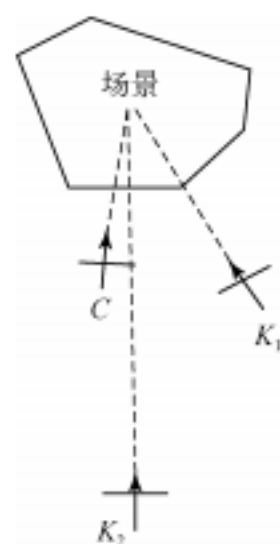


图 5.4 关键帧的选择判据^[7]

可以看出, K_2 与 C 的观察视点更接近。若选择 K_1 作为关键帧与当前帧进行图像匹配, 则可能只获得较少的匹配点对, 这势必影响跟踪定位计算的精度。为修正该问题, 在实际应用过程中, Lepetit 选择了一种基于被跟踪物体外观的关键帧选择判据

$$\sum_{\forall f \in \text{Model}} (\text{Area}(f, A_p[R_p | T_p]) - \text{Area}(f, A_k[R_k | T_k]))^2 \quad (5-8)$$

式中, $\text{Area}(f, P)$ 代表平面 f 经由投影矩阵 P 投影到图像平面后的面积。

采用加速的 OpenGL 将被跟踪物体的模型按照前一帧摄像机的姿态投影到图像平面, 每一个平面采用不同的颜色渲染, 从而可以获得该渲染结果的直方图。横轴为平面的 ID, 纵轴为对应平面颜色的像素数。按式(5-8)将所获得的直方图数据进行运算, 即可选择出最接近当前帧的关键帧。

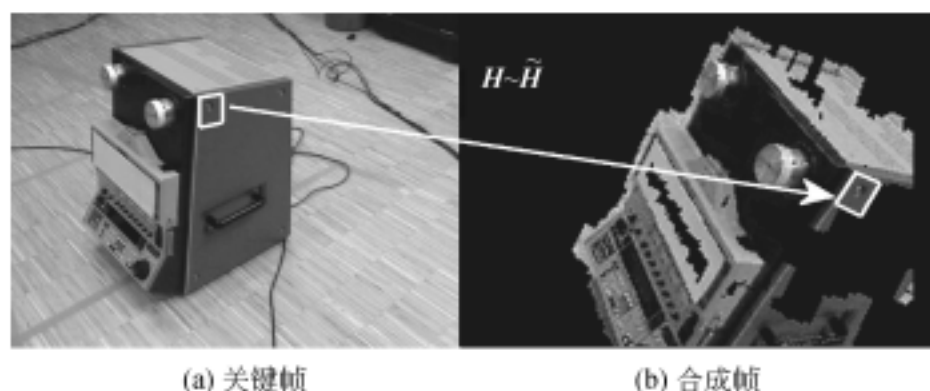


图 5.5 合成中间帧图像^[5,6]

5.2.3 宽基线匹配

将当前帧图像与选取的参考帧图像进行特征匹配时, 两幅图像间的视点仍可能存在较大偏差, 此时两幅图像间的特征匹配为宽基线匹配。为提高特征点的匹配数量, 该算法采用绘制中间帧的方法。首先, 在参考帧图像上的特征点周围提取 15×15 像素大小的图像区域, 利用前一帧图像的位姿参数对其进行旋转平移变换合成中间图像帧。该中间图像的视点采用的是当前图像前一帧的视点参数, 所以这两幅图像间的特征点匹配由宽基线匹配转换为窄基线匹配问题。在进行上述计算时, 其假设特征点周围的像素区域位于同一物体的表面, 因此可视为平面。通过单应性矩阵 H , 参考图像上的特征点 m_0 周围的像素 m_k 可映射到合成的中间图像像素 m_s , 单应性矩阵可通过下式计算:

$$m_s = H \cdot m_k, H = A_k(\delta R - \delta T, n'^T/d')A^{-1} \quad (5-9)$$

$$\delta R = RR_k^T, \delta T = -RR_k^T T_k + T, n' = R_k n; d' = d - T_k^T(R_k n) \quad (5-10)$$

式中, $[R_k | T_k]$ 和 $[R | T]$ 分别为参考图像与当前图像的前一帧的旋转和平移矩阵, n 为特征点像素所处的物体表面的法向矢量, d 为特征点到原点的距离。 H 矩阵将特征点 m_0 附近的像素 m_k 转换到中间帧的像素 m_i 。为节约计算时间, 该变换在 m_0 附近被近似为仿射变换, 其一阶近似为

$$m_i \approx Hm_0 + J_H(m_0)(m_k - m_0) \quad (5-11)$$

式中, J_H 是 H 的雅可比矩阵。图 5.5(a) 所示为一参考图像帧, 白色的边框为参考图像上的一个特征点周围的区域, 该图像区域在单应性矩阵 H 的映射下合成为图 5.5(b) 所示的中间图像。为匹配当前帧与中间合成图像间的特征点, 算法在当前图像的特征点周围开设 7×7 像素大小的窗口, 在中间图像上搜索其相关匹配。采用这一方法搜索到的图像匹配点不够精确, 存在较多误匹配数据, 因此需要采用 RANSAC 算法进一步提高数据匹配的准确度。利用 RANSAC 算法可建立当前帧图像特征点 m_c 与中间帧图像点 m_i 间的对应关系

$$m_c^j \leftrightarrow m_i^i \quad (5-12)$$

合成帧来源于参考图像, 而参考图像是经过标定的。因此, 中间合成图像点 m_i 的三维坐标 M_i 为已知参数。在三维点坐标满足 $M_c^j = M_i^i$ 的情况下, 可以建立三维点坐标 M_c^j 与当前帧图像点 m_c^j 间的映射关系。

5.2.4 摄像机位姿求解

建立起当前图像特征点与其三维空间坐标点的对应关系后, 在摄像机内部参数 K 已知的情况下, 为了进一步提高参数估计精度, 强化算法对错误数据匹配的容忍度, 该算法采用 Tukey M-估计算法计算摄像机的位姿。Tukey M-估计算法利用

最小化误差残差 $\min \sum_{i=1}^n \rho(r_i)$ 获得参数的最优估计, 此处 ρ 为连续对称函数, $r_i = \|m_i - \lambda P_i M_i\|$ 为图像的反投影误差。Tukey M-估计算法中的 ρ 可以表示为

$$\rho(x) = \begin{cases} \frac{c^2}{6} \left\{ 1 - \left[1 - \left(\frac{x}{c} \right)^2 \right]^3 \right\}, & |x| \leq c \\ \frac{c^2}{6}, & |x| < c \end{cases} \quad (5-13)$$

通过迭代优化即可求解摄像机的旋转和平移矩阵, Tukey M-估计算法的初始值为当前图像前一帧的视点参数。

在实际应用过程中, 如果仅利用关键帧计算摄像机的运动参数, 其参数估计结果可能出现大幅跳跃的现象。为此, 可通过加入前后帧信息对计算结果进行约束

的方法加以克服。记 n_t^i 为第 t 帧中点 i 的 2D 图像坐标, N^i 为其对应的未知空间 3D 图像坐标,该算法通过最小化式(5-14)来包含这些附加信息

$$\sum_{j=1}^t r_j + \sum_{i=1}^k \sum_{j \in \Theta^i} \rho_{\text{TUK}}[\|n_j^i - \phi(P_j, N^i)\|^2] \quad (5-14)$$

式中, r_j 对应 P_1 到 P_t 的投影矩阵,即摄像机从初始到当前时刻 t 的投影矩阵, Θ^i 是出现第 i 个特征点的所有帧的集合。为了减小计算量,只计算当前帧和前一帧。于是,问题转化为

$$\min_{P_t, P_{t-1}, N^i} (r_t + r_{t-1} + \sum_i s_t^i) \quad (5-15)$$

$$s_t^i = \rho_{\text{TUK}}[\|n_t^i - \phi(P_t, N^i)\|^2 + \|n_{t-1}^{w(i)} - \phi(P_{t-1}, N^i)\|^2] \quad (5-16)$$

式中, $n_{t-1}^{w(i)}$ 为前一帧中与当前帧特征点 n_t^i 匹配的点。

5.2.5 实验结果

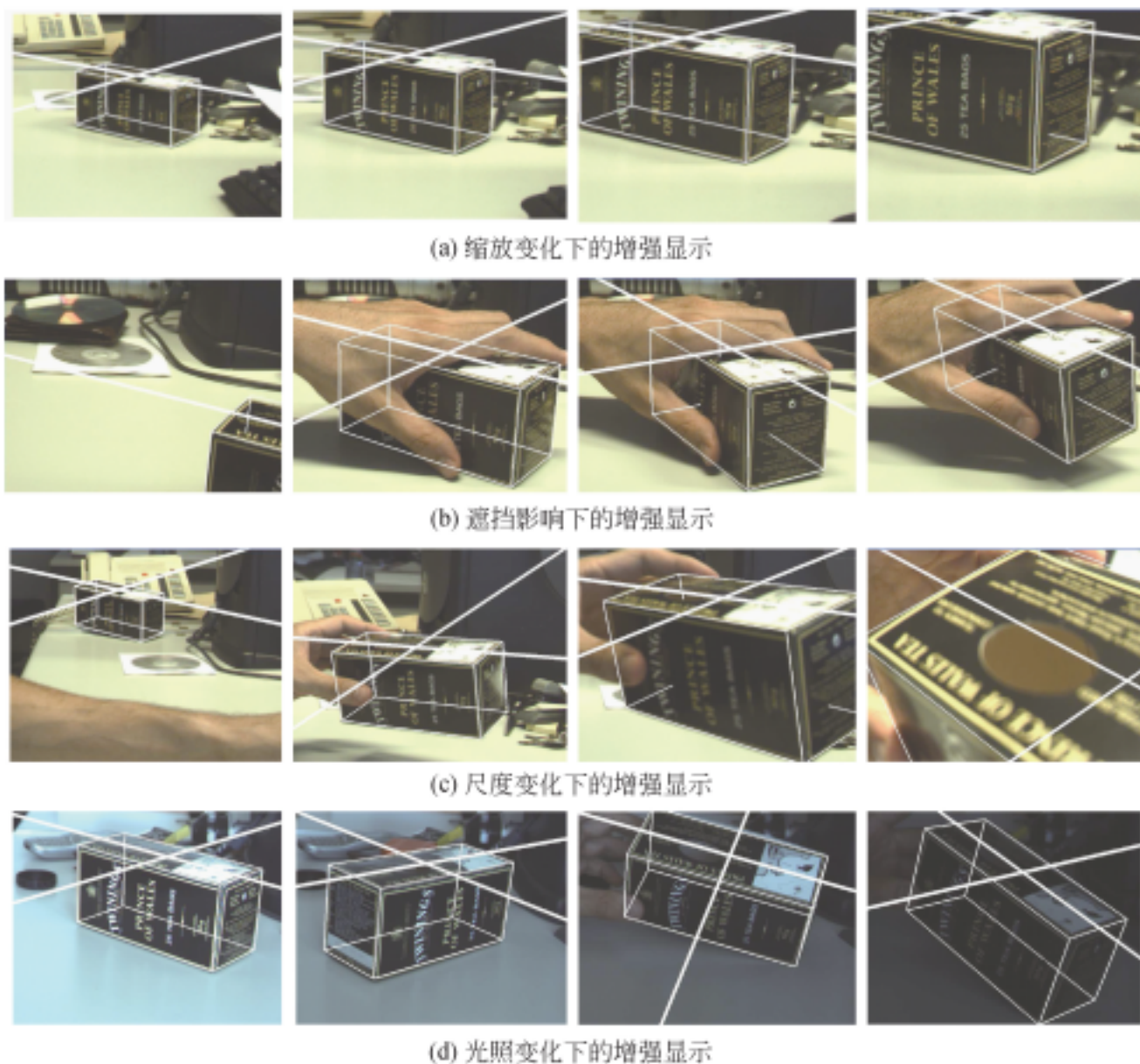
图 5.6 所示为采用 Tukey M-估计算法进行增强现实系统跟踪注册的结果,从图 5.7 可以看到,算法对图像缩放、遮挡、尺度、光照等均能保持一定的鲁棒性。



图 5.6 增强现实系统跟踪注册的结果^[5,7]

5.2.6 算法评价

Lepetit 等提出的联合 3D 模型与关键帧的无标识实时跟踪注册算法,有效解决了注册算法中常常存在的抖动和漂移问题,是一种较优秀的跟踪注册算法。但是,由于该算法是在 2003 年提出的,当时对于解决特征的宽基线匹配问题,还需要绘制中间帧将其转换为窄基线匹配问题。随着尺度不变性特征提取匹配算法的提出(如 SIFT 算法^[9]、SURF 算法^[10]以及 Ferns 算法^[11]),这些特征提取算法自身就具有对旋转、缩放以及大视角变化的不变性。因此,不需要将宽基线匹配转换为窄基线匹配。此外,算法在选择参考图像帧时,采用 OpenGL 绘制目标物体的模型,并将其投影至图像平面上统计直方图。用直方图匹配的方式选择离当前帧视点最接近的参考帧图像。这样做计算量较大,针对小场景还可胜任。若是户外的大范围复杂场景,其计算量将相当可观。

图 5.7 算法在不同环境下的增强显示^[6]

无论如何,该算法提供了一个较完善的增强现实跟踪注册解决方案,此后的大多数增强现实跟踪注册算法大多是在此基础上加以改进的。例如,针对参考帧选择方面,Zhang 等^[12]提出参考帧图像不仅要与当前帧图像视点接近,还需要能够在其上提取到较多生命周期长的特征点。针对特征提取匹配方面,Lowe 提出的目前公认的性能最为优越的 SIFT 算法^[9],以及 Ozuysal 等提出的 Ferns 特征识别算法^[11]均可解决特征的宽基线匹配问题。

5.3 基于图像匹配的无标识跟踪注册方法

为迎接 2004 年的奥运会盛典,希腊多家研究机构和政府部门成立了联合小组设计开发了基于增强现实的文化古迹导游系统 Archeoguide^[13~15]。图 5.8 为该小组在希腊的奥林匹亚神庙进行现场实验的效果图,该系统利用 GPS 实现粗定位,然后采用图像匹配的方式实现精确的无标识跟踪注册。



图 5.8 奥林匹亚神庙的增强效果图及现场测试场景^[13]

该算法采用 2D 图像匹配算法获得真实场景与用户间的相对位姿信息,需要保证真实场景中的待跟踪物体或景物与用户间的距离较远,一般需大于 5m。算法具体步骤如下。

- (1) 创建来自不同视点的真实场景参考图像数据库。
- (2) 计算当前图像与图像库中的参考图像间的互相关系数,确定与用户当前视点最匹配的参考图。
- (3) 采用 Fourier-Mellin 变换算法计算出真实场景与参考图像间的相对位姿信息。
- (4) 渲染融合显示。

5.3.1 关键帧匹配

算法在关键帧选取时,采用的是计算两幅图像间的归一化互相关系数的方式。根据相关系数的计算结果,可以选择与当前帧图像视点最接近的关键帧图像。然后利用 2D 图像匹配算法,计算当前图像与关键帧图像间的视点偏差,估计用户当前的位姿信息。

5.3.2 2D 图像匹配算法

在2D刚性图像配准中,需要获得两幅图像之间的尺度 S 、旋转 R 和平移 T 三个参数,该算法利用 Fourier-Mellin 变换实现这三个参数的检测。其核心思想是利用两幅图像 Fourier 谱的相位相关性实现图像的平移检测,同时结合极坐标变换和 Mellin 变换将旋转和尺度参数的检测问题转换为平移检测问题,实现 R 、 S 、 T 三个参数的综合检测。

1. 图像平移检测的相位相关检测法

设 f_s 为输入图像, f_r 为参考图像,如果二者之间存在相对平移量 $(\Delta x, \Delta y)$,即

$$f_s(x, y) = f_r(x - \Delta x, y - \Delta y) \quad (5-17)$$

设 F_s 、 F_r 是两幅图像所对应的 Fourier 谱,根据 Fourier 变换的相移定理,有

$$F_s(u, v) = e^{-j2\pi(u \cdot \Delta x + v \cdot \Delta y)} F_r(u, v) \quad (5-18)$$

经 Fourier 变换后相对平移信息转移到相位因子中。为求解 $(\Delta x, \Delta y)$,定义 F_s 和 F_r 的归一化互相关功率谱 R 如下:

$$R = \frac{F_s(u, v) \cdot F_r^*(u, v)}{|F_s(u, v) \cdot F_r(u, v)|} = e^{j2\pi(u \cdot \Delta x + v \cdot \Delta y)} \quad (5-19)$$

式中, F^* 是 F 的复共轭。对 R 取 Fourier 反变换,可得到 xOy 平面上的冲激函数,即

$$F^{-1}(R) = \delta(x - \Delta x, y - \Delta y) \quad (5-20)$$

通过检测冲激函数 $\delta(x - \Delta x, y - \Delta y)$ 的峰值位置,即可获得两幅图像间的相对平移量 $(\Delta x, \Delta y)$ 。

2. 图像平移和旋转参数的检测

若输入图像 $f_s(x, y)$ 是由参考图像 $f_r(x, y)$ 经过角度为 $\Delta\theta$ 的相对旋转以及 $(\Delta x, \Delta y)$ 的相对平移后得到的,那么它可以表示为

$$f_s(x, y) = f_r(x \cdot \cos\Delta\theta - y \cdot \sin\Delta\theta - \Delta x, x \cdot \sin\Delta\theta + y \cdot \cos\Delta\theta - \Delta y) \quad (5-21)$$

根据 Fourier 变换的性质,其 Fourier 谱 F_s 和 F_r 之间存在如下关系

$$F_s(u, v) = e^{-j2\pi(u \cdot \Delta x + v \cdot \Delta y)} F_r(u \cdot \cos\Delta\theta - v \cdot \sin\Delta\theta, u \cdot \sin\Delta\theta + v \cdot \cos\Delta\theta) \quad (5-22)$$

对式(5-22)两边取模得到

$$|F_s(u, v)| = |F_r(u \cdot \cos\Delta\theta - v \cdot \sin\Delta\theta, u \cdot \sin\Delta\theta + v \cdot \cos\Delta\theta)| \quad (5-23)$$

这意味着 $|F_s|$ 中只包含旋转角度信息 $\Delta\theta$ 。在极坐标系中, $|F_s|$ 和 $|F_r|$ 之间的这一旋转变换可以表示成角度参数的平移,即

$$|F_s(\rho, \theta)| = |F_r(\rho, \theta - \Delta\theta)| \quad (5-24)$$

采用相位相关检测即可得到相对旋转角 $\Delta\theta$ 。获得 $\Delta\theta$ 以后,对输入图像进行反向旋转校正,可得到 f'_s 。 f'_s 与参考图像 f_r 之间仅存在相对平移量 $(\Delta x, \Delta y)$,再次使用相位相关检测即可得到 $(\Delta x, \Delta y)$ 。

3. 图像平移、旋转和缩放参数的检测

若输入图像 f_s 是参考图像 f_r 经由尺度缩放因子 k 、旋转因子 $\Delta\theta$ 以及平移因子 $(\Delta x, \Delta y)$ 变换而来的,即

$$f_s(x, y) = f_r[k \cdot (x \cdot \cos\Delta\theta - y \cdot \sin\Delta\theta) - \Delta x, k \cdot (x \cdot \sin\Delta\theta + y \cdot \cos\Delta\theta) - \Delta y] \quad (5-25)$$

则 Fourier 谱 F_s 和 F_r 之间具有如下关系

$$F_s(u, v) = \frac{1}{k^2} \cdot e^{-j2\pi(\frac{u\Delta x}{k} + \frac{v\Delta y}{k})} \cdot F_r\left(\frac{u \cdot \cos\Delta\theta - v \cdot \sin\Delta\theta}{k}, \frac{u \cdot \sin\Delta\theta + v \cdot \cos\Delta\theta}{k}\right) \quad (5-26)$$

对式(5-26)两边取幅度谱

$$|F_s(u, v)| = \frac{1}{k^2} \cdot \left| F_r\left(\frac{u \cdot \cos\Delta\theta - v \cdot \sin\Delta\theta}{k}, \frac{u \cdot \sin\Delta\theta + v \cdot \cos\Delta\theta}{k}\right) \right| \quad (5-27)$$

令 G_s 和 G_r 是 F_s 和 F_r 相应的幅度谱,则式(5-27)可简化为

$$G_s(u, v) = \frac{1}{k^2} \cdot G_r\left(\frac{u \cdot \cos\Delta\theta - v \cdot \sin\Delta\theta}{k}, \frac{u \cdot \sin\Delta\theta + v \cdot \cos\Delta\theta}{k}\right) \quad (5-28)$$

令 H_s 和 H_r 是幅度谱 G_s 和 G_r 从笛卡儿坐标系转换到极坐标系后的结果,即

$$H_s(\rho, \theta) = \frac{1}{k^2} \cdot H_r(\rho/k, \theta + \Delta\theta) \quad (5-29)$$

其中令

$$\begin{cases} \rho = (u^2 + v^2)^{1/2} \\ \theta = \tan^{-1}(u/v) \end{cases} \quad (5-30)$$

对 ρ 引入对数变换,即取如下对数-极坐标变换

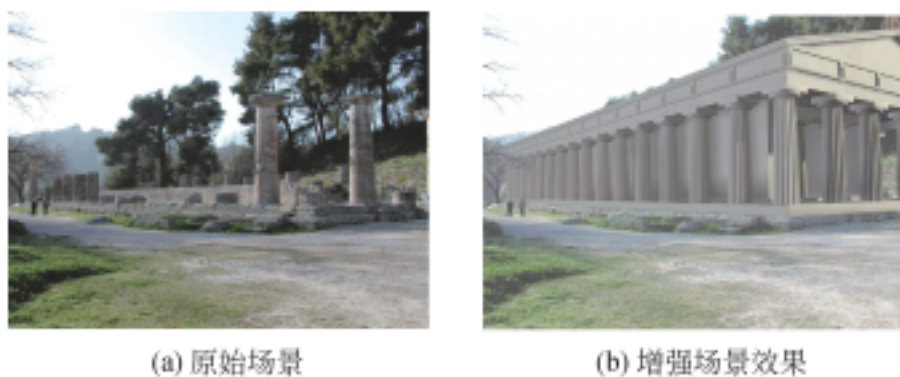


图 5.10 遗址上的增强现实显示实验结果^[14,15]

5.4 基于 SIFT 特征的无标识点跟踪注册方法

从本质上讲,该算法采用的也是基于稀疏点模型的跟踪注册算法。稀疏的场景 3D 模型是通过提取匹配序列帧图像上的 SIFT 特征点构建的。事实上,特征提取是实现增强现实跟踪注册的基础,也是计算机视觉领域的重要研究内容。目前在增强现实系统的跟踪注册算法中,常用的图像特征包括点特征、线特征和面特征等。从序列图像中提取尽可能多而定位精确的特征是完成增强现实系统虚实场景注册的关键步骤。

关于特征提取与匹配问题的研究,计算机视觉领域已有大量文献报道。特征提取匹配的典型流程是首先在待配准的图像上选取显著特征,如封闭区域、边缘、线段交点、角点、兴趣点等,然后确定这些特征在图像间的对应关系。用于特征匹配的特征描述子和相似性度量标准很多,包括灰度相关、二值图像相关、链码相关、结构匹配、不变矩之间的距离、动态规划和松弛法等。然而由于增强现实应用场景的复杂性,这些技术仍然难以可靠地解决增强现实跟踪注册系统中图像特征的对应问题。例如,Harris 特征点提取算子^[16]具有运算量小,自动化程度高,对环境光照变化、噪声能够保持一定鲁棒性的优点。该算子常被用于增强现实系统中,实现真实场景中的自然特征提取。但是 Harris 算子的定位精度较低,只能检测到像素级精度的角点特征,同时对图像的仿射变换不具有不变性。

2004 年 Lowe 总结了现有的基于不变量技术的特征检测方法,提出了一种基于尺度空间的对图像缩放、旋转甚至仿射变换保持不变性的图像局部特征描述算子——SIFT(scale invariant feature transform)算子^[9],即尺度不变特征变换。该算法被公认为目前性能最优越的特征提取算子,被广泛应用于目标跟踪、运动结构重建、模式识别、图像匹配等领域。因此,本节首先介绍 SIFT 算法,然后讨论基于该特征算子的增强现实跟踪注册算法。

5.4.1 SIFT 特征描述算子

SIFT 算子是一种能对图像旋转、尺度缩放、仿射变换、视角变化、光照变化等图像变化因素保持一定的不变性,而对物体运动、遮挡、噪声等因素也保持较强鲁棒性的图像局部特征描述算子。SIFT 算子是将图像的二维平面空间和 DoG (difference of Gaussian) 多尺度空间检测到的局部极值作为特征点,以使特征具备良好的独特性和稳定性的一种算子。

1. 极值搜索

一幅二维图像 $I(x, y)$, 其在某一尺度空间的表示可由图像与高斯核卷积获得, 即

$$J(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (5-32)$$

式(5-32)中的二维高斯核为

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (5-33)$$

式中, (x, y) 代表图像的像素位置; σ 称为尺度空间因子, 大尺度对应于图像的概貌特征, 小尺度对应于图像的细节特征; J 代表图像的尺度空间表示。多尺度空间算子 DoG 定义为两个不同尺度的高斯核的差分

$$D(x, y, \sigma) = [G(x, y, k\sigma) - G(x, y, \sigma)] * I(x, y) \quad (5-34)$$

SIFT 算子针对图像上的每一点, 计算其在每一尺度下的 DoG 算子的响应值。然后将 DoG 尺度空间中的每个点与相邻尺度和相邻位置的点逐个比较, 如图 5.11 所示, 获得的局部极值位置即关键点所处的位置和对应的尺度。

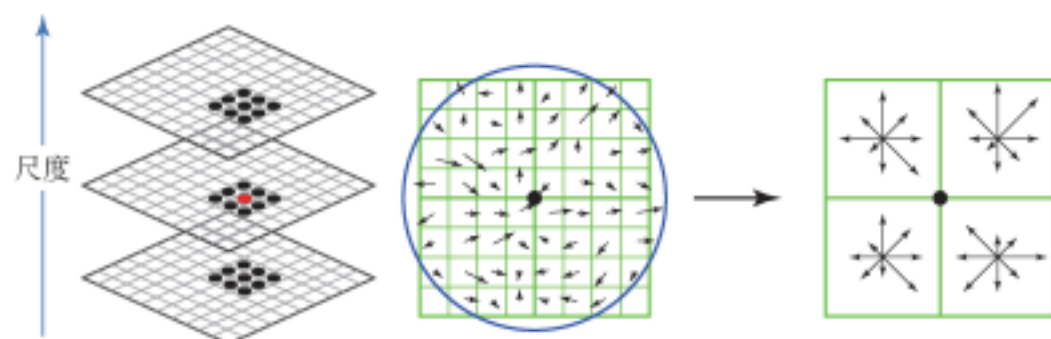


图 5.11 极值搜索过程(图像邻域和关键点描述)^[9]

2. 精炼关键点

在获得大量图像特征点的基础上, 对局部极值点利用三维二次函数拟合的方

法对关键点进行进一步精确定位。同时去除低对比度的关键点和不稳定的边缘响应点,以增强后续图像特征匹配的稳定性,提高算法的抗噪能力。

3. 指派关键点方向

SIFT 特征描述算子利用关键点邻域像素的梯度方向分布特性为每个关键点指定方向参数,使算子具备旋转不变性,点 (x, y) 处梯度的模值 m 和方向 θ 可通过下式表示

$$m(x, y) = \sqrt{[J(x+1, y) - J(x-1, y)]^2 + [J(x, y+1) - J(x, y-1)]^2} \quad (5-35)$$

$$\theta = \tan 2 \{ [J(x, y+1) - J(x, y-1)] / [J(x+1, y) - J(x-1, y)] \} \quad (5-36)$$

式中, J 所用的尺度为每个关键点各自所在的尺度。实际计算时,在以关键点为中心的邻域窗口内采样,并用直方图统计邻域像素的梯度方向,直方图的峰值即代表该关键点处邻域梯度的主方向。

4. 建立关键点描述

以 32 维 SIFT 描述算子的构造过程为例, SIFT 描述算子为每个关键点赋予一个主方向,该主方向是指关键点邻域内各点梯度方向的直方图中最大值所对应的方向。对任意一个关键点,在其所在的尺度空间取以关键点为中心的 8×8 像素大小的邻域,再将此邻域均匀地分为 2×2 个子区域,每个子区域大小为 4×4 像素。对每个子区域计算其梯度方向直方图(直方图均匀分为 8 个方向)。然后对 2×2 个子区域的 8 方向梯度直方图根据位置依次排序,构成一个 $2 \times 2 \times 8$ 维的向量,该向量就是 SIFT 特征描述算子,如图 5.11 所示。

5.4.2 SIFT 关键点的匹配

图像的 SIFT 特征向量生成后,需要完成序列图像间的特征匹配。为此, Lowe 采用关键点特征向量的欧式距离作为两帧图像中的关键点相似性测度。首先,根据特征向量的最近邻和次近邻距离之比,选择可靠性较高的匹配点对来计算几何约束模型;然后在几何模型的约束下对其余可能的匹配点对进行验证,扩展更多的匹配点对;最后去除由于图像自相似或对称性造成的可能的错误匹配点对。具体步骤如下。

(1) 对最近邻与次近邻之比设定阈值,如果描述算子之间的欧氏距离小于该

阈值,则确定为可能的匹配点对。

(2) 应用 RANSAC 鲁棒方法通过几何一致性检验去除野点,初步估计几何约束模型。

(3) 在几何模型的约束下对其余可能的匹配点对进行验证,扩展更多的匹配点对。

(4) 去除由于图像自相似或对称性造成的可能的错误匹配点对。

5.4.3 基于 SIFT 特征的摄像机位姿估计

在提出 SIFT 特征描述算子的基础上,Skrypnik 等于 2004 年提出一种基于 SIFT 算子的场景建模、识别与跟踪的注册算法,并将其应用于增强现实系统中^[17]。该算法首先对离线拍摄的场景图像序列进行稀疏场景结构重建,然后利用稀疏点模型以及场景的前后帧相关信息进行三维跟踪注册。该方法与本章 5.2 节介绍的基于模型和关键帧的算法相类似,在此不再作详细介绍,只给出实验结果。图 5.12 所示为三维运动结构重建结果,图 5.13 所示为不同条件下的增强跟踪注册显示效果。

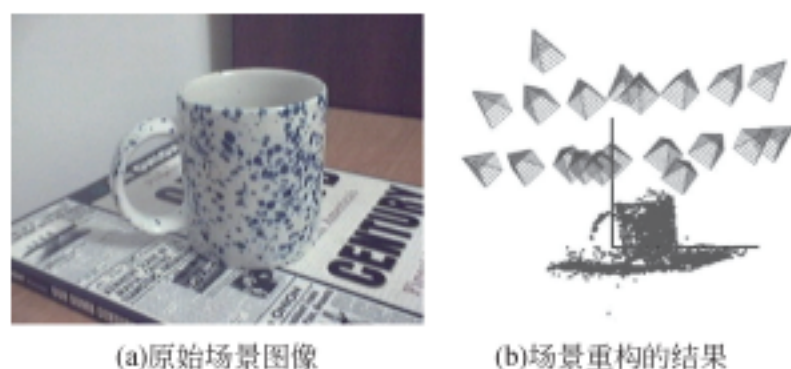


图 5.12 三维运动结构重建结果^[17]



图 5.13 虚实场景融合显示实验结果^[18]

5.5 基于运动结构重建的跟踪注册算法

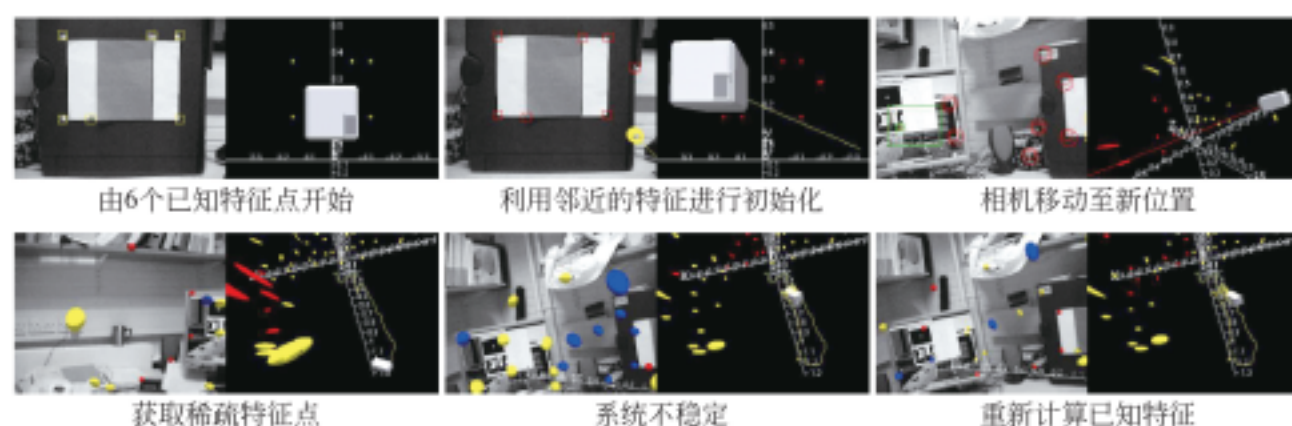
早在 1998 年,英国牛津大学机器视觉组的 Davison 就开始从事实时运动结构重建(real-time structure and motion)方面的研究工作,并于 2003 年在 ICCV 大会上发表文献[2],首次提出了基于主动重建的跟踪定位方法。该方法的思想源自机器人导航领域的同步定位和地图创建算法(simultaneous localization and mapping, SLAM)。本质上,SLAM 问题可以简单描述为:移动机器人在未知环境中运动时逐步构建周围环境的地图,同时运用该地图对机器人的位置和姿态进行估计。近年来,实时可视化定位和地图创建算法取得了很大的进步。比较典型的代表为 Davison 的单目摄像机 SLAM 算法、Montemerlo 等在粒子滤波框架下的 FastSLAM 2.0 算法^[19]以及 Klein 等的 PTAM 算法^[3]。本节就 Davison 等提出的 MonoSLAM 算法以及 Klein 等提出的 PTAM 算法在增强现实系统中的应用作简单介绍。

5.5.1 MonoSLAM 算法

Davison 在文献[2]和文献[20]中提出,其 SLAM 算法首先通过获取环境中放置的已知尺寸的黑色正方形来求取初始摄像机姿态矩阵。系统运行过程中,在估计当前摄像机位置的同时,对新观察到的特征点进行三维重建,以恢复其在水系坐标系下的位置,并供后续位姿估计过程使用。该方法的优点是可以根据主动重建得到的特征点,将用户的视野范围拓展到未知空间,很大程度上提高了系统的可用性。然而,算法也存在一些缺陷。首先该算法采用扩展卡尔曼滤波(extended Kalman filter, EKF)技术,将摄像机位姿和特征点三维位置相结合作为状态变量 $\hat{x}$

$$\hat{x} = \begin{pmatrix} \hat{x}_c \\ \hat{y}_1 \\ \hat{y}_2 \\ \vdots \end{pmatrix}, P = \begin{bmatrix} P_{xx} & P_{xy1} & P_{xy2} & \cdots \\ P_{y1x} & P_{y1y1} & P_{y1y2} & \cdots \\ P_{y2x} & P_{y2y1} & P_{y2y2} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (5-37)$$

式中, $\hat{x}_c$ 代表摄像机的运动状态, $\hat{y}_1, \hat{y}_2, \dots$ 表示空间环境中待构建特征点的水系坐标, P 为协方差矩阵。待构建特征点数的增加,将导致系统协方差矩阵计算的时间开销增大。因此,为保证系统的实时性,重建特征点的数目将受到严格限制。且该方法未采用“集束调整”等误差抑制策略,系统的重建误差将会随时间的推移发生误差积累,从而导致跟踪定位失败。图 5.14 为系统运行时的摄像机轨迹和地图创建过程。

图 5.14 摄像机轨迹和地图创建过程^[2]

文献[21]~文献[23]中沿用了文献[2],文献[20]中的思想,设计了一种基于粒子滤波与主动重建技术相结合的虚实配准方法。该方法采用粒子滤波技术对摄像机运动过程建模,以每一个粒子代表一种摄像机的可能位姿,以归一化互相关算法为基础计算每个粒子的权值,以所有粒子的加权和作为系统输出。这种方法的缺点是需要大量的粒子才能精确地模拟摄像机的位置和姿态。粒子数量越多,系统的运行速度越慢。因此,使用者需要在运行速度和注册精度(粒子数量)两个方面进行折中。同时该方法仍然没有摆脱误差积累问题,因而并不适于实际的增强现实应用。

2007年,Klein等在Davison的工作基础上提出了并行的跟踪定位与地图创建(parallel tracking and mapping,PTAM)算法,并将其应用于小范围的增强现实系统中^[3],引起了广泛的关注。该算法将摄像机定位与地图创建过程拆分为两个独立的过程,并行运算。利用计算机的双核处理能力,一个线程用于鲁棒地跟踪定位摄像头的姿态,另一个线程用于创建摄像头捕获的视频场景内的特征点地图。

5.5.2 PTAM 算法

与Davison的EKF框架下的MonoSLAM以及Montemerlo粒子滤波框架下的FastSLAM 2.0算法不同,Klein提出的PTAM算法具有如下特点。

(1)将定位与地图的创建分为两个独立的模块并行运行。定位算法将不再受地图创建方面的约束限制,从而可以采用更加鲁棒的跟踪算法,如由粗到细的跟踪算法。此外,由于定位和地图创建分为独立的线程,算法可充分利用计算机的并行处理能力,采用计算量较大的局部集束调整算法构建地图。

(2)地图初始化采用的是双目立体匹配5点算法^[24],系统创建的地图中的特征点数可达上千个。

简单地讲,PTAM 就是将 SLAM 分为跟踪和制图两个独立线程。跟踪的任务是在给定的地图中估计摄像机的位置,制图的任务是利用跟踪获得的摄像机位置将图像上的特征点三角化得到其空间坐标,同时利用集束调整对地图进行整体优化,进而得到空间点和摄像机位置的最优估计。定位和地图创建的流程如图 5.15 所示。

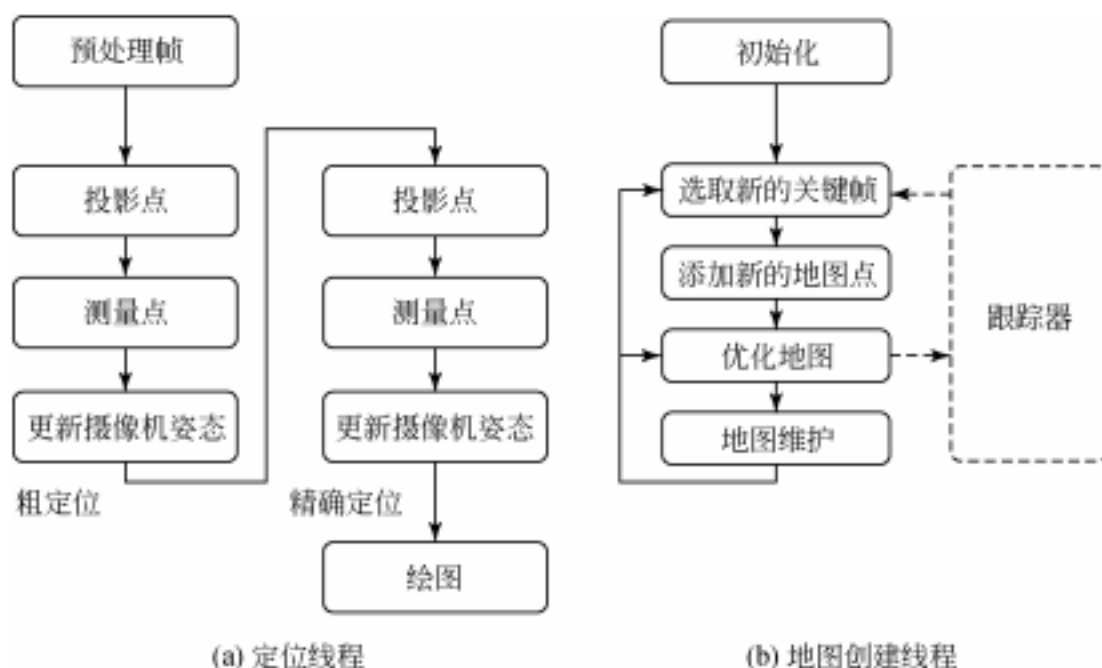


图 5.15 PTAM 算法的定位和地图创建流程

1. 跟踪定位

系统使用 Fire-i 摄像头获取每秒 30 帧 640×480 像素的灰度图像,在获取的灰度图像上采用降采样方法生成 4 层图像金字塔。在每层金字塔图像上利用 FAST-10 快速提点算法获得大量的特征点,从而避免对图像作较耗时的高斯差分滤波,以提高系统实时性。根据减速摄像机模型,估计当前帧摄像机的姿态信息。其具体算法步骤如下。

- (1) 根据摄像机运动模型获得前一帧摄像机的姿态信息。
- (2) 根据当前帧的摄像机姿态,选择地图中的 50 个点投影至图像坐标系。
- (3) 在当前帧图像上搜索与之匹配的特征点集合。
- (4) 根据 2D/3D 对应关系,计算出摄像机的姿态信息进行更新。
- (5) 将 1000 个地图中的点,利用更新的摄像机姿态再次投影至图像坐标系中,并建立其与当前帧图像中的特征点匹配集合。
- (6) 重新建立 2D/3D 对应关系,进一步精确估计摄像机的姿态信息。

将地图上的点投影至图像坐标系,用 (u_i, v_i) 表示构建地图点的图像坐标,用 p_{iw} 表示其三维空间坐标,则可建立如下方程:

$$\begin{pmatrix} u_i \\ v_i \end{pmatrix} = \text{CamProj}(E_{cw}p_{iw}) \quad (5-38)$$

式中, E_{cw} 为世界坐标系至摄像机坐标系的变换矩阵 $[R | T]$, 摄像机模型采用具有桶型径向畸变的针孔摄像机模型,可表示为

$$\text{CamProj} = \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = \begin{pmatrix} u_0 \\ v_0 \end{pmatrix} + \begin{bmatrix} f_x & 0 \\ 0 & f_y \end{bmatrix} \frac{r'}{r} \begin{bmatrix} x/z \\ y/z \end{bmatrix} \quad (5-39)$$

$$r = \sqrt{\frac{x^2 + y^2}{z^2}} \quad (5-40)$$

$$r' = \frac{1}{\omega} \arctan\left(2r \tan \frac{\omega}{2}\right) \quad (5-41)$$

建立了图像 2D 特征点与其三维空间点的 3D 对应关系后,则采用 Tukey M-中值估计算法更新地图点的投影误差

$$u' = \operatorname{argmin} \sum_{j \in s} \operatorname{obj}\left(\frac{|e_j|}{\sigma_j}, \sigma_r\right) \quad (5-42)$$

式中,重投影误差 e_j 表示为

$$e_j = \begin{pmatrix} \dot{u}_j \\ \dot{v}_j \end{pmatrix} - \text{CamProj}(\exp(\mu)E_{cw}P_j) \quad (5-43)$$

式中, $\exp(\mu)$ 为摄像机姿态的指数地图表示法,具体定义可参见文献[3]。

2. 地图创建

地图创建过程可分为两个阶段:地图初始化阶段和地图扩展阶段。在地图初始化阶段,将摄像机对准工作场景拍摄一帧关键帧图像,利用 FAST 算法提取该图像的 1000 个特征点。平移摄像机到另外一个视角,拍摄第二帧关键帧图像。建立两帧图像间的特征点匹配,采用 5 点算法估计本质矩阵 E ,并利用三角化法获得初始地图。

地图扩展阶段:随着摄像机的运动,算法动态地选择新关键帧和特征点,更新构建的地图。首先,算法需要选择新的关键帧,关键帧的选择需满足一定的约束条件。为了使当前帧估计的摄像机姿态信息准确,当前帧距离上一关键帧间的间隔

要大于 20 帧,且摄像机要距地图上的点有一定的运动。一旦确定了当前帧是需要加入的新关键帧,则选择距当前帧位置距离最近的一关键帧作为另一帧图像构建双目立体视觉匹配,估算相应图像特征的三维点位置。因此,地图构建可以分为三个主要步骤:选定关键帧,计算关键帧中特征点的三维位置和地图更新控制,图 5.16 所示为地图创建流程。

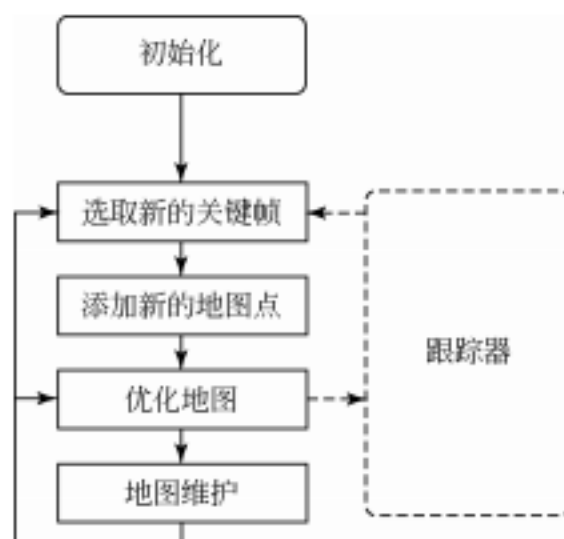


图 5.16 地图创建流程

3. 实验结果

为验证系统的工作性能,Klein 等分别利用从真实场景中采集的视频数据^[3]以及利用计算机合成的数据,对算法进行跟踪定位和地图创建方面的测试。计算机的配置为 Intel 双核 2.66 GHz 处理器,NAVIDA 显卡,Linux 操作系统。图 5.17 所示为利用计算机合成数据对该算法和 EKF-SLAM 算法进行对比的实验结果。

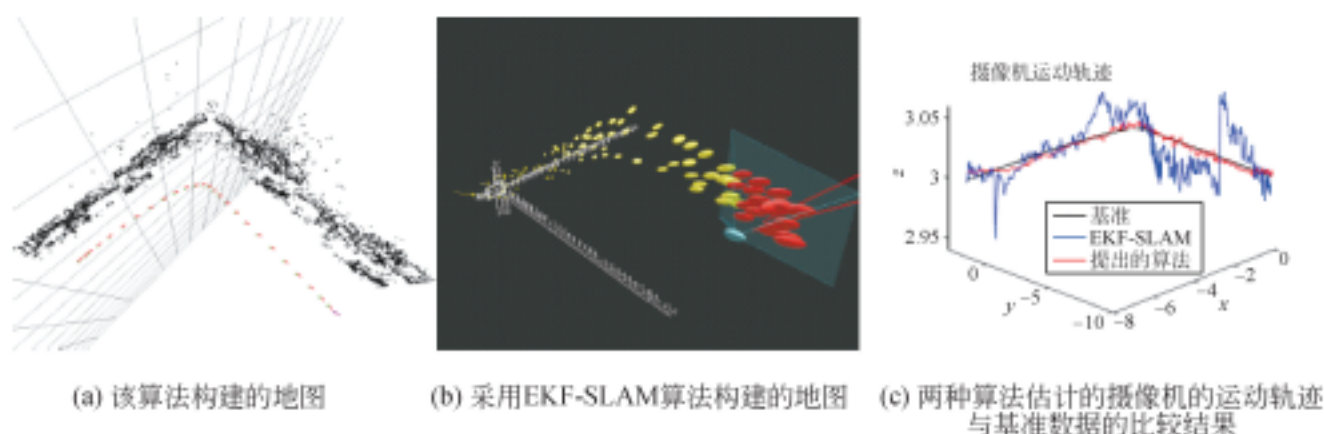


图 5.17 与 EKF-SLAM 算法的对比^[3]

由图 5.17 可以看到, Klein 等提出的 PTAM 算法可构建较为稠密的地图, 可达 6600 个地标点。EKF-SLAM 算法构建地图的地标较稀疏, 为 114 个。与基准数据的比较也表明, 该算法的摄像机跟踪定位精度和稳定性都较 EKF-SLAM 鲁棒。

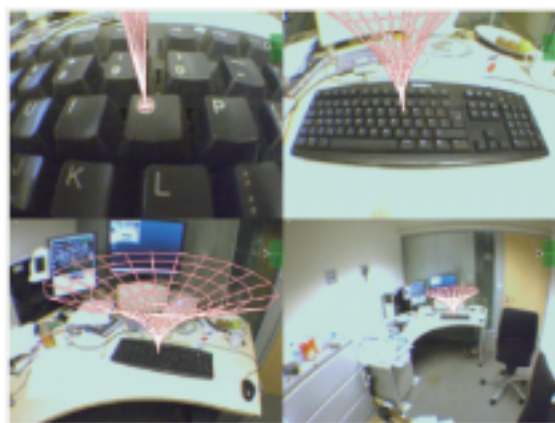


图 5.18 算法尺度适应性^[3]



图 5.19 增强现实射击游戏^[3]

图 5.18 所示为算法对尺度适应性的测试, 随着摄像机距离工作场景的远近变化, 虚拟添加的 3D 模型始终叠加在正确的空间位置。图 5.19 所示为利用该算法开发的增强现实射击游戏。虚拟的 3D 士兵以用户的办公桌为战场, 与用户展开一场激烈的射击游戏。

5.6 混合跟踪注册

前面详细讨论了基于视觉的注册跟踪算法, 下面进一步总结视觉注册算法的优缺点。视觉跟踪的优势表现在三方面。

- (1) 只对感兴趣的物体进行跟踪注册, 而不需要额外放置发射器、接收器。
- (2) 跟踪精度与被跟踪物体的成像分辨率成正比。
- (3) 能够最小化系统注册误差。

但是基于视觉的跟踪注册算法也有其局限性, 主要表现在以下三方面。

(1) 通常 CCD 的采样频率为 30Hz。因此, 视觉跟踪系统适用于低频位姿变化的测量。一旦用户或被跟踪物体发生快速、突然的运动, 则采用视觉跟踪的增强现实系统注册误差将增大。

(2) 图像处理、识别的计算量较大, 系统鲁棒性较差, 周围环境和用户的运动范围受限制。大多数基于视觉的注册算法均采用最小化迭代的方式取得全局最优。虽然利用线性方法求解方程能够获得唯一的解, 但这是以提取大量标识点或

图像信息为代价的。图像的识别与匹配是十分耗时的过程,一旦标识点或特征信息因为遮挡、模糊或从摄像机的视野中消失,则基于视觉的算法将产生严重的注册误差,甚至导致算法失效。因而在这种情况下,摄像机的连续实时跟踪无法实现。

(3) 受环境光照的影响较大。当环境光照强度改变时,将影响到标识点或图像的特征提取,造成较大的注册偏差。

硬件跟踪设备则不存在以上问题,硬件跟踪器具有较强的鲁棒性和实时性。对于一般的惯性传感器而言,其采样频率可达到 1kHz。因此,特别适用于感知运动物体的快速位姿变化。但是,硬件跟踪设备存在测量精度较低的缺点。Durlach 等曾得出结论^[25],由于系统的不精确性和系统延时方面的限制,目前单一的跟踪技术不可能很好地解决增强现实应用系统的方位跟踪问题。为此,国外著名大学和研究机构的研究人员相继提出采用混合跟踪(hybrid tracking)的方法对增强现实系统进行三维跟踪注册。所谓混合跟踪是指采用不同种类的跟踪设备,取长补短共同完成增强现实系统的注册任务。目前常采用的硬件跟踪设备包括机电跟踪器、电磁跟踪器、光电跟踪器、惯性跟踪器、超声波跟踪器、陀螺仪、GPS 等。

1999 年, Azuma 等提出将视觉跟踪与其他跟踪方法相结合,以克服室外视觉跟踪缺陷的新思路^[26,27],使得混合跟踪成为最近十年来增强现实的一个重要研究方向。混合跟踪是利用多传感器融合技术将来自不同类型跟踪器的输出姿态数据进行合并或融合的头部位姿跟踪方法。经过组合或者融合后的混合跟踪系统能够分享不同类别跟踪器的优点,在输出特性上达到优势互补。例如,视觉跟踪与非视觉跟踪的混合跟踪中,视觉跟踪具有跟踪精度高、鲁棒性差的特点,适合于低频运动跟踪;非视觉跟踪器具有鲁棒性好、精度低的特点,适合高频运动的跟踪。因此,将两者结合可以获得较为满意的跟踪效果。

5.6.1 根据传感器类型分类

按照跟踪器类型的不同,混合跟踪的跟踪器大致分为惯性-超声波混合^[28,29]、惯性-GPS^[30,31]、GPS-罗盘混合^[32]、惯性-罗盘混合^[33]、视觉-GPS 混合^[34]、视觉-电磁混合^[35~38]、视觉-惯性混合等类型的跟踪器。其中视觉-惯性混合跟踪器集成了视觉跟踪器精度高和惯性跟踪器高频工作鲁棒性好的优点,被广泛应用于机器人跟踪、运动与结构恢复、室内与室外增强现实系统等。IO 光学跟踪器和 OI 光学跟踪器的混合跟踪则兼具了 IO 跟踪器方向精度高和 OI 跟踪器位置精度高的优点,在室内增强现实系统也得到了广泛应用。

1) 多硬件传感器混合跟踪注册

1994年, Azuma等^[39]开发的六自由度方位跟踪器由线性加速度器和陀螺仪构成, 增加了预估头部位置的准确性。1998年, Emura等^[40]采用电磁跟踪器与陀螺仪相结合的方法进行运动跟踪, 利用陀螺仪补偿电磁跟踪器测量的时间延时。2000年, 荷兰代夫特大学的研究小组利用GPS、惯性跟踪器、陀螺仪等几种不同的方位跟踪设备完成户外增强现实系统的三维注册跟踪任务^[41]。该混合跟踪设备的跟踪精度较高(旋转角度误差为 1° , 平移误差为1mm), 然而设备体积较大, 造价高昂。同年, Yokokohji研究小组^[42]开发出由6个加速度仪构成的方位跟踪设备, 用以预测用户头部的运动, 从而补偿系统延时造成的动态注册误差。

2) 视觉-惯性混合跟踪注册

1999年, USC集成媒体系统中心的You等使用惯性陀螺仪, 通过提供给视频摄像机相邻帧之间的相对方向提高视觉系统的计算效率和鲁棒性, 并使用视觉跟踪器校正惯性跟踪器的累积漂浮^[43~45]。2000年, 日本奈良科技大学的Kanbara等提出了一种立体摄像机和惯性跟踪器构成的鲁棒跟踪方法^[46]。2001年, 日本Canon公司混合现实系统实验室的Sato等建立了一个室外便携式AR系统^[47], 此系统使用陀螺仪测量头部的三自由度方向, 使用头盔上的摄像机跟踪室外的自然特征。2004年, USC的Jiang建立了一个陀螺仪与基于线特征的视觉混合跟踪的室外AR系统^[48,49]。Sato等提出了一种使用陀螺仪和俯视(birds eye)摄像机的新型跟踪方法^[50]。日本大阪大学的Maeda等建立了一个可携带的AR系统^[51]。2002年, 奥地利格拉茨大学的Lang等开发了一个轻量级便携式惯性-视觉跟踪系统^[52,53]。2003年, 瑞典皇家技术学院的Rehbinder等使用一个速率陀螺仪和一个基于线特征的视觉跟踪器估计头部姿态^[54~56]。2005年, 维也纳技术大学的Gemeiner等使用惯性-视觉传感器估计头部姿态运动和恢复目标结构^[57,58]。葡萄牙科英布拉大学的系统与机器人学院在移动机器人研究中一直使用惯性-视觉混合跟踪技术^[59,60]。其他研究人员, 如Klein^[61]、Choi^[62]等在各自的研究应用中也使用了视觉-惯性混合跟踪技术。2006年, 剑桥大学Reitmayr等利用视觉以及混合重力、惯性、GPS等传感器在手持设备上开发了一套鲁棒的基于模型跟踪的户外增强现实系统——Going Out^[63], 如图5.20所示。

最近研究视觉-惯性混合跟踪比较典型的项目是欧洲的MATRIS, 该项目的研究目标是为增强演播室、文化遗产保护等增强现实系统等开发一个实时测量摄像机运动的跟踪系统^[64,65]。图5.21所示为该项目所使用的IMU和视觉传感器的集成单元、使用该实时摄像机跟踪系统实现的文化遗迹增强和演播室增强效果图。



图 5.20 户外手持增强现实系统^[63]



图 5.21 MATRIS 研究项目中的惯性-视觉混合跟踪^[65]

3) IO-OI 光学混合跟踪

IO-OI 光学混合跟踪器也称为 inside-outside-in 跟踪器。1998 年, Hoff 在其所开发的医学增强现实系统中将一个 IO 跟踪器和一个 OI 跟踪器的姿态数据进行了融合^[66,67]。2003 年, Satoh 等将一个普通摄像机作为 IO 跟踪器, 一个俯视摄像机作为 OI 跟踪器用于头部跟踪^[68]。2004 年, InterSense 公司的 Foxlin 等提出将 inside-outside-in 跟踪模式用于飞行模拟器的 FlightTracker 座舱头部跟踪系统^[69]。剑桥大学的 Klein 等将一个以 LED 为跟踪目标的 OI 跟踪器与一个以边缘线为跟踪目标的 IO 跟踪器进行融合, 应用于桌面 AR 系统^[70]。Yamazoe 等^[71]在多人运动跟踪系统中融合固定在环境中的摄像机和固定在 HMD 上的摄像机的观察信息。

5.6.2 根据传感器融合方式分类

1997 年, Brooks 等将用于定位跟踪的传感器融合分为三类: 互补式传感器融合、竞争式传感器融合以及协作式传感器融合。传感器融合在增强现实混合跟踪技术中的应用策略也可分为这三类。

1) 互补式传感器融合

如果多个传感器相互独立而且输出状态估计为不同类型,则所使用的传感器融合称为互补式传感器融合。互补式传感器融合组合了多个传感器的输出数据,可以得到更完整的系统状态估计。例如,Feiner 等^[30]使用磁力计测量头部偏航角,用双轴倾斜计测量头部俯仰角和滚转角,用 GPS 测量头部位置,从而获得完整的六自由度姿态。Foxlin 等^[72]使用 InterSense IS-300 惯性跟踪器测量方向,用 Pegasus FreeD 超声波跟踪器测量位置。另外一个比较典型的互补式传感器融合就是 Welch 等开发的 SCAAT 异步互补策略^[73],其原理是,在使用 EKF 进行融合时,针对每一个传感器建立一个测量模型和测量协方差矩阵,只要任何一个传感器的测量数据可用,就立即使用该传感器对应的测量变换矩阵和协方差矩阵进行预测和校正。南加利福尼亚大学的室外混合跟踪项目^[45,48]也使用了 SCAAT 互补式传感器融合策略。图 5.22 显示了这两种互补式传感器融合方法。

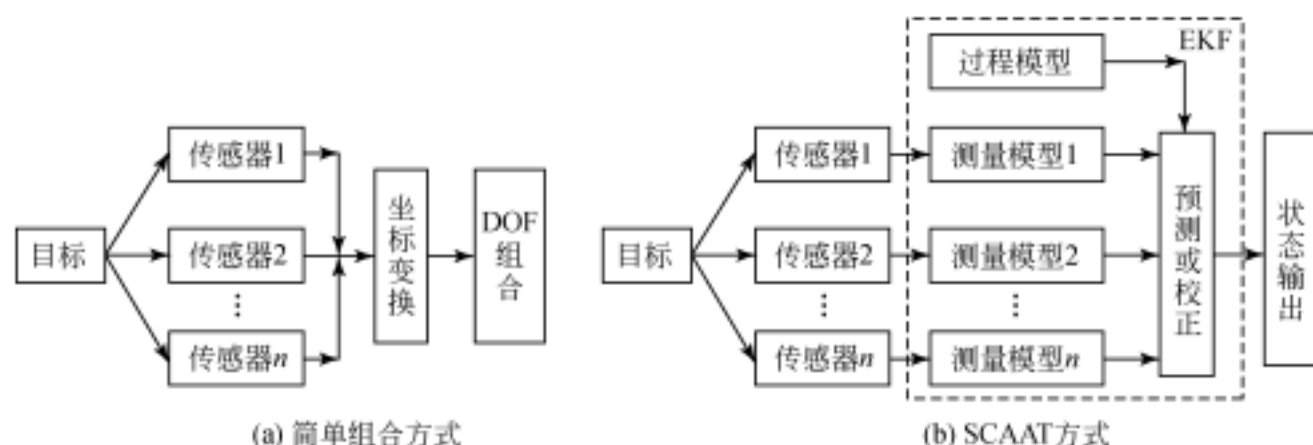


图 5.22 混合跟踪的互补式传感器融合示意图

2) 竞争式传感器融合

如果多个传感器能够在相同的空间坐标系中,给出一个不同测量类型的状态估计,则所使用的传感器融合称为竞争式传感器融合,如图 5.23 所示。在此融合方式下,必须考虑多个传感器引入的冗余,并且需要融合不同误差分布的传感器,以最小化测量误差。由于传感器的测量类型不同,竞争式融合主要使用卡尔曼滤波器或者扩展卡尔曼滤波器进行融合,以解决多个传感器不同测量类型的冲突问题。竞争式传感器融合是比较复杂但是在增强现实中应用较广泛的融合方式,它的典型应用是视觉-惯性混合跟踪。在此视觉-惯性混合跟踪器中,视觉跟踪器的 2D 图像特征点负责提供位置和方向信息,而惯性跟踪器负责提供速度/加速度、角速度/角加速度等刚体运动参数。两者融合的核心思想是,通过刚体运动学方程建

立视觉跟踪器摄像机坐标系中的测量信息与惯性跟踪器惯性坐标系中的测量信息之间的函数变换关系,从而确定扩展卡尔曼滤波器的过程变换矩阵和测量变换矩阵的表达式。竞争式传感器融合的典型例子是欧洲的 MATRIS 项目^[65],在该项目所开发的实时摄像机姿态跟踪算法中,使用四元数描述方向,建立了基于四元数的 EKF 融合框架,将视觉跟踪器的图像特征信息统一在惯性传感器的惯性坐标系下与惯性跟踪器的测量信息进行融合。

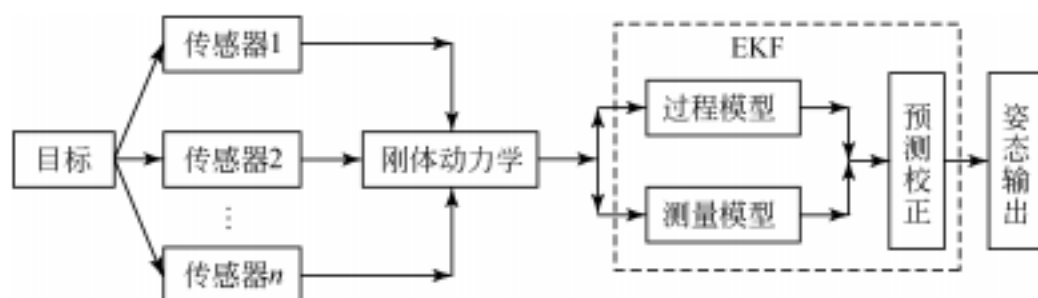


图 5.23 混合跟踪的竞争式传感器融合示意图

3) 协作式传感器融合

如果多个传感器存在紧密的耦合关系,则这类传感器融合称为协作式传感器融合,协作式传感器融合是最复杂的传感器融合。要求解这类传感融合问题,必须获得各个传感器的物理属性信息,并寻求各个传感器测量参数在不同空间下的内在关系。例如,在三维重建中使用计算机视觉算法将多个摄像机的图像信息耦合在一起获得跟踪目标的立体信息。在无处不在的跟踪器系统研究项目中,Wagner^[74]提出的空间关系聚合使用了协作式传感器融合,即如果知道物体 A 相对于 B 的相对关系,物体 B 相对于物体 C 的相对关系,就能够推导出 A 相对于 C 的空间相对关系。

5.7 无标识跟踪注册算法的未来研究方向

本章讨论了多种无标识点增强现实注册算法,尽管它们均在某一领域获得了应用,但也存在局限性。随着增强现实技术以及相关研究领域研究进展,无标识点注册算法的发展方向可能集中在以下几方面。

(1) 更加复杂、耗时的计算方法的引入。在当前的增强现实应用中,能够使用的算法都具有很高的处理速度,以保证系统的整体性能。然而随着计算机处理速度的飞速提升,很多在增强现实早期研究中无法使用的算法逐渐得到了应用,这些算法大都具有良好的性能,但是需要进行大量的计算。可以预见,在不久的将来还

会有更多类似的算法被用于无标识点注册方法中。尽管近年来各种新的无标识点注册算法仍然不断涌现,但是对图像处理和机器视觉领域的原有算法进行整理和改进依然是取得新的进展的一种重要手段。

(2) 基于非标定摄像机的无标识注册算法。目前,已有的无标识点注册算法大多要求在使用之前对注册摄像机进行标定,以获得摄像机的径向失真参数和内部参数矩阵,并且要求在整个注册过程中摄像机的内部参数保持不变,这在一定程度上限制了无标识点注册算法的应用。使用变焦摄像机的难点主要在于变焦系统的摄像机模型难以确定,同时焦距的变化与沿着光轴方向的位置移动难以区分。近年来随着运动结构重建理论不断发展,有望在不久的将来出现在恢复场景结构的同时实现摄像机内外部参数在线标定的实时注册算法。

(3) 高级特征在算法中的使用。出于处理速度与复杂度的考虑,目前已有的各种算法主要使用图像中的兴趣点或是角点作为识别和处理的基本元素。由于点特征本身具有的信息量有限,所以需要大量的点特征才能够满足注册计算的需求。与点特征相比较,直线、椭圆、多边形等高级特征携带更多的环境信息,从而保证跟踪注册算法具有较高的精度和鲁棒性。未来的研究可能集中在如何快速提取和识别这些高级特征,以保证能够快速、准确地得到这些高级特征的先验特征数据。

(4) 主动视觉在跟踪注册算法中的应用。目前,增强现实系统中使用的跟踪摄像机一般都固定安装在头盔显示器上,随着用户头部的运动而运动。在这种情况下,计算机无法预知下一时刻摄像机的运动情况,也就无法预测会有什么样的图像特征出现在图像中。在机器人领域,同样使用摄像机充当机器人的“眼睛”。但是与增强现实系统不同的是,机器人的“眼睛”安装在一对伺服电动机上,可以通过计算机进行控制,主动搜索和寻找目标。如果将这种方法引入增强现实系统中,则注册摄像机能够在一定程度上主动搜索环境中的特征,并且在用户移动头部的时候,保持始终“盯”着某一方向。从而使得系统获取图像特征不再是被动的,其适应环境的能力也将在很大程度上得到提高。但是,这种主动视觉的注册方式增加了系统的复杂程度和实现难度,并且增加了系统误差出现的可能性。相信随着机器人技术的进一步发展和成熟,主动视觉在增强现实领域也会得到应用。

参 考 文 献

- [1] Marchand É, Chaumette F. Virtual visual servoing: A framework for real-time augmented reality// Proceedings of Computer Graphics Forum, Saarebrücken, 2002.
- [2] Davison A J. Real-time simultaneous localisation and mapping with a single camera// Proceedings of

- International Conference on Computer Vision, Nice, 2003.
- [3] Klein G, Murray D. Parallel tracking and mapping for small AR workspaces//Proceedings of International Symposium on Mixed and Augmented Reality, Nara, 2007.
- [4] Simon G, Fitzgibbon A W, Zisserman A. Markerless tracking using planar structures in the scene//Proceedings of IEEE and ACM International Symposium on Augmented Reality, Munich, 2000.
- [5] Lepetit V, Vacchetti L, Thalmann D, et al. Fully automated and stable registration for augmented reality applications//Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality, Tokyo, 2003.
- [6] Vacchetti L, Lepetit V, Fua P. Stable real-time 3D tracking using online and offline information. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(10): 1385 ~ 1391.
- [7] Vacchetti L, Lepetit V, Fua P. Fusing online and offline information for stable 3D tracking in real-time//Proceedings of Computer Vision and Pattern Recognition, Madison, 2003.
- [8] 商业三维运动结构重建软件 Boujou. <http://www.2d3sensing.com> [2009-1-12].
- [9] Lowe D G. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 2004, 60(2): 91 ~ 110.
- [10] Bay H, Tuytelaars T, Van Gool L. Surf: Speeded up robust features//European Conference on Computer Vision, Graz, 2006.
- [11] Ozuysal M, Calonder M, Lepetit V, et al. Fast keypoint recognition using random ferns. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(3): 448 ~ 461.
- [12] Zhang G F, Dong Z, Jia J, et al. Keyframe-based real-time camera tracking//International Conference on Computer Vision, Kyoto, 2009.
- [13] ARCHEOGUIDE-Augmented Reality-Based Cultural Heritage on-site Guide. <http://archeoguide.intranet.gr> [2009-10-12].
- [14] Vlahakis V, Karigiannis J, Tsotros M, et al. Archeoguide: First results of an augmented reality, mobile computing system in cultural heritage sites//Proceedings of the 2001 Conference on Virtual Reality, Archeology and Cultural Heritage, Glyfada, 2001.
- [15] Vlahakis V, Ioannidis N, Karigiannis J, et al. Archeoguide: An augmented reality guide for archaeological sites. IEEE Computer Graphics and Applications, 2002, 22(5): 52 ~ 60.
- [16] Harris C, Stephens M. A combined corner and edge detector//Proceedings of the Alvey Vision Conference, Manchester, 1988.
- [17] Skrypnik I, Lowe D G. Scene modelling, recognition and tracking with invariant image features//Proceedings of the Third IEEE and ACM International Symposium on Mixed and Augmented Reality, Arlington, 2004.
- [18] Gordon I, David G L. Scene Modelling, Recognition and Tracking with Invariant Image Features.

- <http://www.cs.ubc.ca/~lowe/papers/gordon>[2007-9-10].
- [19] Montemerlo M, Thrun S. FastSLAM 2.0: An improved particle filtering algorithm for simultaneous localization and mapping that provably converges//Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, Acapulco, 2003.
 - [20] Davison A J. Real-time 3D SLAM with wide-angle vision//Proceedings of IFAC Symposium on Intelligent Autonomous Vehicles, Lisbon, 2004.
 - [21] Pupilli M, Calway A. Real-time camera tracking using known 3D models and a particle filter//International Conference on Pattern Recognition, Hong Kong, 2006.
 - [22] Pupilli M, Calway A. Real-time visual SLAM with resilience to erratic motion//IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, 2006.
 - [23] Pupilli M, Calway A. Particle filtering for robust single camera localisation//First International Workshop on Mobile Vision, Graz, 2006.
 - [24] Nistér D. An efficient solution to the five-point relative pose problem. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(6): 756 ~ 770.
 - [25] Durlach N I, Mavor A S. Virtual Reality: Scientific and Technology Challenge. Washington DC: National Academic Press, 1995.
 - [26] Azuma R T, Hoff B R, Neely H E, et al. Making augmented reality work outdoors requires hybrid tracking//Proceedings of the First International Workshop on Augmented Reality, San Francisco, 1998.
 - [27] Azuma R T. The challenge of making augmented reality work outdoors//Proceedings of the First International Symposium on Mixed Reality, Yokohama, 1999.
 - [28] Curtis D, Mizell D, Gruenbaum P, et al. Several devils in the details: Making an AR app work in the airplane factory//Proceedings of the 1st IEEE and ACM International Workshop on Augmented Reality, San Francisco, 1998.
 - [29] Foxlin E, Harrington M. WearTrack: A self-referenced head and hand tracker for wearable computers and portable VR//Proceedings of The Fourth IEEE International Symposium on Wearable Computers, Atlanta, 2000.
 - [30] Feiner S, Macintyre B, Hollerer T, et al. A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment//Proceedings of the IEEE First International Symposium on Wearable Computers, Boston, 1997.
 - [31] Huollerer T, Feiner S. Situated documentaries: Embedding multimedia presentations in the real world//Proceedings of the 3th IEEE International Symposium on Wearable Computers, San Francisco, 1999.
 - [32] Piekarski W, Gunther B, Thomas B. Integrating virtual and augmented realities in an outdoor application//Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality,

- San Francisco,1999.
- [33] Azuma R, Hoff B, Neely H I, et al. A motion-stabilized outdoor augmented reality system// Proceedings of IEEE Virtual Reality, Houston,1999.
 - [34] Behringer R. Registration for outdoor augmented reality applications using computer vision techniques and hybrid sensors//Proceedings of IEEE Virtual Reality, Houston,1999.
 - [35] Auer T, Pinz A. Building a hybrid tracking system; Integration of optical and magnetic tracking// Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality, San Francisco,1999.
 - [36] Bajura M, Neumann U. Dynamic registration correction in video-based augmented reality systems. IEEE Computer Graphics and Applications,1995,15(5):52~60.
 - [37] State A, Hirota G, Chen D T, et al. Superior augmented reality registration by integrating landmark tracking and magnetic tracking//Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, New Orleans,1996.
 - [38] Azuma R, Weon Lee J, Jiang B, et al. Tracking in unprepared environments for augmented reality systems,1999,23(6):787~793.
 - [39] Azuma R, Bishop G. Improving static and dynamic registration in an optical see-through HMD// Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques, Orlando,1994.
 - [40] Emura S, Tachi S. Multisensor integrated prediction for virtual reality. Presence; Teleoperators and Virtual Environments,1998,7(4):410~422.
 - [41] Persa S, Jonker P P. Hybrid tracking system for outdoor augmented reality//Proceedings of 2nd Int Symposium on Mobile Multimedia Systems and Applications MMSA2000, Delft,2000.
 - [42] Yokokohji Y, Sugawara Y, Yoshikawa T. Accurate image overlay on video see-through HMDs using vision and accelerometers//Proceedings of IEEE Virtual Reality, New Brunswick,2000.
 - [43] You S, Neumann U, Azuma R. Orientation tracking for outdoor augmented reality registration. IEEE Computer Graphics and Applications,1999,19(6):36~42.
 - [44] You S, Neumann U, Azuma R. Hybrid inertial and vision tracking for augmented reality registration//Proceedings of IEEE Virtual Reality, Houston,1999.
 - [45] You S, Neumann U. Fusion of vision and gyro tracking for robust augmented reality registration// Proceedings of IEEE Virtual Reality, Yokohama,2001.
 - [46] Kanbara M, Fujii H, Takemura H, et al. A stereo vision-based augmented reality system with an inertial sensor//Proceedings of IEEE and ACM International Symposium on Augmented Reality, Munich,2000.
 - [47] Satoh K, Anabuki M, Yamamoto H, et al. A hybrid registration method for outdoor augmented reality//Proceedings of IEEE and ACM International Symposium on Augmented Reality, New

- York, 2001.
- [48] Jiang B. Robust Hybrid Tracking for Outdoor Augmented Reality. Los Angeles; University of North California, 2004.
 - [49] Jiang B, Neumann U, Suya Y. A robust hybrid tracking system for outdoor augmented reality// Proceedings of IEEE Virtual Reality, Chicago, 2004.
 - [50] Satoh K, Uchiyama S, Yamamoto H. A head tracking method using bird's-eye view camera and gyroscope// Proceedings of the Third IEEE and ACM International Symposium on Mixed and Augmented Reality, Arlington, 2004.
 - [51] Maeda M, Ogawa T, Kiyokawa K, et al. Tracking of user position and orientation by stereo measurement of infrared markers and orientation sensing// Proceedings of the Eighth International Symposium on Wearable Computers, Washington DC, 2004.
 - [52] Lang P, Kusej A, Pinz A, et al. Inertial tracking for mobile augmented reality// Proceedings of the 19th IEEE Instrumentation and Measurement Technology Conference, Anchorage, 2002.
 - [53] Ribo M, Lang P, Ganster H, et al. Hybrid tracking for outdoor augmented reality applications. Computer Graphics and Applications IEEE, 2002, 22(6): 54 ~ 63.
 - [54] Rehbinder H, Ghosh B K. Pose estimation using line-based dynamic vision and inertial sensors. IEEE Transactions on Automatic Control, 2003, 48(2): 186 ~ 199.
 - [55] Rehbinder H, Ghosh B K. Rigid body state estimation using dynamic vision and inertial sensors// Proceedings of the 40th IEEE Conference on Decision and Control, Orlando, 2001.
 - [56] Rehbinder H, Ghosh B K. Multi-rate fusion of visual and inertial data// Proceedings of International Conference on Multisensor Fusion and Integration for Intelligent Systems, Kongresshaus, 2001.
 - [57] Gemeiner P, Vincze M. Motion and structure estimation from vision and inertial sensor data with high speed CMOS camera// Proceedings of the 2005 IEEE International Conference on Robotics and Automation, Barcelona, 2005.
 - [58] Gemeiner P, Einramhof P, Vincze M. Simultaneous motion and structure estimation by fusion of inertial and vision data. The International Journal of Robotics Research, 2007, 26(6): 591 ~ 605.
 - [59] Corke P, Lobo J, Dias J. An introduction to inertial and visual sensing. The International Journal of Robotics Research, 2007, 26(6): 519 ~ 535.
 - [60] Lobo J, Dias J. Vision and inertial sensor cooperation using gravity as a vertical reference. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(12): 1597 ~ 1608.
 - [61] Klein G S W, Drummond T W. Tightly integrated sensor fusion for robust visual tracking. Image and Vision Computing, 2004, 22(10): 769 ~ 776.
 - [62] Choi Y M, Suh Y S, Park S K. Pose estimation from landmark-based vision and inertial sensors// SICE-ICASE International Joint Conference, Bexco, 2006.

- [63] Reitmayr G, Drummond T W. Going out: Robust model-based tracking for outdoor augmented reality//Proceedings of IEEE/ACM International Symposium on Mixed and Augmented Reality, Santa Barbara, 2006.
- [64] Chandaria J, Thomas G, Bartczak B et al. Real-Time Camera Tracking in the MATRIS Project. <http://www2.cvl.isy.liu.se/ScOut/Publications/PaperInfo/ibc06.html>[2009-7-8].
- [65] Chandaria J, Thomas G A, Stricker D. The MATRIS project: Real-time markerless camera tracking for augmented reality and broadcast applications. Journal of Real-Time Image Processing, 2007, 2(2,3): 69 ~ 79.
- [66] Hoff W A. Fusion of data from head-mounted and fixed sensors//Proceedings of First International Workshop on Augmented Reality, San Francisco, 1998.
- [67] Hoff W A, Nguyen K, Lyon T. Computer-vision-based registration techniques for augmented reality. Intelligent Robots and Computer Vision XV: Algorithms, Techniques, Active Vision and Materials Handling, Boston, 1996.
- [68] Satoh K, Uchiyama S, Yamamoto H, et al. Robot vision-based registration utilizing bird's-eye view with user's view//Proceedings of The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, Tokyo, 2003.
- [69] Foxlin E, Altshuler Y, Naimark L, et al. FlightTracker: A novel optical/inertial tracker for cockpit enhanced vision//Proceedings of the Third IEEE and ACM International Symposium on Mixed and Augmented Reality, Arlington, 2004.
- [70] Klein G, Drummond T. Sensor fusion and occlusion refinement for tablet-based AR//Proceedings of the Third IEEE and ACM International Symposium on Mixed and Augmented Reality, Arlington, 2004.
- [71] Yamazoe H, Utsumi A, Tetsutani N, et al. Vision-based human motion tracking using head-mounted cameras and fixed cameras//Proceedings of the Sixth Asian Conference on Computer Vision, Jeju Island, 2004.
- [72] Foxlin E, Harrington M. WearTrack: a self-referenced head and hand tracker for wearable computers and portable VR//Proceedings of The Fourth IEEE International Symposium on Wearable Computers, Atlanta, 2000.
- [73] Welch G, Bishop G. SCAAT: Incremental tracking with incomplete information//Special Interest Group on Graphics and Interactive Techniques, Los Angeles, 1997.
- [74] Wagner M. Tracking with Multiple Sensors. Munchen: Technische Universitat Munchen, 2005.

第 6 章 移动增强现实系统

6.1 概 述

移动增强现实(mobile augmented reality, MAR)就是将增强现实技术有效地应用于移动设备终端。移动增强现实系统运行平台的进展如图 6.1 所示。早期的 AR 系统受限于硬件的发展,多选择 PC 作为系统运行平台,要求用户随身佩戴头盔显示器、摄像机、数据手套以及惯性跟踪器等跟踪定位装置和人机交互设备。该阶段的移动计算终端包括便携式计算机、Backpack(后背式计算机)、Tablet-PC 等。随着硬件技术的发展和软件算法的不断研发,小型移动设备(如 PDA 及智能手机)开始普及。



图 6.1 移动增强现实系统运行平台的进展

美国哥伦比亚大学研制的户外 AR 系统“Touring Machine”[图 6.2(a)]、美国海军研究中心研制的战场增强现实系统 BARS(battlefield augmented reality system)[图 6.2(b)]以及南澳洲大学开发的“Tinmith Project”是移动增强现实系统的前期典型代表[图 6.2(c)]。“Touring Machine”是第一套已知的户外移动增强现实系统^[1],系统由可穿戴式计算机作为运行平台,采用磁力计、倾角计以及 GPS 实现用户头部的六自由度跟踪定位,在 HMD 上实现虚实场景的融合显示。由于系统单纯

采用硬件跟踪设备实现用户的跟踪定位,其跟踪定位精度低,位置误差可达数十米。BARS 的目的在于扩展单兵作战人员对环境的感知视野,提高其协同作战能力^[2],该系统采用光学透视式 HMD 作为虚实场景融合显示设备,利用惯性陀螺传感器、数字罗盘以及差分 GPS 提供用户的位置和方向信息,并与基于模型的视觉跟踪定位数据进行融合,实现精确的六自由度方位跟踪定位。系统同样由可穿戴式计算机作为运行平台。BARS 的局限性在于系统初始化需要大量的人工干预。南澳洲大学在其一系列基于 Tinmith 框架的应用中(如 ARQuake 游戏和户外交互实验),同样采用视频透视式 HMD 作为显示设备,利用 GPS、电子罗盘和基于标识物的视觉检测相结合的方法实现跟踪定位^[3]。由于采用了基于标识物的视觉检测方法,严重限制了系统的使用范围。上述移动增强现实系统的整体重量一般为 4 ~ 10kg,在使用和操作上具有一定的局限性。且因成本高、易损坏以及难以维护等致命缺点导致其难以在实验室以外的环境中应用。



图 6.2 户外移动增强现实系统

随着移动计算技术的发展,小型手持式移动设备,如 PDA (personal digital assistant)、智能手机的出现为移动增强现实系统的发展提供了崭新的途径。小型移动设备不仅具有便携的特点,而且由于其内部嵌入了操作系统,具有较强的计算和处理能力;同时配备了触摸屏,能够方便地进行人机交互。目前,大多数主流移动设备都内置了百万甚至千万像素级的 CCD/CMOS 摄像头,内置了与通信基站协同工作的 GPS 定位设备,同时集成了 WiFi、红外通信接口技术以及蓝牙无线接口等高速无线通信网络设备。选择小型移动设备作为增强现实技术的新载体,将其与增强现实技术相结合,能够使增强现实系统脱离体积庞大的普通 PC、摄像设备、头盔显示器等的限制。在人机交互性、便携性方面具有较强的优势。此外,小型移动设备能够支持多用户同时获得、查看增强信息;对所获得的增强信息或场景能够进行静态处理(如放大、缩小操作);通过无线网络与远程服务器连接,方便地下载

和更新软件,这些都为系统的维护和更新提供了较大的灵活性。而将增强现实技术应用于小型移动设备也将扩展产品的现有功能,使其更具实用性和娱乐性。因而,基于手持式移动终端的增强现实系统在某些特别要求便携和人机交互的应用领域,如古迹数字重现、城市漫游、室内导览、危险预警与疏散、市政规划、电子商务等方面,比基于头盔显示器的增强现实系统更具优势。

6.2 移动增强现实系统的研究现状

移动增强现实是近几年来计算机互联网领域讨论比较频繁且受关注度非常高的话题之一,与其相关的软件应用和开发也呈大幅上升趋势。而随着 iOS 和 Android 手机操作系统的推广,特别是从 2007 年 Google 推出了开放的 Android 平台之后,移动增强现实技术更加受到了广泛的关注与推崇。美国《时代》周刊在 2010 年将移动增强现实技术列为引领未来的十大科技趋势之一,该国的巴特尔研究所也将 AR 技术列为 10 个 2020 年最具战略意义的技术趋势之一。Juniper Research 几年前发表的一篇研究报告预测,到 2014 年,移动 AR 市场的年销售收入将从 2009 年的不到 100 万美元增长到 7.32 亿美元,移动 AR 浏览器的安装用户数量也将从 60 万迅速增至 200 万。

由于移动增强现实巨大的商业价值,国内和国外的研究者都瞄准了该市场,以能为消费者使用和接受的移动增强现实应用作为主攻方向,争先占领该市场。40 亿手机用户一旦和 AR 技术相结合,将碰撞出巨大的火花。其应用将有助于用户从一个全新的维度理解周围世界,并与之互动。另外,移动 AR 技术凭借其灵活性、交互性和它所带来的高关注度、高参与度等优势,无疑将成为市场创新应用的热点。相信在不远的未来,移动增强现实的应用将会充斥着人们生活的方方面面。

6.2.1 移动增强现实技术的学术研究

充分利用移动设备以及增强现实技术的优势,将增强现实技术与手持式移动终端相结合已成为许多国外著名企业和研究机构的研究热点。国外移动增强现实技术开展的研究工作较早,发展很快,已经有很多研究成果和技术积累。例如,微软公司、英特尔公司、西门子公司、诺基亚公司、高通公司、谷歌公司以及美国、欧洲、日本、奥地利、瑞典、新西兰等国家和地区的研究机构都在从事这方面的研究工作。本节将详细介绍手持式移动增强现实系统的技术发展历程。

2000 年,由英国剑桥大学 AT&T 实验室的 Newman 等开发的 Batportal 系统

[图 6.3(a)],采用超声波跟踪设备定位 PDA 以及用户的位置和方位信息,实现 2D 与 3D 信息的可视化显示^[4]。该研究为增强现实技术在 PDA 上的应用提供了前期探索,并证明了其可行性。2003 年,维也纳理工大学的 Wagner 等研究了首款能够在 PDA 的内部处理器上独立运行的增强现实系统,如图 6.3(b)所示^[5]。系统将 ARToolkit 软件开发包移植到 Pocket PC 操作平台,用定点数代替浮点数运算,提高了系统的运算效率。该款手持式增强现实系统采用 Pocket PC 2002 (HP iPAQ 5450) 手持设备,设备配置了 400 MHz Intel 袖珍型处理器,320×240 像素的 16 位显示器,320×240 像素的彩色摄像头,64MB RAM 以及 IEEE 802.11b 无线网络接口。该款手持设备的主要缺点是内存小、工作电量有限,在执行图形和视觉任务时不能处理任何浮点和符号类型的数据。此外,该 Pocket PC 没有 3D 图形加速卡,甚至连 2D 图像都不能流畅加速。摄像头也仅能以 6.2Hz 的频率传输图像,比市售的低端 USB 网络摄像头性能还要差。尽管如此,这款手持 AR 系统的研发人员将大量的精力花在了对产品软件部分的研发,使得该手持式平台有限的硬件设备得到了最佳发挥,并能出色地完成交互功能。在此工作基础上,该研发组开发了第一个多用户手持 AR 系统的应用——“无形火车”(invisible train)游戏,如图 6.3(c)所示^[6]。“无形火车”是一款简单的多人游戏,不同玩家可以操纵虚拟火车在真实的木质微型铁轨上行驶,玩家能够通过各自的手持 AR 系统看到虚拟火车,而这些火车在真实世界中并不存在。玩家可通过手持 AR 系统的触摸屏完成切换铁轨和调节虚拟火车行驶速度等操作。



图 6.3 早期基于移动设备的增强现实应用

为了更接近普通消费群体,德国的魏玛包豪斯大学利用普通的诺基亚手机开发了一款手持式增强现实系统,如图 6.4(a)所示^[7]。系统采用基于标识的跟踪注册算法,并通过帧间信息的使用极大地缩小了跟踪区域,排除错误结果的干扰,提高了系统的运行效率。系统支持 16 帧/s、160×120 像素、12bit 颜色深度图像的发

送。识别标识点和读取条形码约需占用 25% 的运行时间。因为缺乏浮点运算功能,所以系统需要快速的定点算法,以避免除法运算。2004 年瑞典和新西兰的几所大学联合开发了基于手机的双人协作 AR 应用系统。系统采用 ARToolkit 进行标识点跟踪,采用蓝牙进行 P2P 数据通信,利用 OpenGL ES 进行三维渲染,如图 6.4(b) 所示,用户通过手机控制虚拟的乒乓球^[8]。

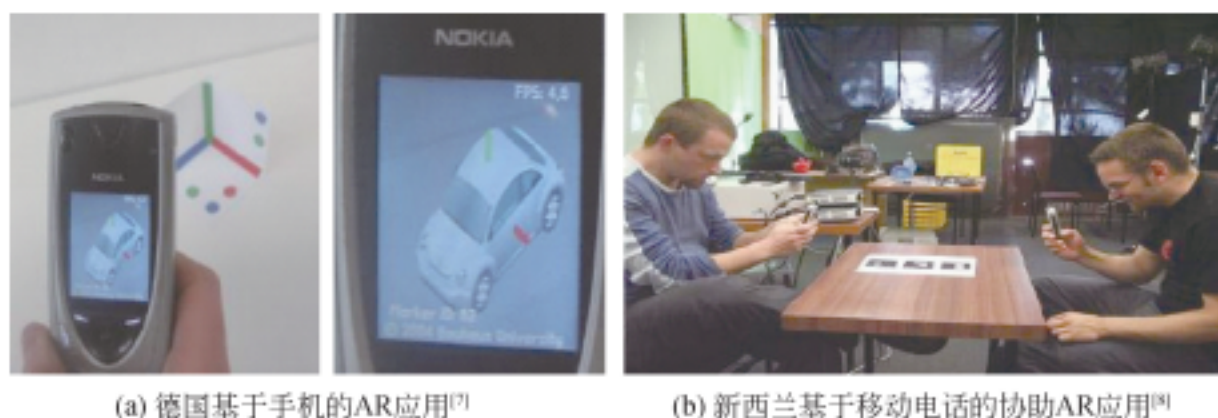


图 6.4 基于标识的增强现实技术在手机上的应用

上述手持移动增强现实系统均采用基于标识的跟踪定位技术,因此系统仅适用于光照变化可控、场景固定且允许放置人工标识的场合。为使手持式增强现实系统能够应用于环境复杂且无法进行人工干预的场合(如户外历史遗址、野外战场、城市街道等),需采用更加智能的跟踪定位方式。这对于推进手持式移动 AR 系统真正得以普及和实用化具有举足轻重的意义。2006 年诺基亚研究中心在其 MARA(mobile augmented reality applications)开发项目中,利用诺基亚 6680 款手机作为增强现实系统运行平台,采用加速计、倾角仪以及 GPS 实现手机的绝对姿态跟踪定位,通过蓝牙无线接口向手机发送测量数据信息。在手机的显示屏上为用户提供其所处户外场景的导航信息,如图 6.5(a) 所示^[9]。同年,英国剑桥大学机器视觉组开发出手持式户外增强现实导航系统 Going Out,该系统融合了来自惯性传感器与视觉测量的数据,跟踪定位用户的位置和方向信息,在 Tablet PC 上显示周围环境的增强场景信息,如图 6.5(b) 所示^[10]。2007 年德国魏玛包豪斯大学的研究组在手机上开发了 AR 博物馆导览系统,如图 6.5(c) 所示^[11]。系统采用蓝牙发射、接收装置定位手机用户的位置信息,在手机上实现基于 SIFT 特征点的无标识目标识别分类功能。

2008 年,Wagner 等对 SIFT 和 Ferns 算法作了较大改进,用 FAST 算子代替 SIFT 和 Ferns 中的 Laplacian 算子,并改进了 SIFT 特征的描述向量的维数以及 Ferns 算法中树簇的大小,提高了二者在智能手机上的计算速度^[12]。2009 年,英国



图 6.5 基于自然场景的增强现实技术在手机上的应用

牛津大学的机器视觉组采用运动结构重建方法,在手机上实现了未知环境下的实时跟踪定位算法 PTAM,如图 6.6(a)所示^[13]。2010 年,U-VR Lab 和 CVLab 联合提出了适合在跟踪识别过程中处理模糊图像的方法,图 6.6(b)所示^[14],此方法中,存储图像数据库阶段需要对图像进行学习训练,对输入的图像依次进行仿射变换、径向模糊、高斯模糊和存储。完成学习训练后,再将手机用于目标的跟踪定位。2013 年,斯坦福大学的 Chen 等在手机上实现了大规模图片的检索,如图 6.6(c)所示^[15,16]。



图 6.6 基于自然场景的增强现实技术在手机上的应用

6.2.2 移动增强现实技术的商业应用

尽管技术还不成熟,但移动 AR 技术发展势头强劲,研发领域主要集中在城市道路与建筑导航、室内游览引导、零售行业、娱乐和游戏应用、汽车与航空航天、工业维修、文化遗产保护、移动电子商务、移动检索等。特别是最近几年,伴随着 iOS、Android 高端手机平台的推出,出现了一些非常有价值的应用。除了上述著名公司

以及研究机构之外,欧美等国家和地区更是诞生了许多专门从事手机移动 AR 开发的研究公司,如 Layar、Metaio、Mobilizy、ARToolworks、Acrossair 等。针对目前流行的商用嵌入式移动平台,如 Nokia 的 Symbian OS(C/C++)、iPhone 的 Cocoa(object-C/C++)、Google 的 Android(Java),都有对应的 AR 实现。

目前,基于手机的移动增强现实技术应用主要分为三大类。一类是基于位置的服务,这类应用首先获取智能手机的 GPS、电子罗盘信息,再通过网络获取增强信息。其中典型的应用包括 Layar 浏览器、Wikitude 浏览器、Yelp iPhone 等。Layar 是全球第一款增强现实感的手机浏览器。用户将手机摄像头对准目标建筑物,即可获得与之相关的各种信息,如图 6.7(a)所示^[17]。Layar 将现实世界跟虚拟的数字化信息内容完美地结合在一起。Wikitude 还可以实现对目标的识别,如图 6.7(b)所示^[18]。通过此类系统应用,人们可以通过手机搜索所在位置附近的饭店、企业和地标等信息,极大地方便了用户的生活。

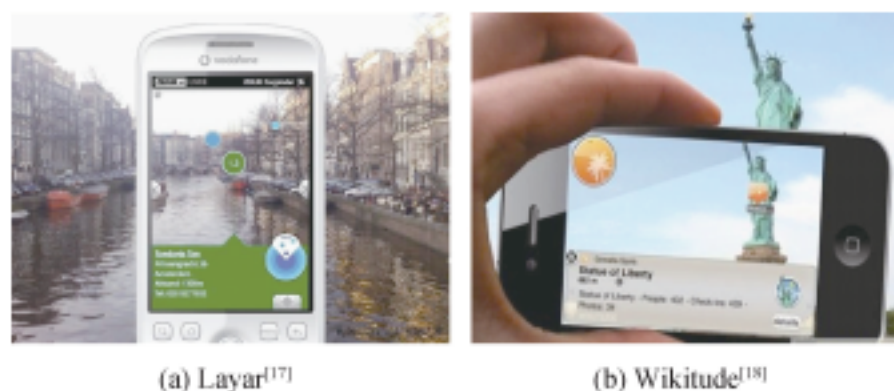


图 6.7 增强现实在手机上的导航应用

第二类是通过跟踪注册目标,应用到游戏开发以及设备维修等方面,如高通公司的 Vuforia^[19]、Metaio 公司的移动 AR SDK^[20]。图 6.8(a)是开发于 iPhone 手机上的一款简单 AR 颠球游戏 ARSoccer^[21],当用户进入游戏状态时,手机屏幕上会自动渲染生成一个虚拟的足球。用户仅需将手机保持腰间高度,摄像头对准脚步就可以玩颠球游戏了。该类应用需要有较强的实时性,对跟踪定位算法的运行速度有较高的要求。2011 年 2 月,著名的软件开发公司 Total Immersion 推出一款帮助消费者选购眼镜的手机 AR 软件,如图 6.8(b)所示^[22],该软件利用带有前置摄像头的手机获取用户的面部图像,对面部进行识别、跟踪定位,并在手机屏幕上显示佩戴虚拟眼镜的增强效果。

第三类主要是利用图像检索或目标识别技术,通过摄像头获取目标图像,利用目标识别算法获得识别结果。进而从网络获取增强信息,并将增强信息叠加



(a) 颠球游戏ARSoccer^[21] (b) 试戴虚拟眼镜^[22]

图 6.8 增强现实在手机上的娱乐应用

在真实的画面中,达到虚拟与真实的完美结合。手机移动增强现实在视觉搜索方面的应用已经商业化,目前已有不少提供移动视觉搜索服务的公司。典型应用有谷歌的 Goggles [图 6.9 (a)]^[23]、Amazon 的 Snaptell [图 6.9 (b)]^[24]、Kooaba [图 6.9 (c)]^[25] 和 Nokia 的 Point&Find [图 6.9 (d)]^[26]。其中,谷歌的 Goggles 是该类系统的典型代表,是一个里程碑式的应用程序。Goggles 可以识别文字、书籍、商标、艺术品等物品。该类系统的优点是识别物品的种类非常多,缺点是受网络通



(a) Goggles^[23]

(b) Snaptell^[24]

(c) Kooaba^[25]

(d) Point&Find^[26]

图 6.9 基于手机平台的图像检索

信延迟以及大数据量识别算法的影响,识别速度较慢。即使实现了实时识别,如 Point&Find,持续网络通信带来的上网流量也会给用户带来不小的话费开支。

此外,2012年4月,谷歌公司发布了一款“增强现实”眼镜 Google Glass^[27],可以对用户所在的真实场景进行增强信息叠加显示,如图 6.10 所示。Google Glass 具有和智能手机一样的功能,它配备了投影显示器、拍摄视频的摄像头、麦克风、喇叭以及各种传感器和陀螺仪,还有多种通信模式,可以通过声音控制拍照、视频通话、辨别方向、上网冲浪、处理文字信息和电子邮件等。

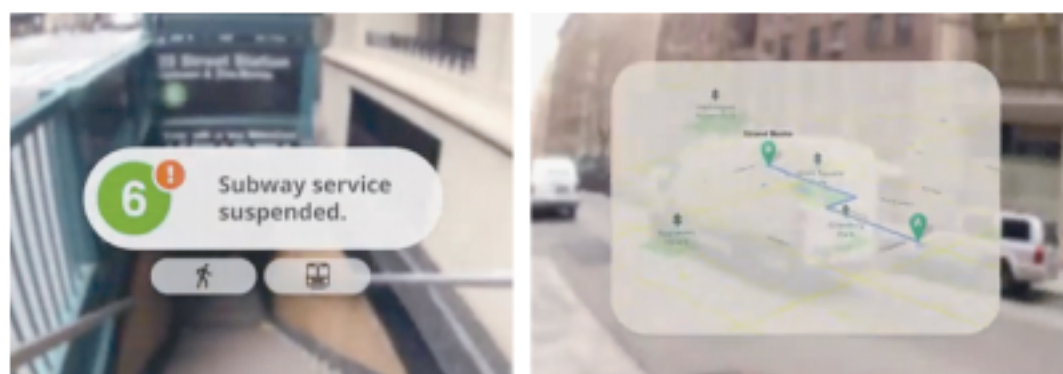


图 6.10 谷歌推出的 Google Glass^[27]

微软公司也相继推出了 Microsoft Glass,其中使用了增强现实技术^[28]。Microsoft Glass 的眼镜镜片是 Windows Glass 的系统显示屏幕,当用户通过眼镜镜片看向外部实物时,系统生成的虚拟物体、场景、提示信息会在屏幕上以叠加的形式与物体同时出现,如图 6.11 所示。

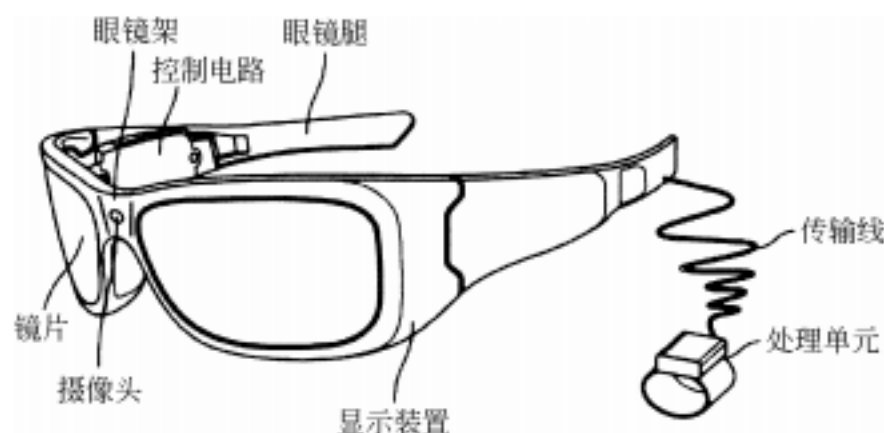


图 6.11 微软推出的 Microsoft Glass^[28]

6.3 面向智能手机的移动增强现实核心技术

将 AR 技术应用于手机终端,意味着系统能够提供以下主要功能模块:视频获取设备、实现高精度定位功能的专用方位跟踪装置、大量复杂的计算机视觉运算功能、3D 渲染功能、人机交互功能,这些功能的实现将受处理器、内存等手机硬件问题的约束。由于目前手机芯片在处理速度、存储容量以及色彩显示等方面仍然无法与 PC 相比拟。将 AR 技术应用于其上,就需要综合考虑采用何种系统主体结构框架,如何实现快速高效的跟踪定位算法、大场景范围下的多目标识别、手机与用户间的自然交互、网络传输延时以及手机耗电量等多方面的问题。

(1)在主体框架方面,可以将系统应用分为基于客户端/服务器(C/S)的方式和完全基于手机的方式。总的来说,前者虽然可以使用 PC 或大型工作站分担部分计算负载,但由于无线网络数据传输的延时、较窄的带宽等因素,该技术无法达到实时处理的效果。而完全基于手机的方式,需将跟踪识别目标算法和渲染显示图像技术全部在手机运行,大大增加了手机的运算量和耗电量。因此,需要研究手机电池克服其电量不足的难题。

(2)在硬件的发展上,智能手机存在一些缺点,如有限的屏幕尺寸和存储空间、缺少浮点运算单元(FPU)等,使得 PC 上运行的算法无法被有效移植,大大影响了在手机上对真实场景增强的效果。对于此,硬件研究人员也在积极地研制,使智能手机与 PC 的差距逐渐缩小,具备更强大的运算能力。

(3)在移动增强现实的算法设计方面,基于 PC 的增强现实技术的算法移植至手机上,是一种被动的形式,迫切需要对原有的算法进行优化,降低其对运算能力的要求,缩短运算时的存储空间,提高运算速度,从而使其更好地应用在手机上。

(4)针对智能手机研制的算法,尤其是多用户之间可以进行交互的模式,复杂性较高,所占的内存较大,可以考虑将其打包压缩后移植到手机上,以滞留出足够的空间满足运算要求。

(5)在室内、室外不同场景下使用时,需考虑光照对目标物体的影响。智能手机用户在户外使用时,太阳光的明暗变化十分显著,手机摄像头拍摄到的景物光线也会随之变化,这会对基于自然特征点的跟踪算法产生极大的干扰,影响目标特征匹配的准确率。因此,需要根据不同的场景需求,综合考虑各种环境的变化。

6.3.1 主体结构框架

手机移动 AR 技术整体架构最初受限于手机处理能力,主要计算运行于服务器端,手机只作为图像采集和显示终端。随着手机处理能力的大幅提高,智能手机和 PC 分别承担处理任务的客户端/服务器方式以及手机完全脱离 PC 进行独立处理的方式被越来越多的研究人员所采用。图 6.12 显示了四种不同的客户端/服务器实现功能模块的划分方式。通过对比可以看出,功能强大的手机移动 AR 系统无疑是集图像采集、姿态估计、图像处理、图像渲染以及显示功能于一体的系统,如图 6.12(a)所示。如果将该流程中的一部分任务交给服务器解决,往往可以提高系统的运行速度。但是,此时必须谨慎考虑所需要的网络带宽,如图 6.12(b)~图 6.12(d)所示。图 6.12(b)和图 6.12(c)显示的流程是将跟踪的任务交给了 AR 计算机服务器。为了达到视觉跟踪的目的,可向服务器端传输经特别压缩预处理的图像数据,然后计算姿态信息,再传输至客户端。另一种方式是将跟踪和图像渲染、绘制全都交给服务器处理,如图 6.12(d)所示,在这种情况下,手持设备只是一个小巧轻薄的客户显示终端。

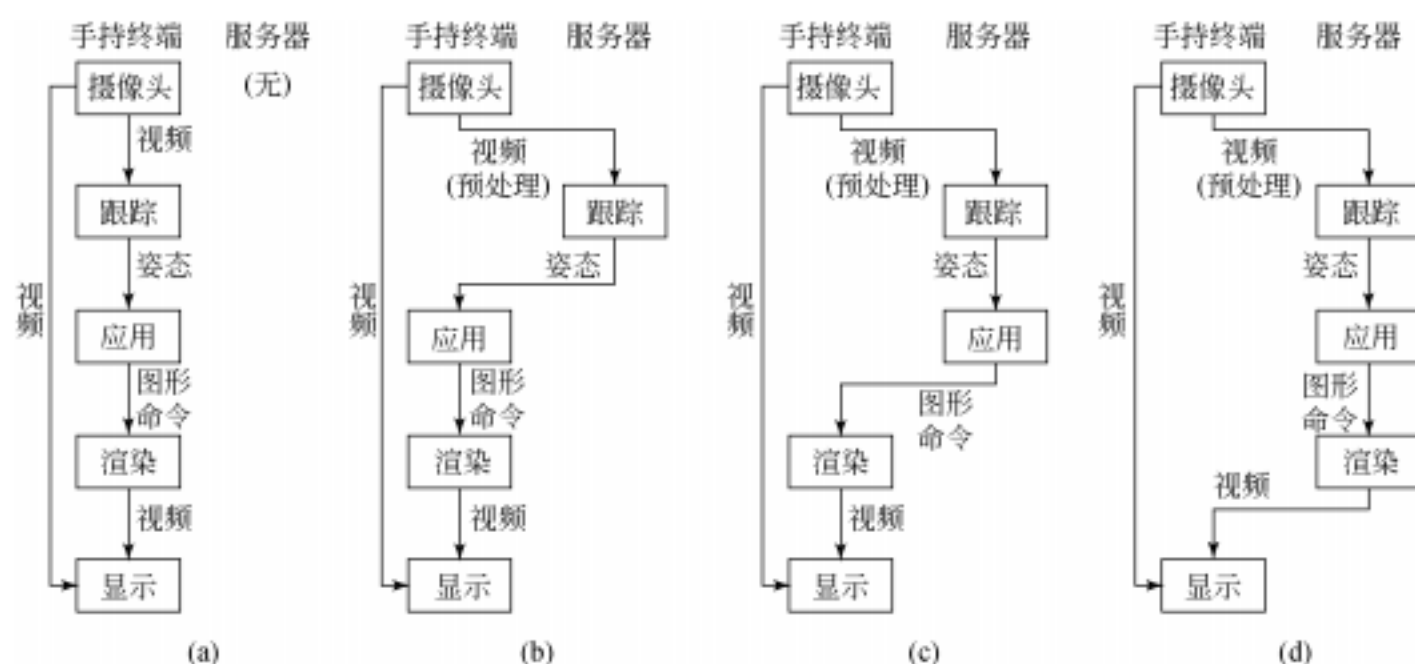


图 6.12 客户端/服务器实现功能模块的划分方式

早期 AR 技术的开发和应用都以 PC 或工作站作为系统运行平台,其主要原因是 PC、工作站拥有强大的计算能力以及图形显示能力。而早期的智能手机在处理能力和存储空间上都较为有限,无法处理在 PC 上运行的跟踪算法。因此,早期的以及后来的一些手机移动 AR 系统采用的多为客户端/服务器架构方式。

该架构的特点是手机只作为图像获取和显示设备,所有的视觉处理算法都由远程服务器负责完成,二者通过无线网络进行数据通信。早在 2003 年,西门子研究院的 AR-Phone 项目就采用该架构构建了基于智能手机平台的 AR 应用。系统通过蓝牙将手机所拍摄的图片上传至服务器,由服务器处理姿态估计算法,并将虚实融合的场景图像发送回手机客户端进行效果显示。受限于手机视频处理的功能,系统处理的时间较长,从图片上传至增强场景显示需要 2 ~ 16s^[29]。

随着手机芯片处理速度的提高,研究人员开始采用客户端与服务器分别承担计算处理任务的方式。文献[30]中的系统将渲染和显示部分移植到了手机中,如图 6.13(a)所示。文献[31]中的系统利用服务器进行目标的初识别,然后将初始识别结果传至客户端,在客户端实现目标的精确识别和渲染显示,如图 6.13(b)所示。基于 C/S 结构的系统作为传统 AR 技术向智能手机平台的初始过渡,仍然没有脱离 PC。虽然基于客户端/服务器结构的系统具有网络传输延时的缺点,限制了系统的进一步发展。但随着 3G 甚至 4G 无线网络的发展,一些基于此架构的企业级应用必将得到发展,该架构也契合了“云服务”的概念,有着较好的应用前景。目前基于手机的浏览器 Layar 和实景导航 Wikitude 的开发就是基于此架构,用户体验效果较好。

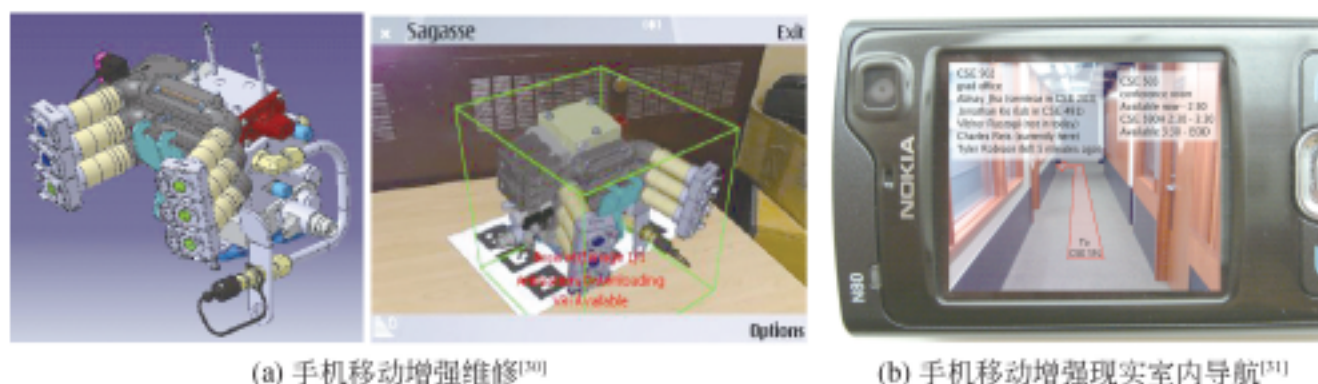
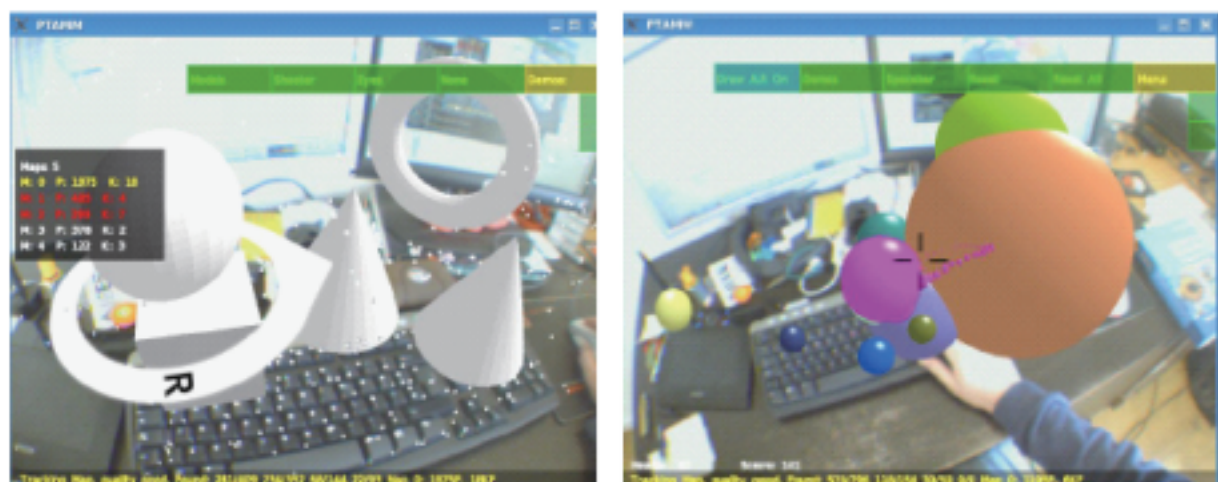


图 6.13 采用客户端/服务器模式的增强现实手机应用

近年来,研究人员也开始着手研究能够使 AR 技术完全摆脱 PC 而独立运行于手机上的各种处理方法,并取得了一定的进展。例如,Mark 小组将 ARToolkit 技术进行改进,并完全移植至 Symbian 手机操作系统中独立运行^[32]。2009 年,牛津大学提出了在手机上对未知环境进行三维重建并实现实时跟踪定位的思想。它完全脱离了 PC 和大型工作站,可以在 iPhone 手机上独立运行 PTAM 系统,此系统的目标跟踪、识别算法和渲染显示图像全部独立运行于手机,图 6.14 所示^[13]。尽管在手机芯片处理能力方面存在不足,但是依然可以独立对未知环境进行实时增强,在手机移动 AR 领域具有较大影响。

图 6.14 PTAM 运行结果^[13]

6.3.2 跟踪定位算法

研究人员针对 AR 跟踪定位算法的研究已开展多年,形成了相对成熟的算法框架。但是将 AR 技术移植到智能手机平台独立运行的研究工作才刚起步。将常用的 AR 跟踪定位算法移植到智能手机平台是实现手机移动 AR 技术的主要途径。基于人工标识算法移植方面,2003 年 Wagner 等将 ARToolkit 软件包移植到 Windows CE 平台,首次实现了能够完全独立运行于移动设备的 AR 应用系统^[5]。2005 年 Henrysson 等将 ARToolkit 移植至 Nokia 手机的 Symbian 操作系统,并成功开发出交互式乒乓球游戏^[8]。

为使手持式增强现实系统能够应用于环境复杂且无法进行人工干预的场合,需要采用更加智能的跟踪定位方式。自 2007 年起,研究人员开始将基于自然特征的跟踪算法移植或应用于智能手机的研究工作。现阶段的自然特征点提取算法主要有 SIFT 算法、SURF 算法和 FAST 角点探测算法等。2005 年, Rosten 等提出了 FAST(features from accelerated segment test)角点探测算法,该算法选择图像中任意一个像素点和以它为中心的一个圆形区域,通过角点响应函数判断候选点是否为角点。该算法运行速度快,在智能手机上使用可以大大缩短运算时间^[33]。2006 年,Bay 等提出 SURF(speeded up robust features)算法,使用了积分图的概念,借助积分图像,使图像与高斯二阶微分模板的滤波转化为对积分图像的加减运算。积分图运算与滤波模板的尺寸无关,因此,该算法在运算速度上比 SIFT 算法快 3 倍左右,综合性能与 SIFT 算法相当^[34]。

2007 年,Nokia 研究中心的 Chen 等将 SURF 这一较常用的鲁棒特征提取算法

应用于智能手机,对 SUFR 算法进行了改进,降低了算法所需的存储空间^[35]。2008 年 Daniel 等对 SIFT 和 Ferns 算法作了较大改进,用 FAST 算子代替 SIFT 和 Ferns 中的 Laplacian 算子,并改进了 SIFT 特征的描述向量的维数以及 Ferns 算法中的树簇的大小,提高了二者在智能手机上的运算速度,并成功应用于 Nokia N95 手机和 Motorola Q9 手机上,达到 10 帧/s 左右的处理速度^[12]。

为解决基于视觉的跟踪定位算法的鲁棒性以及稳定性问题,目前也有研究人员利用硬件跟踪定位设备进行 AR 跟踪定位。MARA 是由 Nokia 公司开发的城市导游系统,它利用诺基亚 6680 款手机作为增强现实系统运行平台,采用加速计、倾角仪以及 GPS 实现手机的绝对姿态跟踪定位,并通过蓝牙无线接口向手机发送测量数据信息^[9]。虽然早期该系统看起来很笨拙,但是其跟踪定位理念被后来取得较大成功的 Layar 以及 Wikitude 发扬光大,实现了商业化运营。

文献[36]中的 PAC-LAN 是一个多人城市游戏系统,用户的空间位置是由 RFID 技术进行定位的。首先在游戏进行的真实环境中布置多个 RFID 接收器,每个用户的智能手机上都安装一个 RFID 发射器,通过该方式确定用户的空间位置[图 6.15(a)所示为一个玩家在城市的角落放置 RFID 接收器,图 6.15(b)所示为整个游戏的路线图]。与 GPS 不同,RFID 不受地形和高大建筑物遮挡的影响,可以在环境中布置多台发射设备。由于配备 RFID 发射器的智能手机在市场上还不多见,而带有蓝牙技术已经成为智能手机中的必备工具。因此,采用蓝牙定位也成为一些系统的选择。Bruns 等构建了一个博物馆导览系统,该系统在博物馆内均匀布置蓝牙发射器,每个蓝牙发射器的有效范围可以覆盖至一些特定展品。当用户接近某一展品时,用户的手机将接收到发射器发送的蓝牙信号,并由此估计用户的当前位置信息(图 6.16)^[37]。

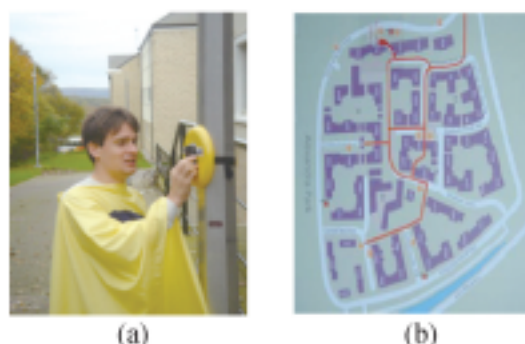
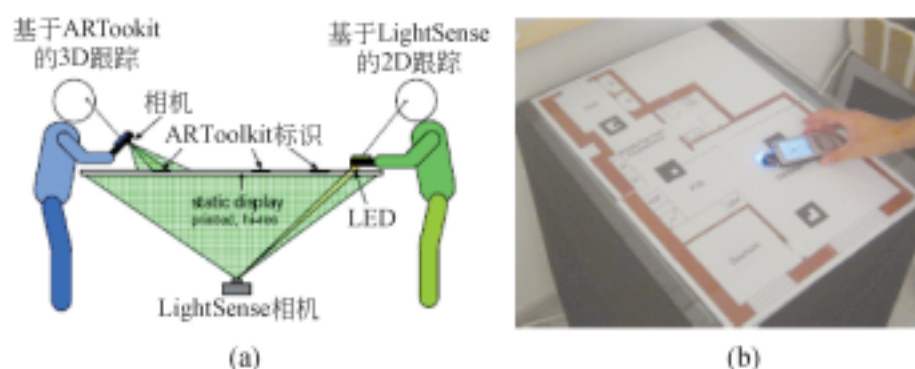


图 6.15 使用 RFID 进行空间定位的 PAC-LAN 城市游戏系统^[36]

此外,Olwal 等构建了 LightSense 系统,利用固定在智能手机背面的 LED 灯以及环境中的灯光传感器来跟踪手机的空间位置。系统中,用户所面对的是固定在

图 6.16 使用蓝牙 RFID 进行空间定位的博物馆导览系统^[37]

玻璃板上的可以透射的平面地图或者地铁路线图。玻璃板的后面是灯光传感器或定位摄像机。当用户持手机对着地图表面移动时,定位摄像头通过分析所拍摄图像的灰度变化区域来判断手机所在的位置,如图 6.17(a)所示,并通过蓝牙将与这一位置和姿态所对应的虚拟信息传给手机进行显示,如图 6.17(b)所示^[38]。

图 6.17 使用 LED 灯和 ARToolkit 标识点进行空间定位的桌面投影增强现实系统示意图和用户操作结果图^[38]

6.3.3 海量场景的目标识别

当手机移动 AR 技术被用在大范围环境或者需要对场景中的物体进行信息增强时,有时系统不需要跟踪手机的三维位置,而只需识别当前场景中的目标,并将相应的虚拟信息叠加显示在图像中。此时,目标识别或图像检索技术是必不可少的。利用特征匹配实现目标识别,最简单的做法是将每幅图片用一个特征集合表示,输入需要识别的图片,将该图片的特征集合与数据库中每张图片的特征集合进行比较,特征匹配数目最多的几张图片作为查询结果返回。当数据库中的图片数量较多时,消耗在特征匹配上的时间会急剧增加,无法满足手机上的应用要求,因此需要研究更高效的检索算法。现阶段用于检索的主流高效算法主要有 kd-树算法、Spill-树、词汇树算法、随机树算法等。kd-树算法^[39]是一种由二叉搜索树推广

而来的用于多维检索的树结构形式(k 即空间的维数)。由于对数据建立了树型结构,因此能有效减少检索的数据量,提高检索速度。在此基础上,改进后的 BBF (best-bin-first) 算法可以运用到高维数据集上。Spill-树算法采用冗余分割方式,在 s 维度上有两个分割超面。分割超面 LR 左边为左子空间,分割超面 LL 右边是右子空间,两个子空间有一部分是重叠的,即重叠范围内的数据同时属于两个子空间^[40]。当查询点接近分界面 L 时,邻域实际上已经包含在它所处的节点一侧,不需要回溯搜索另一个节点,从而提高了搜索效率,如图 6.18 所示。词汇树算法将从训练集中提取到的特征进行 K -Means 聚类,生成的每个簇集定义为一个词汇,每个词汇再关联一个倒排文件^[41]。算法采用分层聚类的方法生成词汇树,使特征量化时不必遍历所有词汇,极大地缩短了量化所需的时间,如图 6.19 所示。随机树算法是将一些本来需要在实时检测阶段的运算转换至训练阶段完成^[42]。算法运行时,使用训练好的分类器对特征点进行匹配,该方法的关键是对分类器进行离线训练,这样可以大量减少实时系统的负担,提高系统的实时性能。

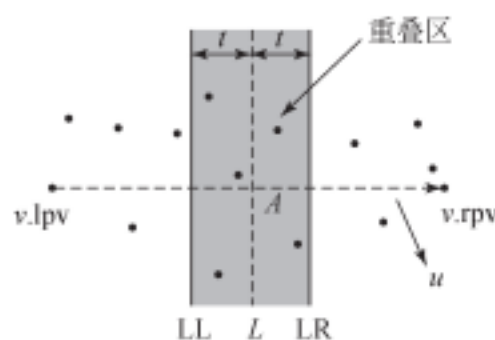


图 6.18 Spill-树空间划分示意图^[40]

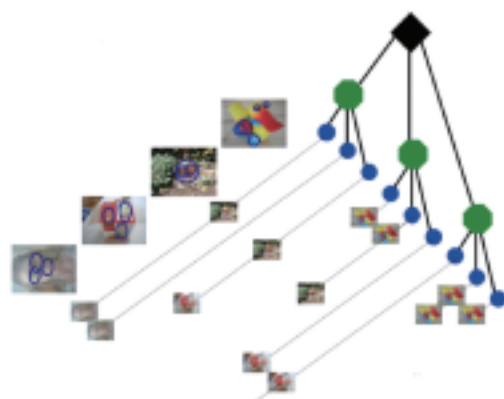


图 6.19 词汇树示意图^[41]

目前,在智能手机上实现 AR 技术多使用 GPS、摄像头和指南针等设备,如在前面提到的 Metro AR 软件和百度手机地图 Android 版软件,通过手机内置的硬件设备,粗略地估计用户的位置。之后,运用自然特征的跟踪方法确定其姿态。文献[43]中,由 GPS 初步确定用户的空间位置,由服务器在数据库找出与该位置相关的几个待识别物体,然后通过比较图像相似度达到识别物体的目的。两年后,该小组对系统中的物体识别算法进行了较大改进,使用 SIFT 特征进行物体识别,得到了更高的识别率^[44]。德国魏玛包豪斯大学的 Oliver 小组^[11]采用神经网络方法对所拍摄的博物馆展品进行识别。法国的 Inria 实验室近年来在移动视觉检索方面的工作相当引人注目,他们针对词汇树算法进行改进,先后提出了汉明嵌入法^[45]、VLAD 特征量化法等算法^[46]。汉明嵌入法可进一步提高基于词汇树的大规

模场景识别精度,VLAD 特征量化法可在减少聚类视觉单词的基础上保证其相应的识别精度。此外,斯坦福大学在移动视觉检索^[47]和哥伦比亚大学在移动终端大规模产品检索^[48]方面也有突出的贡献。

6.3.4 人机交互方式

与人机交互技术结合的移动 AR 系统摆脱了传统 AR 系统的单一交互模式,用户可通过语音、手写、肢体语言、表情等多样化的动作通道与计算机生成的虚拟信息进行交互,使得用户更自然地融入场景之中,并增强其用户体验感。智能手机平台上的人机交互技术发展日新月异,从 iPhone 的多点触控技术、Siri 语音控制技术到谷歌眼镜的滑动触控技术以及眼球跟踪技术,新型人机交互技术彻底颠覆了传统手机的定义。使智能终端具有了“视”、“听”、“触”的功能,掀起了移动互联网一波又一波的发展高潮,极大地改善了移动智能终端的用户体验,降低了 AR 应用的用户门槛。在视觉交互方面,谷歌公司打造了以移动终端(智能手机、平板计算机、智能眼镜)为平台的“所见即所知”式的移动视觉搜索互联网新服务范式。在手势体感交互方面,微软体感交互设备 Kinect 的面世,为当前的人机交互方式带来了重大变革。随后,美国 Leap 公司发布了其 Motion 手势识别产品,基于用户的生理计算也已逐渐引起人机交互研究者的关注。加州大学伯克利分校、德国马克斯-普朗克研究所、瑞士苏黎世联邦理工大学、英国帝国理工学院、微软研究院等已逐步开展相关研究。随着单通道人机交互技术的不断创新与发展,融合语音、视觉、行为、情感等多模态的自然人机对话交互技术日益受到关注。国际上许多著名的研究机构和跨国公司对此极其重视,形成了一系列相关研究项目,如欧盟第六框架计划项目 Cross Modal on Verbal and Nonverbal Communication 和美国麻省理工学院的 Oxygen 计划等。

基于新型交互技术形成了一个又一个产业开发生态圈,挖掘出移动互联网无穷无尽的价值,新型交互技术已成为影响智能终端和移动互联网应用发展方向的一个关键环节。而随着移动互联网逐渐深入渗透人们的生活,智能终端逐步向“穿戴式”方向发展,各种新型的人机交互技术也将不断涌现。

6.3.5 数据通信和传输

无论基于 C/S 模式的系统还是完全由智能手机独立运行的多用户系统,都需要进行数据的无线通信和传输。在 C/S 模式的系统中,客户端和服务端间的数据传输需通过 GPRS、GSM、WiFi 或 3G 网络完成。传输网络的选择首先要考虑用户

所处的应用环境支持哪种网络,其次要考虑网络的通畅性、带宽、传输速度以及传输距离等因素。其中 3G 网络属于最新的网络类型,具有较好的带宽和传输速度,但是目前只有部分智能手机支持该网络。GPRS 和 GSM 较为常用,但是在速度和带宽方面无法与 3G 网络相比。蓝牙技术虽然传输速度较慢,但是由于大多数手机都支持该技术,使用较为广泛。WiFi 技术与蓝牙一样,也属于短距离传输技术,其无线电波覆盖范围广,但是通信质量不尽如人意。但是无论采用何种无线网络通信,都不可避免地会造成系统处理时间因网络传输而增长,这也是最终要实现智能手机独立于服务器的一个主要原因。手机移动 AR 技术的推广和应用离不开通信技术的发展,信息的提供和传送需要客户端和服务端不断地进行数据传输,WiFi、4G 甚至 5G 网络的发展,为移动增强现实大数据量应用插上了翅膀。

6.3.6 功耗

使用移动增强现实系统的时候,往往要用到 GPS 定位、液晶显示、网络通信等比较耗电的功能,导致终端电力消耗过快,影响了用户体验。所以需要设计出灵活高效的调用方法来减少电量的消耗。对于 GPS,除了需要研究如何降低 GPS 带来的耗电问题之外,还需要研究如何在 GPS 不可用的场合帮助用户定位。例如,可采用各种基站 WiFi 等辅助定位方式,以及在 GPS 暂不可用的情况下用陀螺仪等传感器恢复运动轨迹,继续进行定位。对于液晶屏,则需要研究基于手机的交互、姿态和上下文内容来智能节约电量。在网络传输方面,需要研究架构的优化,缩短用户的网络连接和使用时间,如使用本地高速缓存技术、数据压缩、交互流程简化等。

此外,针对不同的移动增强现实应用,在识别和渲染时需要采集的特征信息有所不同,为此,针对不同的业务特征对传感器进行统筹管理,通过传感器之间的协作,实现功耗最优的传感器调用方案。

虽然目前智能手机 AR 技术的发展已可以实现简单的增强现实应用,但是与 PC 工作平台相比,还面临着诸多技术挑战,如较低的图像分辨率、屏幕尺寸的局限、较低的图形处理能力、有限的存储空间、缺少浮点运算单元以及低端处理器等都为该技术的普及和实用化设置了障碍。未来短时间内,智能手机的性能还无法与 PC 抗衡。虽然可以选择 C/S 构架分担部分计算负载,但是无线网络数据传输的延时、较窄的带宽都使得该技术还无法具备完全的实时处理能力。智能手机在硬件发展上的滞后使得 PC 上运行的算法无法被有效移植,进一步拉大了手机移动 AR 技术与传统 AR 技术在发展速度上的差距。另外,智能手机对于电池的依赖也是影响此类系统发展的一个重要因素。

随着移动终端技术的不断发展和完善,制约 AR 技术移植至手机平台的种种约束必将成为过去,手机作为人们日常生活的一部分,将发挥其更大的技术优势。而将 AR 技术与移动终端相结合也必将开启 AR 领域的一个崭新研究方向。

6.4 Android 系统开发环境

作为目前国内应用最广泛的智能手机操作系统,Android 为 AR 应用提供了一个稳定便捷的开发平台,其国内市场前景广阔,潜力巨大。目前,只有国外少量知名大学的视觉实验室和一些视觉商业公司展示过一些研究成果,如高通公司的跟踪注册 SDK 和德国 Metaio 公司的移动 SDK,但是都没有公开实现细节。因此,本节将详细介绍基于 Android 操作系统的移动 AR 开发所需的软硬件环境,以及部分技术细节。

6.4.1 Android 操作系统^[49]

Android 是 Google 公司于 2007 年 11 月 5 日发布的基于 Linux 2.6 核心的开源智能手机操作系统,它包括三个主要组成部分:操作系统、用户界面和应用程序。Android 系统架构分为四层,每一层都将底层实现封装,并暴露调用接口给上一层,从高到低分别是应用(application)层、应用框架(application framework)层、系统运行库(libraries、Android runtime)层和 Linux 内核(kernel)层。本节将对每一层进行详细介绍。

应用层:应用层是 Android 操作系统的应用层,是 Android 系统自带的以 Java 语言编写的核心应用程序集合,包括电子邮件程序、SMS 短消息程序、日历、地图、浏览器、联系人管理程序等。该层是 Android 系统给用户提供的手机基本功能软件集。

应用框架层:应用框架层是编写 Google 发布的核心应用时所使用的 API 框架,它允许开发人员使用这些框架来开发应用,SDK 所提供的 API 就来自这里。隐藏在每个应用后面的是一系列服务和系统,具体如下。

(1) 丰富可扩展的视图(view),它包括列表(lists)、网格(grid)、文本框(text box)、按钮(button),以及可嵌入的 Web 浏览器。

(2) 内容提供者(content providers)使得应用程序可以访问另一个应用程序的数据(如联系人数据库),或者共享它们自己的数据。

(3) 资源管理器(resource manager)提供非代码资源的访问,如本地字符串、图

形和布局文件。

(4) 通知管理器 (notification manager) 使得应用程序可以在状态栏中显示自定义的提示信息。

(5) 活动管理器 (activity manager) 用来管理应用程序生命周期, 并提供常用的导航回退功能。

系统运行库层: 系统运行库层分为核心函数库和 Dalvik 虚拟机两部分。核心函数库提供了 Java 编程语言核心库的大多数功能, 通过 Android 应用程序框架为开发者提供服务。Android 平台的核心函数库包括以下几个。

(1) Surface Manager: 主要功能是负责管理显示与存取操作间的互动, 在“显示”与“存取操作”间的对应、互动以及实现无缝的 2D/3D 多应用程序之间的整合。

(2) OpenGL ES: 主要是基于 OpenGL ES 1.0 规范所定义的 3D 绘图引擎。该库可以将硬件的 3D 功能作最佳化处理显示, 可以使用硬件 3D 加速或者使用高度优化的 3D 软加速。

(3) SGL: 主要处理有关显示 2D 方面的绘图引擎。

(4) Media Framework: 主要是以 PacketVideo 公司的 OpenCORE 为基础发展的多媒体函数库, 支持多种声音及影像格式, 如 MPEG4、H. 264、MP3、AAC、AMR、JPG、PNG、GIF 等格式。

(5) FreeType: 主要处理有关显示位图及向量字形方面的任务。

(6) SSL (secure socket layer): 主要处理有关保护网页通信的协议。

(7) SQLite 数据库: 是一套开放源代码的关联式数据库, 其特点在于轻量性方面的设计结构, 更适合在手机上使用。

(8) WebKit: 网页浏览器引擎, 让 Android 发展出专属的浏览器, 也可以让开发者利用该引擎完成不同的应用。

(9) Libc: 是 BSD 标准系统 C 函数库, 但是针对 Linux 嵌入式装置进行过最佳化的调整。

Dalvik 虚拟机主要是为 Android 应用程序提供运行环境, 其作用相当于 JVM (Java 虚拟机)。在运行 Android 中的每一个应用程序时, 都拥有一个独立的 Dalvik 虚拟机实例。另外, 为适应手机特殊的运行环境, Dalvik 被设计成一个可以同时高效地运行多个虚拟系统的设备。

Linux 内核层: Linux 内核层是硬件与软件层之间的抽象层, 包含显示驱动、键盘驱动、摄像头驱动、WiFi 驱动等。

6.4.2 手机硬件环境和软件开发环境

基于手机平台的 AR 系统开发,若采用客户端/服务器模式负担系统运行负载,则可由服务器端负责图像处理或视觉任务中部分计算量较大的环节。因此,服务器端配置需要运算速度快、存储容量大以及支持 GPU 加速。客户端不论选择哪种操作系统,一般都需要对用户进行位置跟踪和通信。因此,手机硬件应配置电子罗盘和 GPS 模块,并可以执行部分图像处理和渲染任务。这就要求手机具有高性能的 CPU、良好的摄像功能和完善的三维图形处理能力。因此,建议采用如表 6.1 所示的配置。

手机终端:运行 Android 2.0 以上操作系统的智能手机,支持触摸屏、摄像头、3G、WiFi、GPS、电子罗盘,处理速度达到 1GHz 以上,具有较强的图形处理能力。

服务器:一般的中低端 PC 服务器,服务器需要运算速度快,存储容量大。

网络:支持 3G、WiFi。

表 6.1 参考手机硬件规格

操作系统	Android 2.3
屏幕参数	3.7 英寸 800×480 像素 AMOLED
中央处理器	高通 QSD8250 1GHz
ROM 和 RAM	均为 512MB
摄像头	500 万像素

以安装 Android 系统的 Nexus One 智能手机为例,其 GPS 模块可为用户提供待跟踪注册目标的初始定位,3-axis 加速度计和指南针可用于获取手机设备的朝向信息,较强的 2D/3D 图形处理能力可用于图像处理,内置的 OpenGL ES1.1 可作为三维图形渲染 API,三维模型的格式可选用 OBJ 格式。

在软件开发环境方面,服务器端可采用 Visual Studio 开发平台编写 VC++ 程序,以加快运算速度。为满足后台信息管理需求,可通过 XAMPP 搭建服务器环境。智能手机系统可选择 Android 2.0 以上版本。Android 如今已支持使用多种编程语言开发应用程序,开发者可以使用 Java 编写应用程序,也可以通过 NDK 使用 C/C++ 混合编程,还可使用其他脚本语言实现数据库管理和网络通信功能(如 Python、Lua、PHP 等),另有其他诸如 Qt(Qt for Android)、Mono(Mono for Android)等一些著名编程框架。除此以外,开发者甚至可以使用 C#来实现程序框架和细节。而在移动互联网方面,Android 已经支持 JavaScript、AJAX、HTML5、jQuery、Dojo、PhoneGap

等。因此,系统的软件开发环境可配置如下。

服务器端安装 Eclipse 编译环境、XAMPP 系统软件,搭载 Android 2.0 或更高版本的开发包,支持 ADT 插件。手机端支持 NDK 和 OpenCV 视觉库,以实现算法。

6.4.3 OpenCV 库编译和基于 NDK 的开发

OpenCV (open source computer vision library) 是 Intel 的开源计算机视觉库^[50],它由一系列 C 函数和少量 C++ 类构成,实现了图像处理 and 计算机视觉方面的很多通用算法,如图像数据操作、图像/视频的输入与输出、矩阵/向量数据操作及线性代数运算,支持多种动态数据结构、基本图像处理、结构分析、摄像头定标、运动分析、目标识别、基本的 GUI、图像标注等。使用 OpenCV 进行开发,不需要从零开始写视觉算法代码,极大程度地降低了用户的开发难度。目前 OpenCV 2.3.1 已经支持 Android 平台,Android-OpenCV 是 OpenCV 在 Android 手机上的移植版,这里首先介绍 Android-OpenCV 的安装和使用。

Android-OpenCV 的安装:OpenCV 是基于 C/C++ 的。因此,Android-OpenCV 必须依赖 NDK 才可以实现编译。首先应正确安装 android-ndk-r4-crystax 和 cygwin。确保在系统变量路径中包含 cygwin 的 bin 和 android-ndk-r4-crystax 的路径。运行 cygwin,通过 CD 命令进入 opencv 目录下,输入 sh build.sh 进行编译,若编译成功则生成 armeabi 和 armeabi-v7a 两个文件夹,里面都包含 libandroid-opencv.so。

Android-OpenCV 的使用:在 Android 平台开发基于 OpenCV 的应用,主要有两种方式,一种方式是直接调用 OpenCV 打包好的函数库,另一种方式是通过修改 OpenCV 库实现特定功能的函数,目前的应用还主要是以第一种为主。打开 Eclipse,导入 (import) Opencv 这个工程,该工程位于/opencv/android 目录下。Opencv 这个工程编译通过之后,就可以导出 (export) 它,选择 Java 的 Jar file,导出时去掉 obj 和 libs 这两个文件夹,去掉 AndroidManifest.xml 和 default.properties 文件,输出文件名为 Opencv.jar,程序中加入 Opencv.jar 库即可。

为了实现 Android 平台上的 C/C++ 编程,需要研究基于 NDK 的开发步骤。

- (1) 将自己的源代码放到 <project>/jni 中。
- (2) 创建 <project>/jni/Android.mk 描述本地源代码给 NDK 编译系统。
- (3) 选择 Create <project>/jni/Application.mk。
- (4) 编译本地代码,运行命令 CD 进入工程路径,再运行命令 ndk-build:

```
$ cd <project>
$ <ndk>/ndk-build
```

(5) 运行工具复制共享库到应用程序的恰当位置。

(6) 用 SDK 编译成 apk 发布。

6.4.4 部分增强现实功能实现

1. 视频数据和 GPS 信息获取

Android 平台提供视频数据获取功能,通过实时获取场景的视频图像帧,以及相应的后台图像处理函数,从而实现增强效果的实时性。Android 平台视频帧的画布(surface)显示机制通过类 Hardware.Camera 和 SurfaceHolder.Callback 接口实现。只有在 Callback 接口的回调函数中设置摄像机的开启和预览等功能并获得画布,摄像头捕获的图像才能实时正确地显示在手机屏幕上。函数方法介绍如下。

`public static Camera open ( ) throws IOException;` 获得摄像机实例。抛出 `IOException` 表示摄像头打开失败。

`public final void setPreviewDisplay ( SurfaceHolder holder );` 为视频帧预览设定画布,且必须在 `startPreview` 函数之前设定。参数 `holder` 包含画布的接口,可以通过它来控制画布的大小等。

`public static void startPreview ( );` 初始化摄像头,开启预览帧,即可以在画布代表的手机屏幕上绘制捕获后的视频帧。

`public final void setPreviewCallback ( Camera.PreviewCallback cb );` 设置摄像头预览回调接口,以获取捕获的视频帧。参数 `cb` 预览帧显示回调接口。

`public final void autoFocus ( Camera.AutoFocusCallback afc );` 设置摄像头自动对焦。参数 `afc` 表示自动对焦状态设定回调接口。

`public abstract void onPreviewFrame ( byte[ ] data, Camera camera );` 预览帧回调实现接口函数,在 `open` 函数启动摄像头以后该函数随即开启,并实时获取视频帧数据流。参数 `data` 表示由图像版式类定义的比特数组形式的视频帧数据流, `camera` 表示由 `open` 方法获取的摄像机类。

位置服务(location-based service, LBS),又称为定位服务,融合了 GPS 定位、移动通信、导航等多种技术,提供与空间位置相关的综合应用服务。Android 平台支持提供位置服务的 API,在开发过程中主要用到 `LocationManager` 和 `Location Providers` 的对象。

`LocationManager` 可以用来获取当前位置、追踪设备的移动路线,或设定敏感区域,在进入或离开敏感区域时设备会发生特定警报。`LocationProviders` 是能够提供

定位功能的组件集合,集合中的每种组件以不同的技术提供设备的当前位置,区别在于定位的精度、速度等方面。提供位置服务,首先需要获得 LocationManager 对象,获取 LocationManager 可以通过调用 android.app.Activity.getSystemService 函数实现。在获取到 LocationManager 后,还需要指定 LocationManager 的定位方法,然后才能够通过回调函数 onLocationChanged(Location loc) 函数获取当前位置。LocationManager 支持的定位方法有两种,即 GPS 定位和网络定位。GPS 定位可以提供更加精确的位置信息,但定位速度和质量受到卫星数量和环境情况的影响。网络定位提供的位置信息精度差,但速度较 GPS 定位快。

回调函数获取的 Location 对象中,包含了可以确定位置的信息,如经度、纬度和速度等,通过调用 Location 中的 getLatitude 和 getLongitude 方法可以分别获取位置信息中的纬度和经度。LocationManager 提供了一种便捷、高效的位置监视方法 requestLocationUpdates,可以根据位置的距离变化和时间间隔设定产生位置改变事件的条件,这样可以避免因微小的距离变化而产生大量的位置改变事件。

2. 方位传感器信息获取

加速度计(accelerometer)可以用来获取手机的姿态信息,主要包括横滚角和俯仰角。从设备中读取的数值包括 Gx、Gy、Gz 三个,分别对应手机当前加速度在其体坐标系上的分量。其空间坐标示意图如图 6.20 所示。加速度分为静态加速度和动态加速度。当静止时,默认手机加速度与重力加速度大小相同,方向相反。将加速度传感器倾斜一定角度,重力场会在其感应器上产生加速度分量,通过该分量就可以测量出手机倾斜了多少度,由此实现一些前后左右的控制。例如,在赛车、滑雪等游戏中,加速度传感器甚至可以取代方向键,通过前后左右的角度变换实现转弯、加速、刹车等动作,使游戏更具娱乐性。另外一种就是所谓的动态加速度,可以侦测速度、撞击等。在开发过程中,当手机做非匀速运动时,需要滤去重力加速度产生的影响,以求得手机真正的运动加速度。

电子罗盘(compass)或方位感应器由磁传感器配合加速度计实现,可以用来获取手机设备的朝向,即航向角,以及更精确稳定的横滚角和俯仰角。通过综合 3-axis 加速度计和磁传感器信息,以及求取摄像机内部参数等,可完成简单的辅助性运动跟踪。通常情况下,方向传感器多用作指南针,这也是其称为电子罗盘的原因。在 Android 系统中,往往通过直接调用系统函数的方法获得方向传感器数据,以求取其姿态角,涉及的函数包括 getRotationMatrix、onSensorChanged 等。

三轴陀螺仪(gyroscope)在手机中是一个结构非常精密的芯片,内部嵌有一个

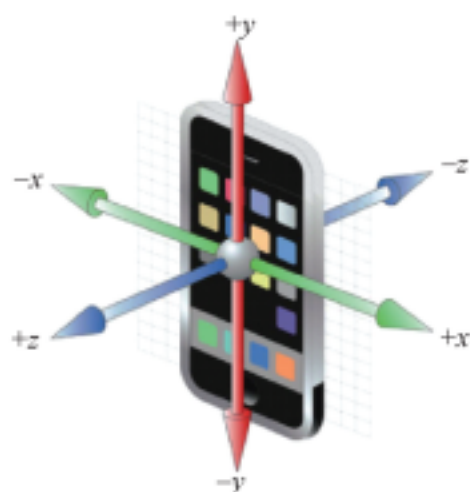


图 6.20 智能手机空间坐标示意图

在与地面垂直方向上转动的十分微小的陀螺。通过记录设备与该陀螺的夹角,可以测量设备自身的旋转运动。陀螺仪对设备旋转角度的检测是瞬时且精准的,能满足一些需要高分辨率和快速反应的场合,如游戏的瞄准等。因此,陀螺仪的优势在于测量设备自身的旋转运动状态,但其不足是无法确定设备的方位。而磁传感器的融入,可以在获得设备运动状态的同时,再获得其朝向信息,而不至于迷失方向。陀螺仪的一个经典应用就是在没有卫星和网络的情况下,配合加速度计进行导航。Android 系统中涉及的参数主要是 TYPE_GYROSCOPE。需要注意的是,绝大多数版本的手机都含有加速度计和磁传感器,但是陀螺仪却因手机型号而异。

3. 图像的特征提取

由于许多计算机视觉算法使用特征提取作为其初级计算步骤,所以研究人员研究了大量的特征提取算法。不同的特征提取算法可在图像上提取不同的图像特征,且它们的计算复杂性和可重复性也差别较大。

OpenCV 中实现了多种角点特征的提取方法,如 Harris 角点、ShiTomasi 角点、SIFT 兴趣点、SURF 兴趣点等,下面介绍如何利用 OpenCV 库中的算法实现这些特征点的检测。在此之前先介绍与特征检测密切相关的一些变换,如 Sobel 算子、拉普拉斯算子、Canny 算子、霍夫变换。之后介绍 OpenCV 中常用的自然特征 SIFT、SURF、FAST、BRIEF 的提取以及与之相应的描述符计算。最后介绍增强现实系统中基于自然特征的跟踪定位算法的一般流程。

1) Sobel 算子

Sobel 算子用多项式计算来拟合导数计算,可以用 Android-OpenCV 中的如下

方法实现。

`void org.opencv.imgproc.Imgproc.Sobel ( Mat src, Mat dst, int ddepth, int dx, int dy, int ksize, double scale, double delta, int borderType )`: `src`-输入图像, `dst`-输出图像, 图像大小和通道数与输入图像一致; `ddepth`-输出图像的深度; `dx`- x 方向求导的阶数, `dy`- y 方向求导的阶数; `ksize`-Sobel 内核的大小; `scale`-计算导数值的尺度因素, 一般使用默认值; `delta`-使用默认值; `borderType`-使用默认值。需要注意的是, `dx` 和 `dy` 中必须有且只能有一个为非零值, 即只能计算 x 方向或者 y 方向的导数。

2) 拉普拉斯算子

拉普拉斯算子可以用于边缘检测, 可以用 Android-OpenCV 中的如下方法实现。

`void org.opencv.imgproc.Imgproc.Laplacian ( Mat src, Mat dst, int ddepth, int ksize, double scale, double delta, int borderType )`: `src`-输入图像, `dst`-输出图像, 图像大小和通道数与输入图像一致; `ddepth`-输出图像的深度; `ksize`-内部调用的 Sobel 算子的内核大小; `scale`-计算拉普拉斯导数值的尺度因素, 使用默认值; `delta`-使用默认值; `borderType`-使用默认值。

3) Canny 算子

Canny 算子也可以用于边缘检测, 可以用 Android-OpenCV 中的如下方法实现。

`void org.opencv.imgproc.Imgproc.Canny ( Mat image, Mat edges, double threshold1, double threshold2, int apertureSize, boolean L2gradient )`: `image`-8 位单通道输入图像; `edges`-输出图像, 其大小和类型与输入图像一样; `threshold1`-低阈值, `threshold2`-高阈值; `apertureSize`-Sobel 算子的大小, 值得注意的是 `apertureSize` 只能取 3、5 或者 7; `L2gradient`-一个标识, 用于标识是否采用更精确的方法计算图像梯度幅值。

4) 霍夫变换

霍夫变换是一种在图像中寻找直线、圆及其他简单形状的方法, 在 OpenCV 中实现了霍夫圆变换和霍夫线变换。

`void org.opencv.imgproc.Imgproc.HoughCircles ( Mat image, Mat circles, int method, double dp, double minDist, double param1, double param2, int minRadius, int maxRadius )`: `image`-输入 8 位单通道灰度图像; `circles`-使用一个集合来表示检测到的圆, 每一个集合 $(x, y, radius)$ 包含 3 个浮点型元素; `method`-使用的检测方法, 现在 OpenCV 中只有霍夫梯度法; `dp`-累加器图像的分辨率, `dp` 的值不能比 1 小; `minDist`-检测到圆心之间的最小距离; `param1`-Canny 边缘函数的高阈值; `param2`-圆心检测阈值; `minRadius`-能检测到的最小圆半径, 默认为 0; `maxRadius`-能检测到的

最大圆半径,默认为0。

`void org.opencv.imgproc.Imgproc.HoughLines( Mat image, Mat lines, double rho, double theta, int threshold, double srn, double stn );` 标准霍夫线变换。`image`-边缘检测的输出图像;`lines`-存储着检测到的直线参数对的容器;`rho`-参数极径,以像素值为单位;`theta`-参数极角,以弧度为单位,一般使用 1° ($CV_PI/180$);`threshold`-检测一条直线所需的最少曲线交点;`srn` 和 `stn` 参数默认为0。

`void org.opencv.imgproc.Imgproc.HoughLinesP( Mat image, Mat lines, double rho, double theta, int threshold, double minLineLength, double maxLineGap );` 统计概率霍夫线变换。`image`-边缘检测的输出图像;`lines`-储存检测到的直线参数对的容器;`rho`-参数极径 r ,以像素值为单位,一般使用1像素;`theta`-参数极角,以弧度为单位,一般使用 1° ($CV_PI/180$);`threshold`-要检测一条直线所需的最少曲线交点;`minLineLength`-能组成一条直线最少的点数量,点数量不足的直线将被抛弃;`maxLineGap`-点成线允许的最大点距。

5) Harris 角点

角点是两个边缘的连接点,它代表了两个边缘变化方向上的点,这种变化可以用来帮助检测角点。可以用 Android-OpenCV 中的如下方法实现。

`void org.opencv.imgproc.Imgproc.cornerHarris( Mat src, Mat dst, int blockSize, int ksize, double k, int borderType );` `src`-输入单通道8位或浮点型图像;`dst`-存储角点的输出图像,类型为 `CV_32FC1`,大小和原图像相同;`blockSize`-邻域大小;`ksize`-Sobel算子的核的大小;`k`-Harris检测器的权重;`borderType`-像素外插方法。

要计算 Harris 角点列表,需要使用 `goodFeatureToTrack` 函数,并传递适当的参数。

6) ShiTomasi 角点

寻找更精确的角点位置(不是整数类型的位置,而是更精确的浮点类型位置)。在默认情况下,`goodFeatureToTrack` 函数用于计算 ShiTomasi 角点;不过如果将参数 `use_harris` 设置为非0值,那么它会计算 Harris 角点。可以用 Android-OpenCV 中的如下方法实现。

`void org.opencv.imgproc.Imgproc.goodFeaturesToTrack( Mat image, MatOfPoint corners, int maxCorners, double qualityLevel, double minDistance, Mat mask, int blockSize, boolean useHarrisDetector, double k );` `image`-输入8位或32位浮点型单通道图像;`corners`-输出检测到的角点容器;`maxCorners`-返回最大的角点数目;`qualityLevel`-图像角点的质量水平;`minDistance`-两个角点间的最小欧氏距离;`mask`-

选择的兴趣区域;blockSize-邻域的大小;useHarrisDetector-是否使用 Harris 角点;k-Harris 角点检测器的权重。

7) SIFT 特征

使用传统 SIFT 特征提取算法,寻找图像尺度空间中的极值点,依据特征点的位置、尺度、主方向构造 128 维特征向量作为特征点的描述符。SIFT 特征对旋转、尺度缩放、亮度变化保持不变性,信息量丰富,适用于在海量特征数据库中进行快速、准确的匹配。

SIFT 特征提取算法需要构造图像的尺度空间金字塔,并且为防止图像反走样,需要将图像扩大一倍再进行处理,所以 SIFT 特征提取算法用时比较长,对建筑物表面相似特征的区分度比较低。

8) SURF 特征

SURF 特征提取算法采用 Hessian 矩阵获取图像局部最值定位特征点,并采用了积分图像、Box 滤波等技术,在时间上优于 SIFT 算法。

9) FAST 特征

FAST 特征检测算法来源于 corner 的定义,这个定义基于特征点周围的图像灰度值,检测候选特征点周围的一圈像素值,如果候选点周围邻域内有足够多的像素点与该候选点的灰度值差别足够大,则认为该候选点为一个特征点。

为了获得更快的结果,FAST 还采用了额外的加速办法。如果测试了候选点周围每隔 90°角的 4 个点,应该至少有 3 个和候选点的灰度值差足够大,否则不用再计算其他点,直接认为该候选点不是特征点。为了提高比较的效率,通常只使用 N 个周边像素来比较,也就是 FAST- N ,使用比较多的是 FAST-9 和 FAST-12。

10) BRIEF 特征

BRIEF(binary robust independent element feature)利用局部图像邻域内随机点对的灰度大小关系来建立局部图像特征描述子,得到的二值特征描述子不仅匹配速度快,而且存储要求内存低,适合在移动终端平台使用。

除了 BRIEF,近几年研究人员还提出了许多二值特征描述子,如 ORB、BRISK、FREAK。上述这些特征描述子都是基于手机平台设计的,也有一些研究试图利用机器学习的方法,通过数据驱动得到想要的特征描述子,这类特征描述子包括 PCA-SIFT、Linear Discriminative Embedding、LDA-Hash 等。当然,除了提到的这些特征描述子之外,还有许多其他的特征描述子。

4. OpenCV 中特征点的提取与匹配

在 Android-OpenCV 中,提取图像上的特征点的方法如下:先建立 FeatureDetector

的对象 `detector`,再使用方法 `FeatureDetector org.opencv.features2d.FeatureDetector.create(int detectorType)` 建立特征检测器,其中,`detectorType`支持 FAST、BRIEF、SIFT、SURF、ORB、BRISK 等特征点的检测。然后使用方法 `void org.opencv.features2d.FeatureDetector.detect(Mat image,MatOfKeyPoint keypoints)` 从图像 `image` 上提取特征点放到容器 `keypoints` 中。

在 Android-OpenCV 中,计算特征点描述符的方法如下:先建立 `DescriptorExtractor` 的对象,然后用 `DescriptorExtractor org.opencv.features2d.DescriptorExtractor.create(int extractorType)` 方法建立描述符提取器,其中,`extractorType` 为对应特征点的描述符提取类型。最后使用方法 `void org.opencv.features2d.DescriptorExtractor.compute(Mat image,MatOfKeyPoint keypoints,Mat descriptors)` 计算图像 `image` 上特征点 `keypoints` 的描述符 `descriptors`。

在 Android-OpenCV 中描述符的匹配方法如下:首先建立 `DescriptorMatcher` 的对象,再使用方法 `DescriptorMatcher org.opencv.features2d.DescriptorMatcher.create(int matcherType)` 创建描述符匹配,其中,`matcherType` 为对应特征点的描述符匹配类型。最后使用方法 `void org.opencv.features2d.DescriptorMatcher.match(Mat queryDescriptors,Mat trainDescriptors,MatOfDMatch matches)` 对训练图像和查询图像的描述符进行匹配。

5. OpenGL ES 和三维模型数据格式

OpenGL ES (OpenGL for embedded systems) 是 OpenGL 三维图形 API 的子集,针对手机、PDA 和游戏主机等嵌入式设备而设计。OpenGL ES 是从 OpenGL 裁剪定制而来的,它去除了 `glBegin/glEnd`、四边形 (`GL_QUADS`)、多边形 (`GL_POLYGONS`) 等许多非绝对必要的特性。经过多年发展,现在主要有两个版本:OpenGL ES 1.x 针对固定管线硬件,而 OpenGL ES 2.x 则针对可编程管线硬件。OpenGL ES 1.0 以 OpenGL 1.3 规范为基础,OpenGL ES 1.1 以 OpenGL 1.5 规范为基础的,OpenGL ES 2.0 则参照 OpenGL 2.0 规范定义,它们分为 `common profile` 和 `common lite profile` 两种模式,`common lite profile` 模式只支持定点实数,而 `common profile` 既支持定点数又支持浮点数。

OBJ 文件是 Alias | Wavefront 公司为其基于工作站的 3D 建模和动画软件 `Advanced Visualizer` 开发的一种标准 3D 模型文件格式,适用于 3D 软件模型之间的互导,也可以通过 Maya 读写。例如,在 3DS Max 或 LightWave 中建了一个模型,想把它调到 Maya 里面渲染或动画,导出 OBJ 文件是一个很好的选择。目前几乎所有

知名的3D软件都支持OBJ文件读写,但其中大多数需要通过插件实现。OBJ文件是一种文本文件,可以直接用写字板打开查看并编辑修改。图6.21所示为OBJ格式文件中的内容。

#3ds Max Wavefront OBJ Exporter	g Box01
v0.94b-(c)2007 guruware	f 1 2 3
# object Box01	f 3 4 1
	f 5 6 7
v -5.00 0.00 5.00	f 7 8 5
v -5.00 0.00 -5.00	f 1 4 6
v 5.00 0.00 -5.00	f 6 5 1
v 5.00 0.00 5.00	f 4 3 7
v -5.00 10.00 5.00	f 7 6 4
v 5.00 10.00 5.00	f 3 2 8
v 5.00 10.00 -5.00	f 8 7 3
v -5.00 10.00 -5.00	f 2 1 5
#8 vertices	f 5 8 2
	#12 faces

图 6.21 OBJ 格式文件示例

文件内容表示的是一个立方体,注释行以符号“#”为开头,“v”表示几何体顶点(geometric vertices),“g”表示组名称(group name),“f”表示面(face),这些文件中有8个顶点、12个面(3个顶点构成一个三角形面),显示的结果如图6.22所示。由此可见,无论多复杂的三维模型都是由最简单的三角面片组成的。

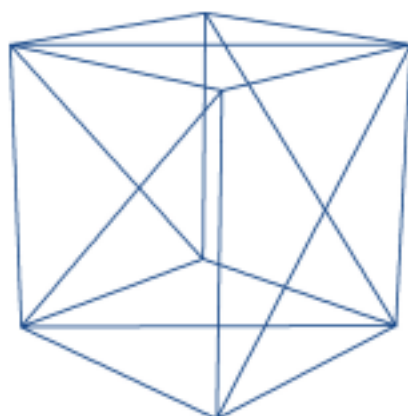


图 6.22 由 OBJ 文件表示的简单图形

6. AR 系统跟踪注册的一般流程

首先进行系统预处理。

(1) 建立好虚拟三维模型。

(2) 预先拍摄关键帧图像,提取特征点,计算描述符,并存入数据库。

然后根据匹配算法生成虚实结合图像。

(1) 系统工作时,在摄像机获取的当前帧图像上提取特征点,计算描述符,对关键帧和当前帧图像特征点进行匹配,得到匹配点对;利用摄像机模型参数估计当前帧摄像机的姿态信息,应用 RANSAC 鲁棒方法去除错误匹配点对。

(2) 使用筛选后的匹配点对和摄像机模型参数,根据 2D/3D 对应关系,再次计算出摄像机的姿态信息并进行更新。

(3) 根据摄像机的姿态信息,将虚拟三维模型叠加到当前帧上。

最后将虚实结合图像输出到显示设备(如智能终端手机等),以实现虚拟与真实场景的融合显示,如图 6.23 所示。

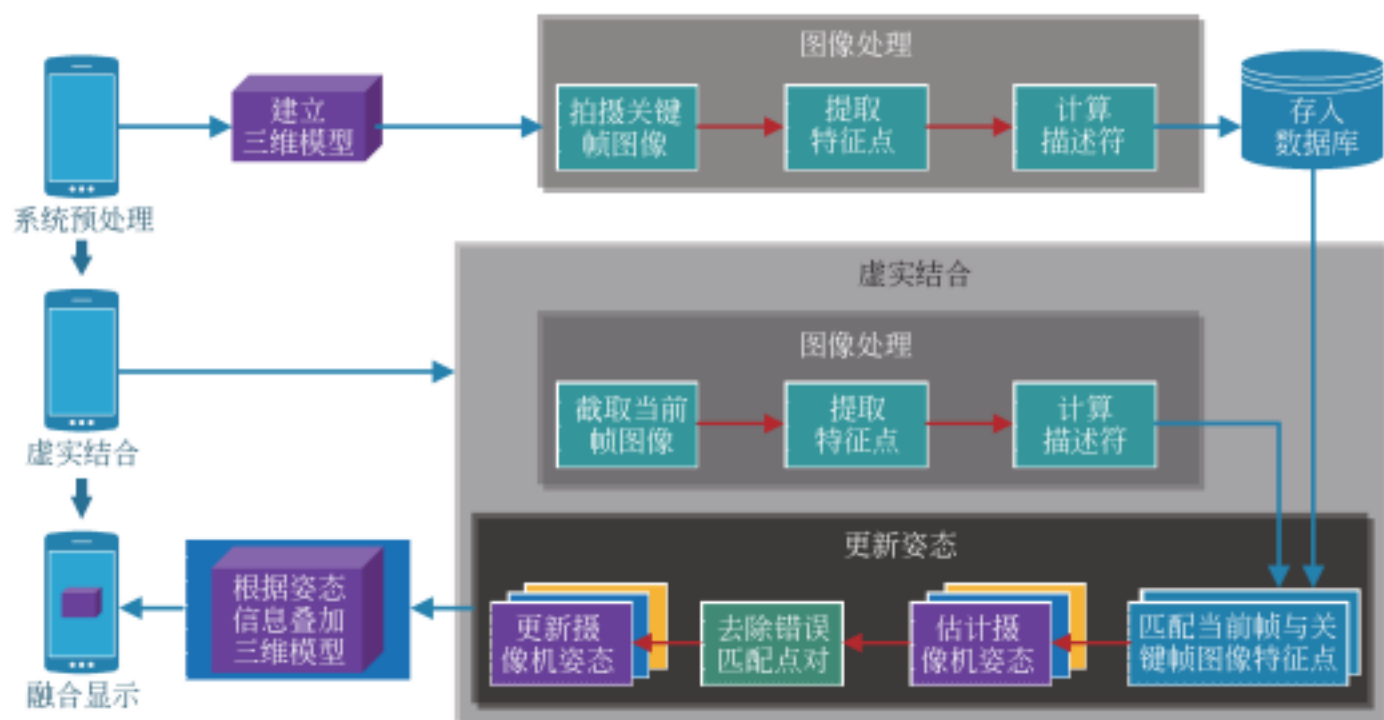


图 6.23 AR 系统跟踪注册一般流程

参考文献

- [1]Feiner S, MacIntyre B, Hollerer T, et al. A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment//Proceedings of International Symposium on

- Wearable Computers, Cambridge, 1997.
- [2] Julier S, Baillot Y, Lanzagorta M, et al. BARS: Battlefield augmented reality system//NATO Symposium on Information Processing Techniques for Military Systems, Istanbul, 2000.
- [3] Thomas B, Close B, Donoghue J, et al. ARQuake: An outdoor/indoor augmented reality first person application//Proceedings of 4th International Symposium on Wearable Computers, Atlanta, 2000.
- [4] Newman J, Ingram D, Hopper A. Augmented reality in a wide area sentient environment//Proceedings of the 2nd IEEE and ACM International Symposium on Augmented Reality, New York, 2001.
- [5] Wagner D, Schmalstieg D. ARToolkit on the pocket PC platform//Proceedings of the Second IEEE Intl. Augmented Reality Toolkit Workshop, Tokyo, 2003.
- [6] Wagner D, Pintaric T, Ledermann F, et al. Towards massively multi-user augmented reality on handheld devices//Proceedings of the Third International Conference on Pervasive Computing, Munich, 2005.
- [7] Optical Tracking and Video See-Through AR on Consumer Cell-Phones. http://www.jku.at/cg/content/e60566/e155460/e156630/CAR_VRAR_Workshop.pdf [2012-5-15].
- [8] Henrysson A, Billinghurst M, Ollila M. Face to face collaborative AR on mobile phones//Proceedings of the Fourth IEEE and ACM international Symposium on Mixed and Augmented Reality, Vienna, 2005.
- [9] Kähäri M, David J M. MARA-sensor based augmented reality system for mobile imaging device//Proceedings of the Fifth IEEE and ACM International Symposium on Mixed and Augmented Reality, Santa Barbara, 2006.
- [10] Gerhard R, Drummond T. Going out: Robust model-based tracking for outdoor augmented reality//Proceedings of the Fifth IEEE and ACM International Symposium on Mixed and Augmented Reality, Santa Barbara, 2006.
- [11] Brans E, Brombach B, Zeidler T, et al. Enabling mobile phones to support large-scale museum guidance. IEEE Multimedia, 2007, 14(2): 16 ~ 25.
- [12] Wagner D, Reitmayr G, Mulloni A, et al. Pose tracking from natural features on mobile phones//Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, Cambridge, 2008.
- [13] Klein G, Murray D. Parallel tracking and mapping on a camera phone//Proceedings of International Symposium on Mixed and Augmented Reality, Orlando, 2009.
- [14] Lee W, Park Y, Woo W, et al. Point-and-shoot for ubiquitous tagging on mobile phones//Proceedings of IEEE International Symposium on Mixed and Augmented Reality, Seoul, 2010.
- [15] Chen D, Girod B. Memory-efficient image databases for mobile visual search. IEEE MultiMedia Magazine, 2014, 21(1): 14 ~ 23.

- [16] Chen D, Tsai S, Chandrasekhar V, et al. Residual enhanced visual vector as a compact signature for mobile visual search. *Signal Processing*, 2013, 93(8): 2316 ~ 2327.
- [17] Laya 浏览器. <http://www.layar.com> [2012-5-12].
- [18] Wikitude 浏览器. <http://www.wikitude.com> [2012-5-15].
- [19] ARSDK. <http://developer.qualcomm.com/mobile-development/mobile-technologies/augmented-reality> [2012-6-2].
- [20] Metiao mobileSDK. <http://www.metaio.com/software/mobile-sdk> [2012-5-20].
- [21] AR 颠球游戏 ARSoccer. <https://itunes.apple.com/cn/app/arsoccer-augmented-reality/id381035151?mt=8> [2012-07-15].
- [22] TryLive Eyewear. <http://www.t-immersion.com> [2012-08-14].
- [23] Nokia Point & Find. http://pointandfind.nokia.com/main_publisher [2012-8-19].
- [24] Google Goggles. <http://www.google.com/mobile/goggles/#text> [2012-5-22].
- [25] SnapTell- Visual Product Search. <http://www.snaptell.com> [2012-8-25].
- [26] Kooaba: Make Smart Visuals-Image Recognition Products for Mobile and Desktop. <http://www.kooaba.com> [2012-9-12].
- [27] Google Glass. <http://www.google.com/glass/start> [2012-9-18].
- [28] Microsoft Files Patent for Augmented Reality Smart Glasses. <http://www.bbc.com/news/technology-20462840> [2012-10-6].
- [29] Assad M, Carmichael D J, Cutting D, et al. AR phone: Accessible augmented reality in the intelligent environment//*Proceedings of OZCHI, Brisbane*, 2003.
- [30] Hakkarainen M, Woodward C, Billinghurst M. Augmented assembly using mobile phone//*Proceedings of the 7th IEEE International Symposium on Mixed and Augmented Reality, Cambridge*, 2008.
- [31] Harlan H, Gaetano B. Information overlay for camera phones in indoor environments. *Lecture Notes in Computer Science*, 2007, 4718: 68 ~ 84.
- [32] Mark O, Anders H. UMAR: Ubiquitous mobile augmented reality//*Proceedings of the 3rd International Conference on Mobile and Ubiquitous Multimedia, Maryland*, 2004.
- [33] Rosten E, Drummond T. Machine learning for high-speed corner detection//*Proceedings of European Conference on Computer Vision, Graz*, 2006.
- [34] Bay H, Tuytelaars T, Gool L V. SURF: Speeded up robust features//*Proceedings of European Conference on Computer Vision, Graz*, 2006.
- [35] Chen W C, Xiong Y G, Gao J, et al. Efficient extraction of robust image features on mobile devices//*Proceedings of IEEE/ACM International Symposium on Mixed and Augmented Reality, Nara*, 2007.
- [36] Rashid O, Bamford W, Coulton P, et al. PAC-LAN: Mixed reality gaming with RFID enabled

- mobile phones. Computers in Entertainment, 2006, 4 (4) : 1 ~ 17.
- [37] Bruns E, Brombach B, Zeidler T, et al. Enabling mobile phones to support large-scale museum guidance. IEEE MultiMedia, 2007, 14(2) : 16 ~ 25.
- [38] Olwal A. LightSense: Enabling spatially aware handheld interaction devices//Proceedings of IEEE/ACM International Symposium on Mixed and Augmented Reality, Santa Barbara, 2006.
- [39] An Introductory Tutorial on kd-Trees. http://www.ri.cmu.edu/pub_files/pub1/moore_andrew_1991_1/moore_andrew_1991_1.pdf [2012-11-23].
- [40] Liu T, Moore A, Gray A, et al. An investigation of practical approximate nearest neighbor algorithm//Proceedings of Advances in Neural Information Processing Systems (NIPS), Vancouver, 2004.
- [41] Nister D, Stewenius H. Scalable recognition with a vocabulary tree//Proceedings of Computer Vision and Pattern Recognition, New York, 2006.
- [42] Amit Y, Geman D. Shape quantization and recognition with randomized trees. Neural Computation, 1997, 9(7) : 1545 ~ 1588.
- [43] Fritz G, Christin S, Paletta L, et al. Mobile vision for tourist information systems in urban environments//Proceedings International Conference on Mobile Learning, Rome, 2004.
- [44] Gerald F, Christin S, Lucas P. A mobile vision system for urban detection with informative local descriptors//Proceedings of the Fourth IEEE International Conference on Computer Vision Systems, 2006.
- [45] Hervé J, Matthijs D, Cordelia S. Improving bag-of-features for large scale image//International Journal of Computer Vision, 2010, 87(3) : 316 ~ 336.
- [46] Hervé J, Florent P, Matthijs D, et al. Aggregating local images descriptors into compact codes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 43 (9) : 1704 ~ 1716.
- [47] Stanford 移动视觉检索. <http://www.stanford.edu/~dmchen/mvs.html> [2012-10-24].
- [48] He J F, Feng J Y, Liu X L, et al. Mobile product search with bag of hash bits and boundary rerankings//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Providence, 2012.
- [49] Android 操作系统. [Http://developer.android.com/index.html](http://developer.android.com/index.html) [2013-3-15].
- [50] OpenCV 开源计算机视觉库. <http://opencv.org> [2013-5-10].

第7章 增强现实技术的应用实例

——轻量级飞行模拟器

本章主要介绍基于增强现实技术的轻量级飞行模拟器系统,该系统利用增强现实技术创建了一个虚实融合的座舱视景,实现了人与座舱的自然交互。系统使用三个固定的摄像机阵列,分别摄取座舱底部的三个交互区域的图像,利用四点纹理映射技术将真实图像的兴趣区域逐帧添加到虚拟座舱的多边形模型上。在空军某飞行学院对系统进行了测试,初步测试结果表明,该系统采用的虚拟融合方法提高了飞行员与座舱环境的交互性,增强了对座舱环境的感知理解,在进行编队等多机模拟训练时具有一定的优势。

7.1 增强现实轻量级飞行模拟器

7.1.1 系统功能

该轻量级飞行模拟器系统主要针对飞行教员的需要,其实现的主要功能如下。

- (1) 设定训练科目,包括起飞、降落、编队训练等。
- (2) 设置环境参数,如夜航、雾天等特定气候条件。
- (3) 飞行参数数据实时存储与显示。
- (4) 飞行数据回放。
- (5) 定止和定位操作。
- (6) 飞行航迹显示。

针对飞行学员的需求,其实现的主要功能如下。

- (1) 通过 HMD 以自然的方式观察位于不同视角方向的飞行训练环境。
- (2) 通过 HMD 能够观看到与真实座舱的油门、方向操纵杆等设备的交互动作。
- (3) 根据虚拟仪表上的指针和数字信息控制飞机飞行,可以进行起飞、降落、特技、编队等训练。
- (4) 监视和观察座舱端飞行学员正在模拟驾驶的飞机的三维姿态,观察飞行

学员训练时的各种特技动作,在编队飞行训练时观察和比较长机和僚机的飞行姿态。

7.1.2 系统结构

基于增强现实技术的轻量级飞行模拟器系统主要由主控制台教员子系统、学员座舱子系统、座舱监视子系统和一个编队服务器构成。系统可以根据需求在客户端增加数台相同的座舱子系统和座舱监视子系统。图 7.1 为轻量级飞行模拟器结构示意图,其中客户端座舱子系统主要由简易座舱、高性能计算机、HMD、头部跟踪器以及多固定摄像机采集系统构成,是飞行学员模拟训练的操作平台。监视端系统主要用于从不同的观察角度监视和观察飞行学员正在模拟驾驶的飞机的三维姿态。主控制台主要用于记录飞行学员模拟飞行训练的飞行参数数据,供教员对学员的每一次飞行训练进行管理,同时也可对其对应的座舱端发送控制命令。编队服务器主要用于飞行教学中的编队科目等训练,发送控制命令,记录和分析数据等。系统通过一个局域网将各个子系统连接在一起,采用 TCP/IP 在主控制台、座舱端、座舱监视端以及编队服务器之间进行数据传输。

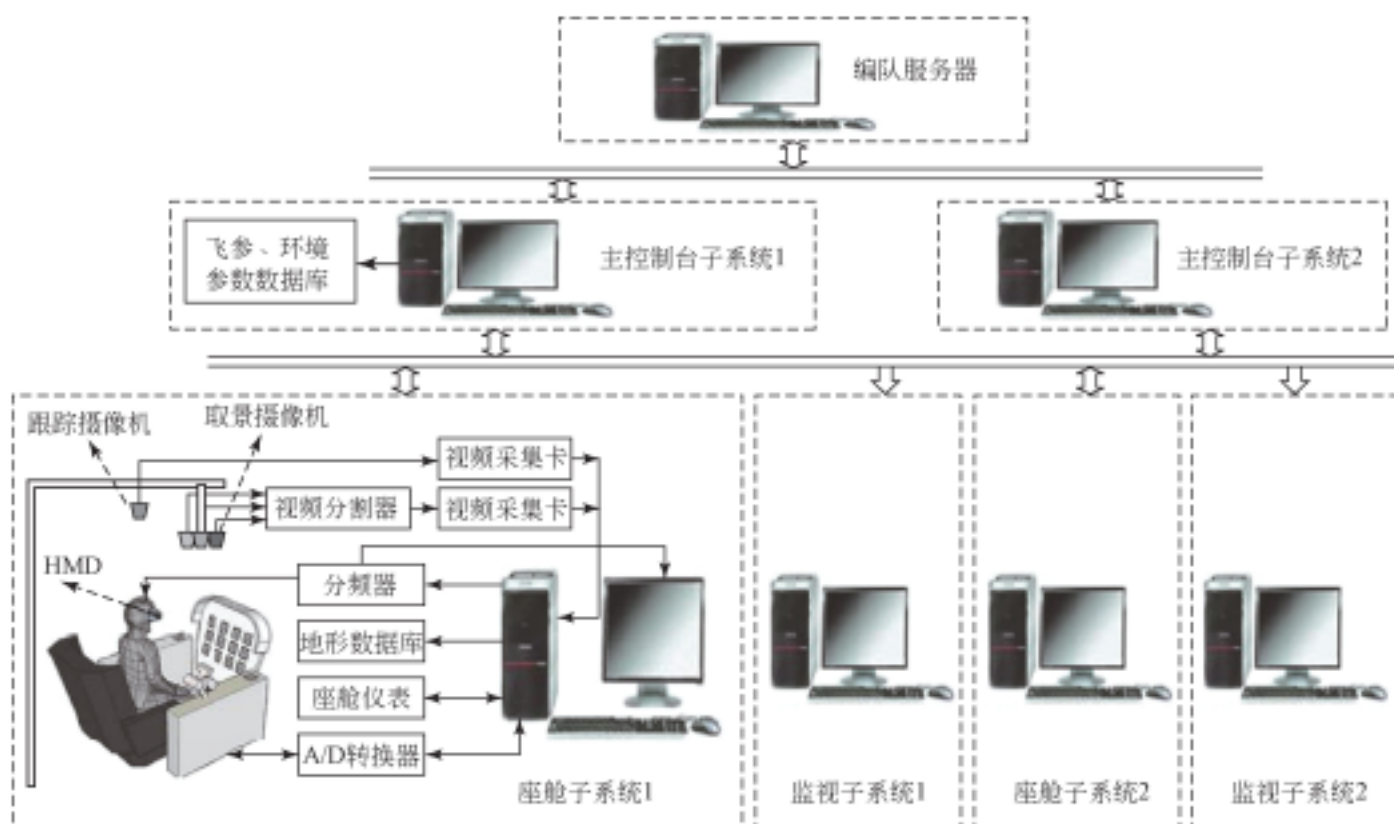


图 7.1 轻量级飞行模拟器硬件系统结构示意图

1) 简易座舱

飞行座舱是一个造价低廉的简易座舱,按照某型飞机实际座舱同比例制造。如图 7.2(a)所示,该简易座舱未使用六自由度运动系统和复杂的仪表设备,仅保留油门操纵杆、方向舵操纵杆、脚踏板、刹车等操作设备,分别通过机械传动装置与座舱底端的传感器连接。各传感器与客户端计算机之间通过一个带 USB 2.0 接口的单片机连接,计算机通过单片机读取各个操作工具传感器的模拟量数据或者数字量数据,作为飞行动力学模型计算的输入量,进行飞行动力学计算,获得模拟飞行时的飞参数据。

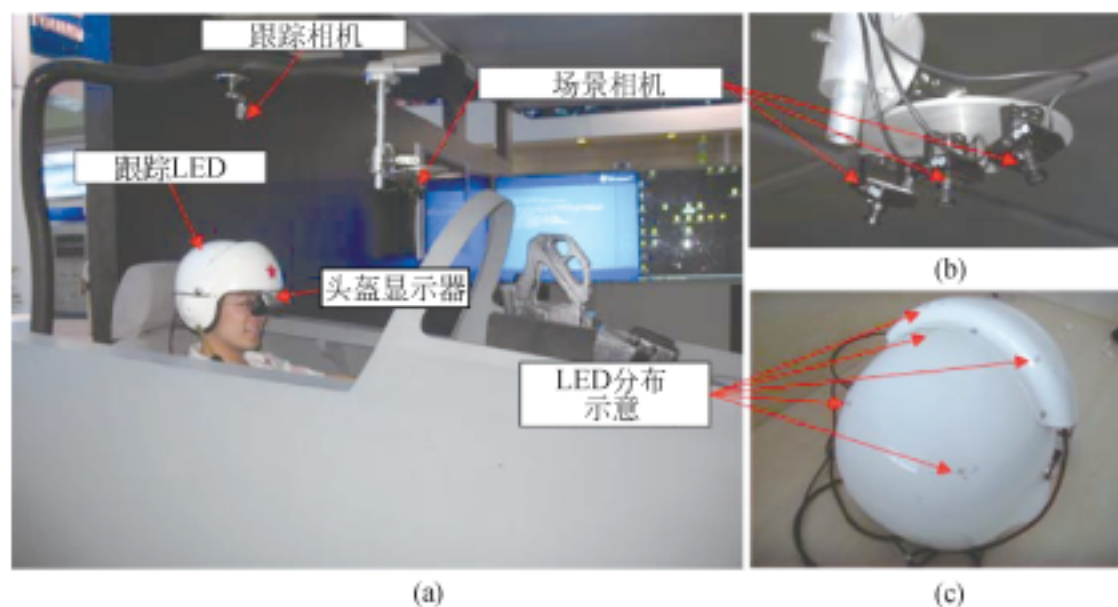


图 7.2 轻量级飞行模拟器座舱系统实物图

2) HMD 头部跟踪器

头部位置姿态跟踪器主要由 HMD、MicroView V200(DS) 视频采集卡、固定在飞行头盔上的 5 个红外灯以及固定在支架上的摄像机构成。所使用的 HMD 型号为 3Dvisor 公司的 Z800,其分辨率为 800×600 像素,视场角为 60° ,重量仅 0.5kg。为跟踪头部运动,如图 7.2(c)所示,在飞行头盔顶上安装了一个位于同一平面上的 5 个红外灯,呈正方形排列。4 个主动跟踪用的红外灯位于正方形的顶点,另外一个辅助计算用的红外灯位于正方形的一条对角线上,并靠近其中一个主红外灯。跟踪摄像机带有广角摄像头,使得飞行学员身体前倾、后仰或者左右侧身时,头上的 5 个红外灯光点均位于摄像机的视场内。摄像机的镜头内加装了一个波长范围为 $700 \sim 800\text{nm}$ 的滤光片,使摄像机仅获取红外灯图像。带广角镜头的摄像机在使用前要进行非线性标定,以获取摄像机内部参数。设定世界坐标系的原点在四个主标识点的中心,则可采用经典的平面单应性算法进行头部姿态的估计,也就

是计算摄像机的外参数,即世界坐标系原点相对于摄像机坐标系的位置和方向。计算得到的外部参数数据经过滤波处理后,用于控制飞行观察者的位置和方向。采用该方法可以跟踪飞行学员头部转动的范围为水平约 $\pm 135^\circ$,俯仰约 $\pm 45^\circ$ 。

3) 多固定摄像机场景采集系统

多固定摄像机(multiple fixed cameras, MFC)场景采集系统由三台内部参数相同性能一致的摄像机、一个四通道视频分割器和一个 MicroView V200(DS)采集卡组成。如图 7.2(b)所示,这三个摄像机并行排列,固定在与支架连接的调节装置上。三台摄像机通过视频线与四通道视频分割器的三个输入通道连接。四通道视频分割器将三台摄像机的三幅图像合成为一个通道的图像,从而可以在一幅图像内对各个通道输入的图像进行操作。

7.1.3 座舱视景的虚实融合原理

飞行座舱中可实现人-座舱交互的主要真实场景区域,是用户正下方的操纵杆操作区域、用户左侧的油门操作台面区域以及座舱右侧台面区域。图 7.3(a)为座舱的简化几何结构图。座舱油门操作台面区域和座舱右侧台面区域分别在座舱的左右两个台面上。操纵杆操作区域平面比左右两个台面低,在其周围还有左右两个内侧面以及与仪表显示区比邻的前内侧面,这三个内侧面与操纵杆区合称为座舱底部区域。另外,在创建虚拟座舱模型时将这三个区域简化为三个比邻的四边形面。如图 7.3(b)所示,左侧四边形用来显示座舱左侧用于油门操作台面区域的真实视频图像,右侧四边形用来显示座舱右侧台面区域的真实视频图像,中间的四边形用来显示座舱底部区域的真实视频图像。

图 7.4 是采用多固定摄像机的座舱视景虚实融合的流程。座舱底部区域、油门操作台面区域、座舱右侧台面区域的一个重要特征就是这三个区域由直线区别开来,该直线在分别用多个摄像机成像后依然是一条直线。因此,在虚拟座舱底部的三个四边形面上添加真实座舱三个区域的图像时,需要确保真实座舱三个交互区域的比邻边界(直线)的像(直线)始终与虚拟座舱底面的两个比邻面边界(直线)重合。

采用计算机图形学中的 2D 四点纹理映射技术,将真实的视频图像添加到虚拟座舱模型底部的三个比邻四边形面上后,即可实现座舱视景系统的虚实图像融合。在座舱真实场景采集系统中,视频分割器输出的图像包括三个固定摄像机采集到的图像,但并不是所有的图像信息都是有用的。例如,对于座舱底部区域,有用的图像数据必须是座舱底部交互区域成像的那部分图像,这个有用区域称为图

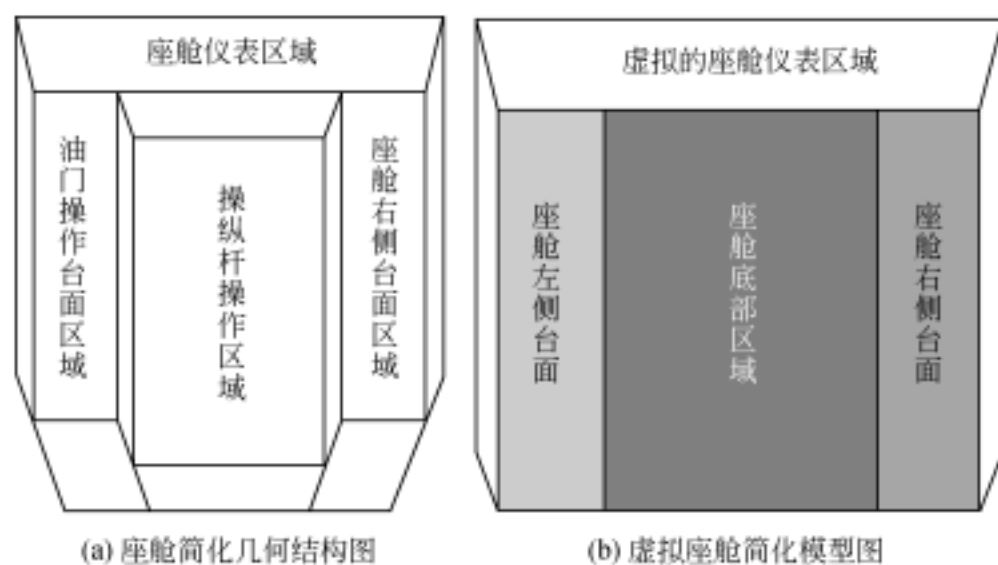


图 7.3 座舱视景

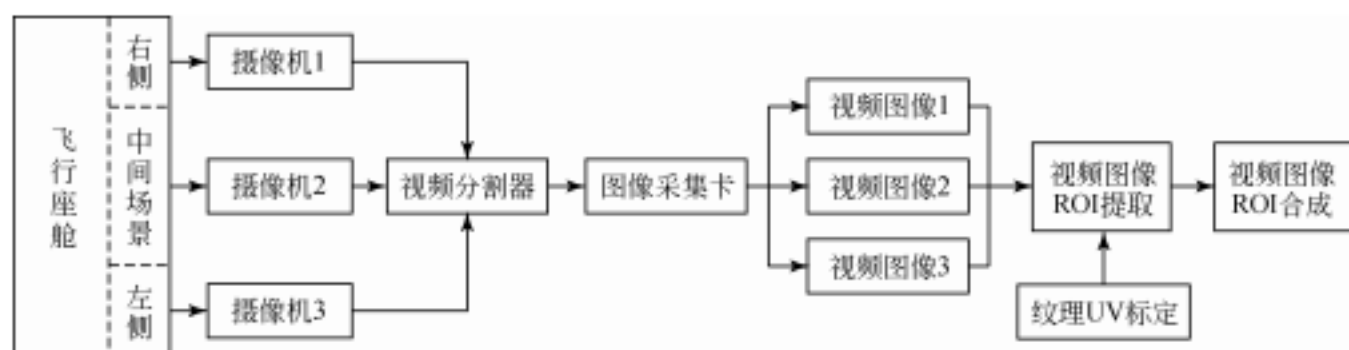


图 7.4 多固定摄像机的虚实融合流程图

像 ROI。对于每一个摄像机采集到的图像,它的兴趣区域都可以用纹理空间中四个点的纹理坐标包含的四边形区域来表示,如图 7.5 所示。

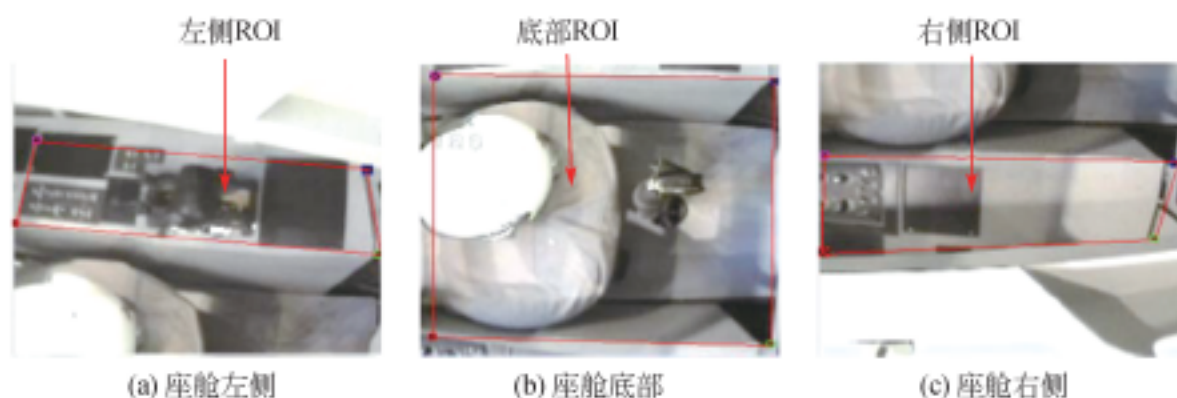


图 7.5 座舱三个区域图像中的 ROI

在系统使用前对三个兴趣区域四个点的纹理坐标进行标定的具体步骤如下。

(1) 稳固支架,调节好三台摄像机的镜头方向,确保座舱左右两侧及中间区域

分别在各自对应的摄像机视场内。

(2) 截取一幅四通道视频分割器输出的图像作为纹理空间的初始纹理,用四点纹理映射方法将三个通道图像中的 ROI 分别映射到对应的三个四边形面上。

(3) 对 ROI 的四点纹理坐标进行精细调节,确保在三个四边形面过渡边界上显示出平滑、连续的视觉效果。

当系统运行时,多固定摄像机座舱真实场景采集系统实时获得稳定的视频图像。系统依据标定好的四点纹理坐标,以实时刷新的视频图像为纹理,将该纹理图像中各个区域的 ROI 实时地添加到虚拟座舱的三个相邻的底面四边形面上。当操作者戴上飞行头盔低头时就能够通过 HMD 看到真实场景中发生的变化,既可以看到座舱左右两侧以及底部的真实结构,又能够看到自己的手操作油门和操纵杆的动作。图 7.6 显示了系统在运行过程中,由用户观察到的虚实融合、人-座舱自然交互的视景效果。

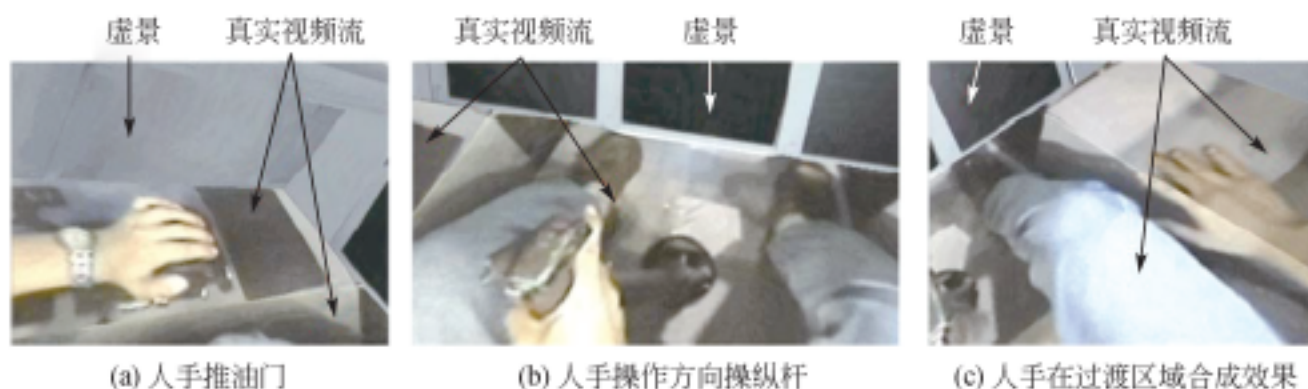


图 7.6 虚拟座舱的虚实融合图像以及人-座舱交互的示意图

7.1.4 样机系统实现与测试

系统在开发过程中使用了 VC++ 2005、Vega Prime、OpenGL、DirectX、OpenCV 等函数库实现了轻量级飞行模拟器的样机,样机的功能模块示意图如图 7.7 所示。编队服务器程序模块主要负责配置参数和飞行参数的数据传输和交换。主控制台模块负责设定训练科目、基本仿真参数、仿真速度,选择监视端监控对象,发送定止和定位操作命令,显示监控对象的实时飞行参数数据,显示监控对象的 2D 飞行航迹,对飞行训练结果进行评估等。监视端模块的主要功能是大地形数据库的管理、视景渲染,使用鼠标改变观察视角,实时监控座舱飞行的 3D 姿态和运动轨迹。座舱端程序模块实现的功能主要分为五部分。

(1) 视景仿真与管理:包括管理大地形数据库,渲染飞行环境场景图形,模拟

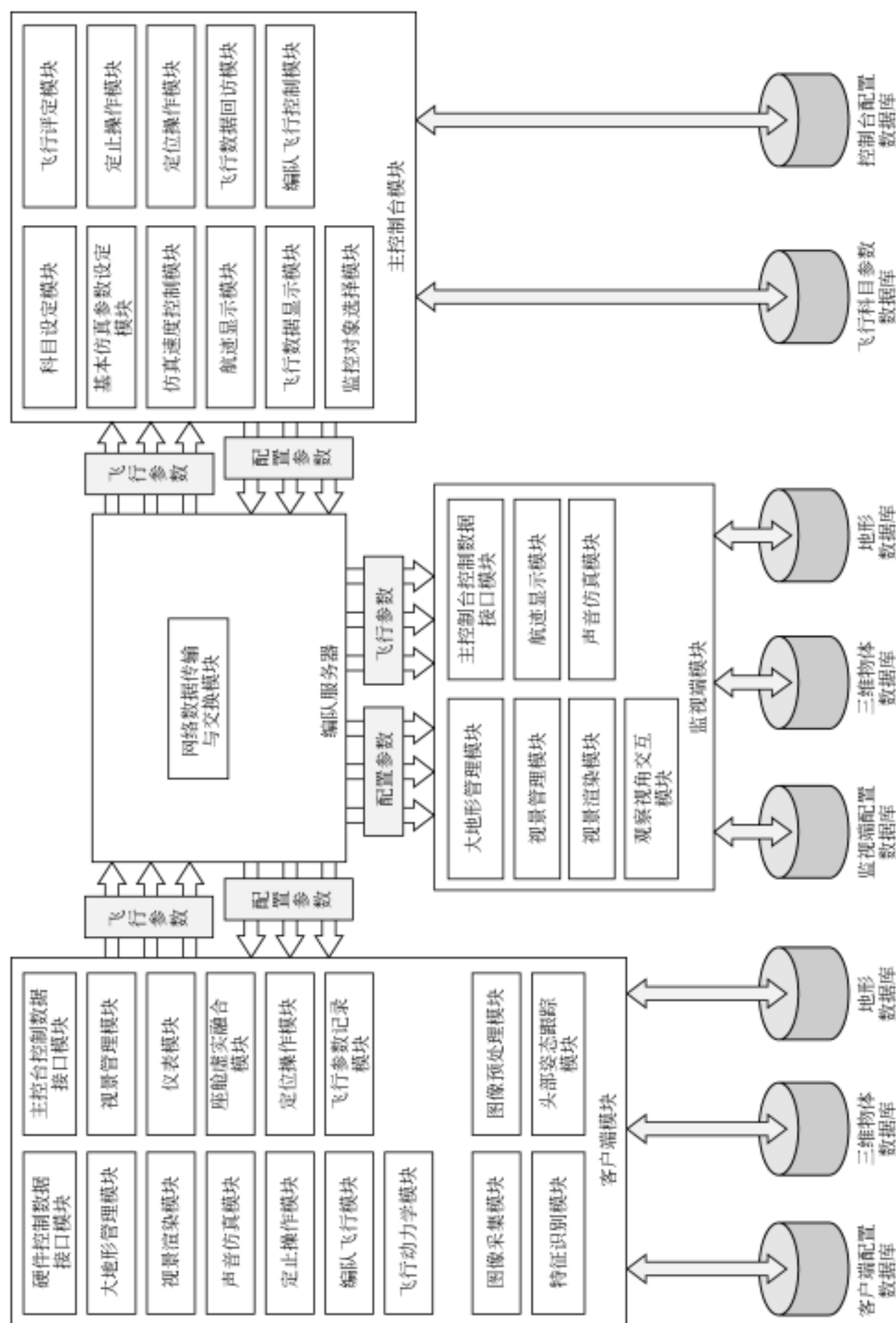


图 7.7 轻量级飞行模拟器样机系统功能模块示意图

座舱发动机声音和座舱飞行尾迹,实时显示座舱前仪表板上各个仪表参数数据。

(2) 座舱飞行驱动与管理:包括座舱使用 DSP 实时读取传感器模拟数据和写入参数数据以控制传感器状态,使用传感器解码数据进行飞行动力学计算,并实时向客户端和监视端发送飞行参数数据。

(3) 观察视角控制与交互:主要包括实时采集光学头部跟踪器 LED 的视频图像,实时提取 LED 特征点,使用计算机视觉领域的姿态估计算法计算跟踪器的姿态数据。

(4) 座舱视景虚实融合:主要包括使用多固定摄像机实时采集座舱底部区域的视频图像,使用纹理映射技术实现虚实融合。

(5) 飞行训练管理:包括接收主控制台的科目训练命令,进行定止和定位操作训练,进行双机编队飞行训练。

轻量级飞行模拟器样机系统飞行效果如图 7.8 ~ 图 7.10 所示。图 7.8 显示了飞行员在座舱端进行飞行训练过程中头部左转、正视以及右转从 HMD 中所观察到的飞行场景。图 7.9 显示了从监控端观察到的座舱在飞机场跑道上滑行、起飞以及收起起落架的视景图。图 7.10(a) 显示了进行双机编队飞行训练时,从座舱端观察到的相邻飞机在跑道上滑行的视景图,图 7.10(b) 显示了从监视端观察到的双机在跑道上滑行的视景图,图 7.10(c) 显示了从监视端观察到的双机起飞视景图。

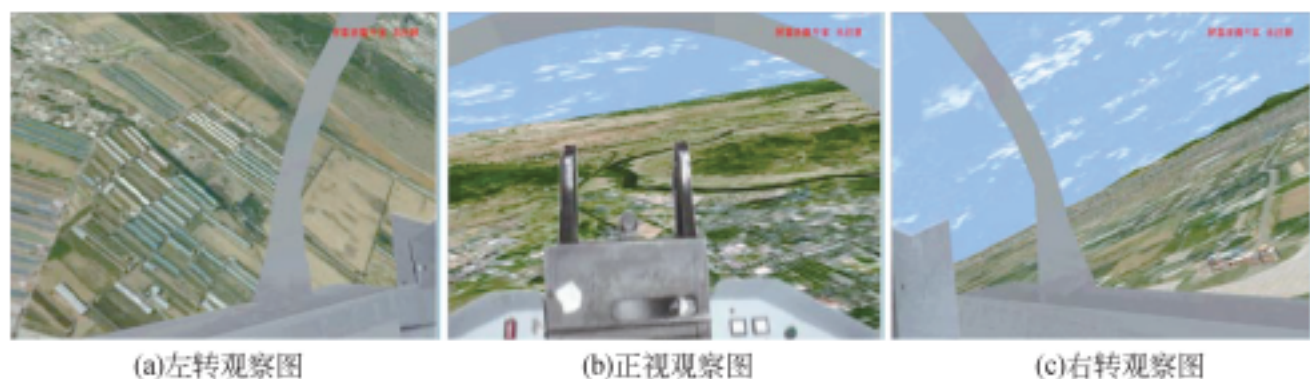


图 7.8 座舱客户端的飞行效果图

在完成上述样机系统的基础上,与空军某飞行学院合作对所开发的轻量级飞行模拟器进行了初步测试,并与该学院教学中使用的某大型模拟器和某桌面式模拟器进行了比较。测试用的大型模拟器使用六自由度运动系统的座舱系统和环幕显示设备,视景显示的最大角度为 120° ;桌面式模拟器则使用游戏操作杆驱动飞机飞行,使用三台 LCD 实现三屏显示。测试中组织 10 名飞行员和 10 名飞行教员依

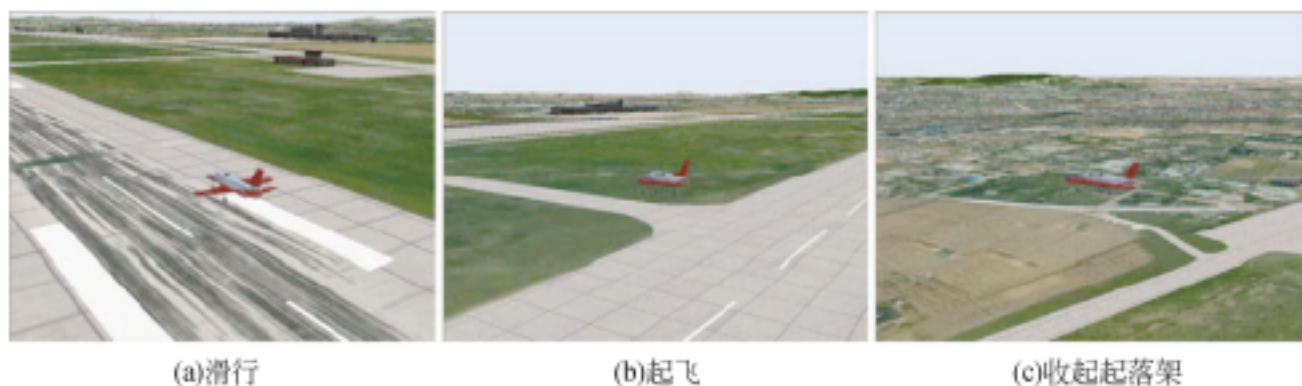


图 7.9 监视端的飞行效果图

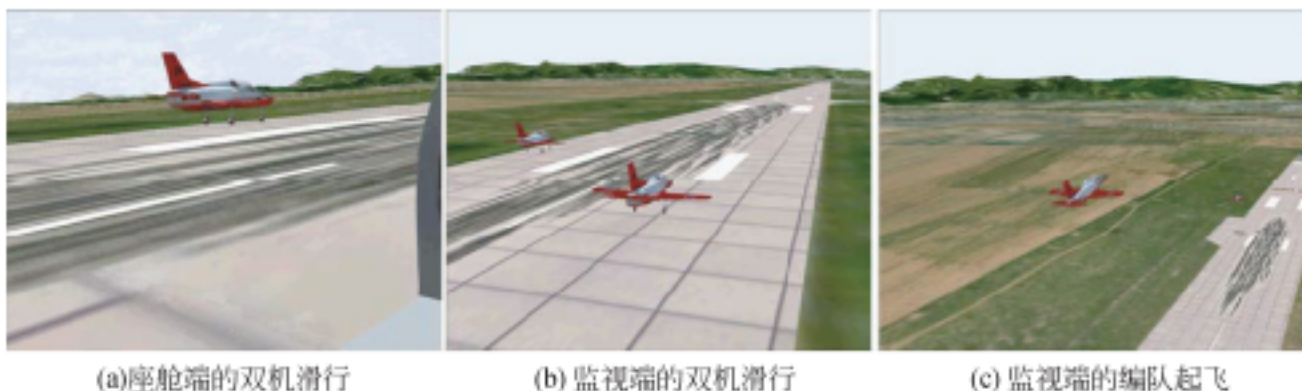


图 7.10 编队飞行的飞行效果图

次使用桌面式、大型以及轻量级飞行模拟器进行起飞、降落、特技等基本飞行训练以及编队训练。每一次测试后,让测试人员针对所测试的系统的人-座舱交互、沉浸感和多机配合等三个评估项进行满意度评定。

测试所设定的三个评估项的满意度评定标准均按照百分制来衡量,将评估分数分为三个等级,高于或等于 80 分为“非常满意”,60~80 分为“满意”,低于 60 分为“不满意”。在所有测试人员对系统的各个评估项打分后,对评估分数进行统计。统计结果为各评估项的分数在某一评分等级中的人数占总人数(20 人,10 名飞行员和 10 名飞行教员)的比例。统计结果表明,在人-座舱交互评估项中,大型模拟器在“非常满意”等级优势明显,但有多达 15 人对轻量级模拟器“满意”,仅有 6 人对小型模拟器“满意”。在沉浸感评估项中,大型模拟器优势也很明显;尽管没有人对桌面模拟器“非常满意”,但有 15 人选择“满意”;而对轻量级模拟器选择“满意”的人仅有 12 人,选择“不满意”的人多达 8 人。在多机配合评估项中,轻量级模拟器优势明显,有 7 人表示“非常满意”,13 人表示“满意”,而大型模拟器和桌面模拟器分别有多达 9 人表示“不满意”。

除了定量打分,让测试人员在每次系统测试后针对三个评估项填写对系统的感受,进行定性评估,解释自己给定的评估分数的缘由。例如,对桌面模拟器在编队配合评估项“不满意”的测试人员,认为在该系统中观察者转向、搜索邻机时需要依赖额外的键盘来控制,影响编队模拟训练。测试人员针对本系统的定性评估概括如下。

(1) 飞行学员基本认可当前的人-座舱交互方式,虽然没有大型模拟器那样全真实交互,但与油门、操纵杆的交互足以控制飞行。对人-座舱交互评估项不满意的测试人员期望能够与前仪表盘区域的起落架等操作设备交互。

(2) 系统的头部跟踪器使模拟飞行驾驶更贴近真实飞行驾驶,可在模拟飞行过程中任意时刻看到不同视线方向的飞行环境。特别是能逼真地模拟真实的编队飞行,可以灵活地搜寻邻机,获取邻机的相对位置,而这是大型模拟器所不具备的功能。

(3) 由于 HMD 显示分辨率的限制,使用系统现有配置的 HMD 所看到的虚拟仪表参数有限,期待能够看清楚更多的虚拟仪表参数,同时 HMD 视场角的限制导致观察视景时的沉浸感不足。

上述测试结果表明,该轻量级飞行模拟器与测试用的其他两个模拟器相比,在多机配合方面具有突出优势,而沉浸感是最弱的;在人-座舱交互方面虽然比大型模拟器差,但与桌面模拟器相比依然具有很大优势。

7.2 改进型增强现实轻量级飞行模拟器

7.2.1 改进型增强现实飞行座舱系统设计

7.1 节介绍的轻量级飞行模拟器,采用的是多固定摄像机虚实融合方法,仅能够实现座舱底部区域的虚实融合。如果需要实现人与座舱前仪表盘上特定部件的交互,则多固定摄像机方法将无法同时保证座舱的底部区域和前仪表盘区域的融合效果。解决这一问题的有效方法是改进原型系统中的头部姿态跟踪系统,增大系统的人-座舱交互区域,使得飞行学员可以与座舱内部全区域进行交互。图 7.11 为改进型轻量级飞行模拟器座舱系统结构示意图,在改进的座舱系统中,设计了一个由 IO 跟踪器和 OI 跟踪器组成的新型座舱姿态跟踪系统,一个门框形状的固定支架用于固定 OI 跟踪器的两台摄像机。选择十字形分布的 LED 作为 OI 跟踪器的跟踪目标,并将其固定在 HMD 顶端,IO 跟踪器摄像机固定在 HMD 前端。

选择 LED 或者 ARToolkit 标识板作为 IO 跟踪器的跟踪目标,并将其固定在座舱前仪表盘各仪器间的间隔区域。如果使用 ARToolkit 标识板,则 IO 跟踪器摄像机也可同时用作取景摄像机;如果使用 LED 作为 IO 跟踪器跟踪目标,则必须在 HMD 前端额外增加一台取景摄像机。改进型轻量级飞行模拟器座舱系统将不再使用三台固定摄像机实现场景的虚实融合和人-座舱交互,而是使用新型座舱姿态跟踪系统。新型座舱姿态跟踪系统的输出姿态不仅用于控制观察的视角,而且用于对准座舱轮廓内的真实视频图像和虚拟飞行环境,从而使得飞行员透过 HMD 能够观看到座舱轮廓内真实视频图像与座舱轮廓外虚拟飞行环境融合的视景。这样,当飞行员低头或正视座舱内部时就能够通过座舱轮廓内的真实视频图像看到自己操作座舱设备的动作,从而实现飞行员与座舱之间的全区域交互。

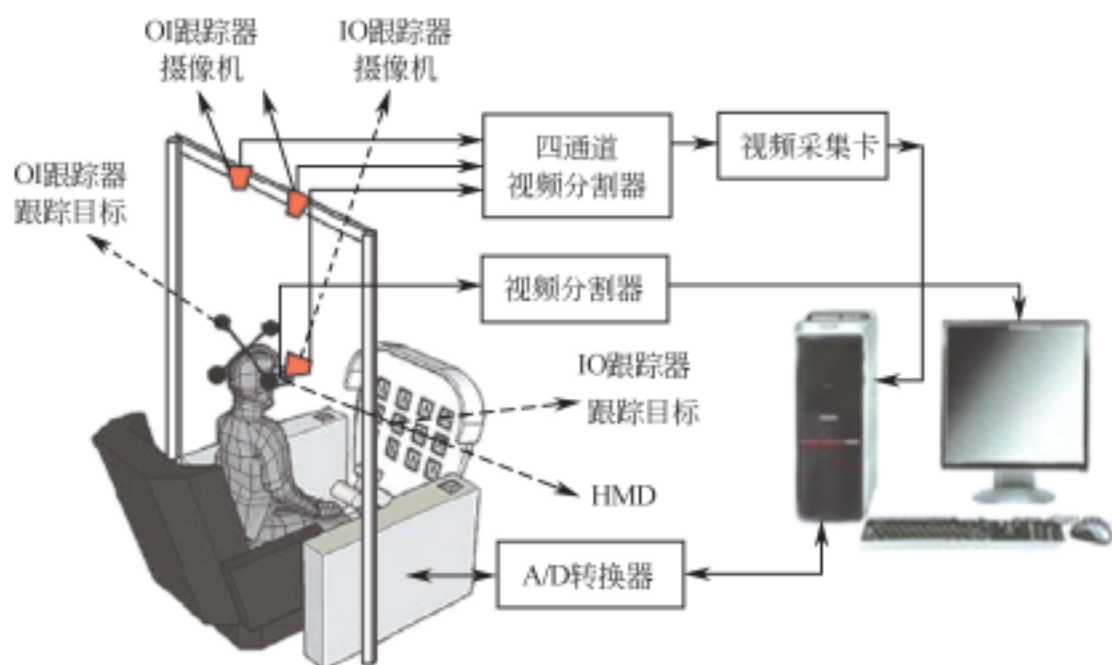


图 7.11 改进型轻量级飞行模拟器座舱系统结构示意图

7.2.2 座舱头部跟踪系统的实验模拟

根据座舱姿态跟踪系统的鲁棒性要求,采用基于航迹融合的传感器融合方法计算座舱姿态跟踪系统的输出姿态,如图 7.12 所示。姿态融合算法的两个主体为两个 EKF 和一个融合滤波器。前者分别用于 OI 跟踪器和 IO 跟踪器正常工作下的姿态信息平滑滤波,后者将来自两个 EKF 滤波后的姿态进行数据融合,融合后的姿态作为座舱头部跟踪系统的姿态输出。当只有 IO 跟踪器或者 OI 跟踪器工作时,经过 EKF 滤波的姿态数据直接作为座舱跟踪系统的姿态输出。

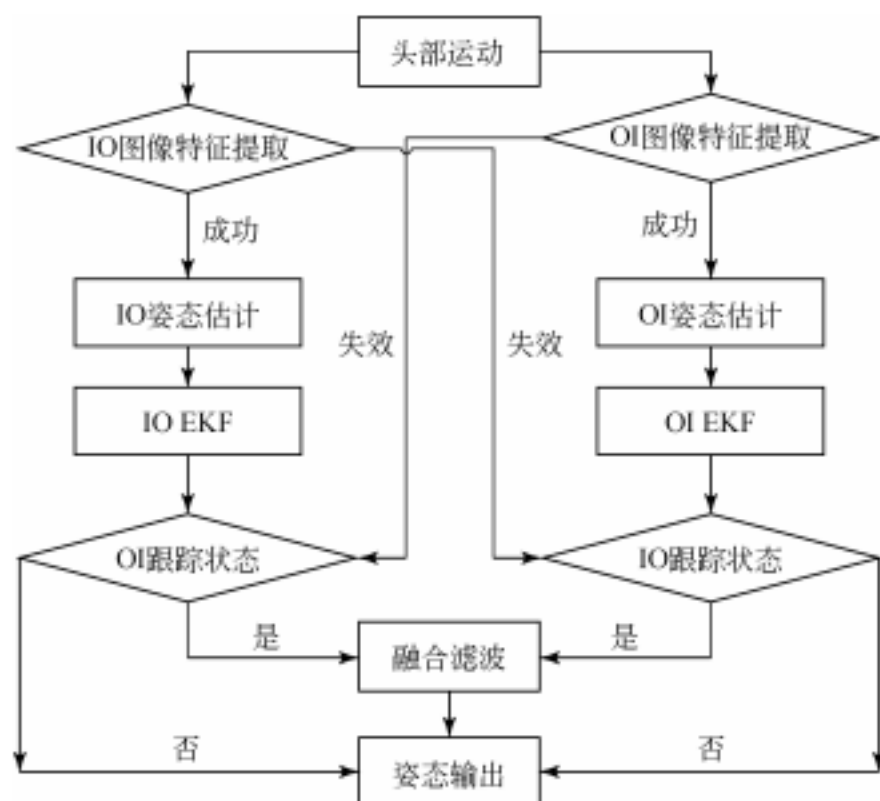


图 7.12 座舱姿态跟踪系统的姿态融合算法

7.2.3 跟踪器的实验对比结果

与 2004 年开发的 InterSense 座舱跟踪器相比^[1],该座舱跟踪系统具有以下特点。

(1) 使用更加简单的跟踪器结构扩展跟踪范围,增强鲁棒性。仅使用 IO 跟踪器和 OI 跟踪器组合结构,而 InterSense 座舱跟踪器增加了惯性跟踪器。

该座舱跟踪系统中,由于 OI 跟踪器的两台摄像机均位于 OI 跟踪目标上方,这使得即使 IO 跟踪器跟踪失败,还可以使用 OI 跟踪器进行跟踪。在 InterSense 座舱跟踪器中(图 7.13),由于其 OI 跟踪器的两台摄像机与 IO 跟踪器的摄像机是正面相向的,跟踪范围小,只有通过增加一个惯性跟踪器来扩展跟踪范围,增强鲁棒性。

(2) 可满足轻量级飞行模拟器系统中快速运动的跟踪需求。通过与飞行员交流,发现飞行员在使用轻量级座舱模拟训练时,精力主要集中于与座舱操纵杆以及仪表的交互,飞行员的头部运动大部分时间内都是比较慢速的运动,仅存在零星的快速运动。实验结果表明当运动目标运动比较快时,由于 IO 跟踪器的摄像机转动一个很小角度都会导致 IO 跟踪器跟踪目标的图像点产生很大的偏移,IO 跟踪器的跟踪目标成像点容易产生变形甚至模糊而导致跟踪误差或者跟踪失败。然而,

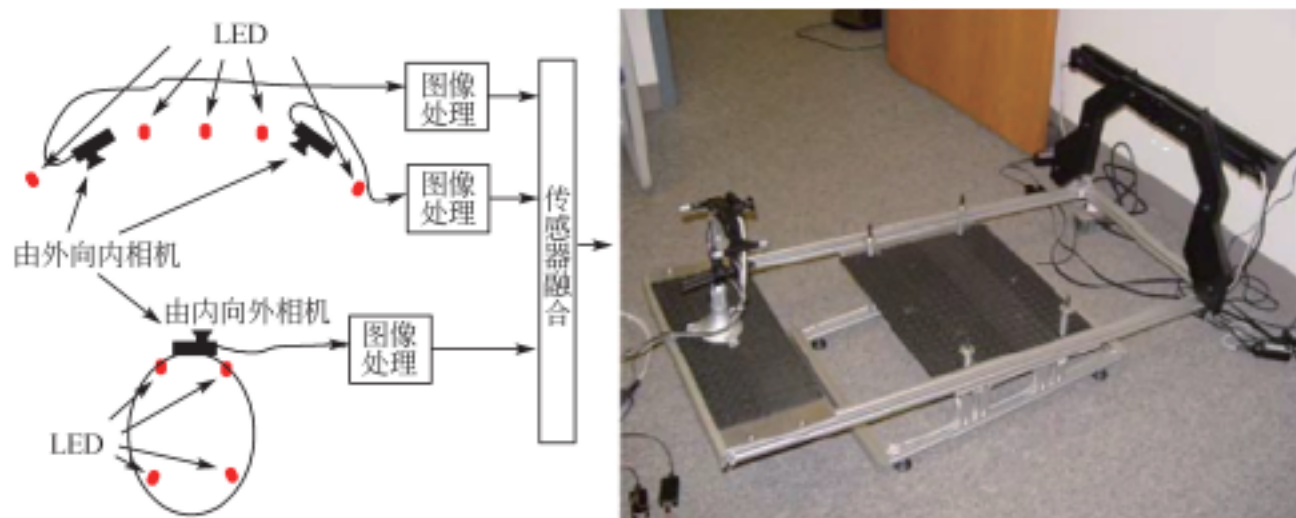


图 7.13 Foxlin 等所使用的 inside-outside-in 结构(左图)和实验装置(右图)

即使在 IO 跟踪器跟踪失败的情况下, OI 跟踪器的结构决定了其还能正常工作。这主要是由于 OI 跟踪器运动目标的平移和旋转都仅导致在 OI 跟踪器两个摄像机像平面上的图像点产生很小的偏移, 纵使零星的、较快的运动也不会使成像点产生模糊。此外, 在 IO 跟踪器间或跟踪失败的条件下, 可以使用前一帧的姿态数据以及跟踪算法中的 EKF 进行姿态预测, 然后与 OI 跟踪器的输出姿态进行滤波融合输出。这与 InterSense 座舱跟踪器中增加一个惯性跟踪器以实现快速运动跟踪的原理不同。

参考文献

- [1] Foxlin E, Altshuler Y, Naimark L, et al. Flight Tracker: A novel optical/inertial tracker for cockpit enhanced vision//IEEE and ACM International Symposium on Mixed and Augmented Reality, Arlington, 2004.

第8章 增强现实技术的应用实例

——圆明园定点观景器

本章主要介绍基于增强现实技术的圆明园定点观景系统。圆明园是我国清朝自康熙年间起,集中全国人力和物力历时百余年兴建的一座规模空前的山水离宫,素有“东方凡尔赛宫”和“万园之园”的美称。1860年,圆明园毁于第二次鸦片战争。如今的圆明园是一个遗址公园,是中国近代屈辱历史的见证,对于了解中国历史和进行爱国主义教育具有非常重要的意义。新中国成立以后,对圆明园的修复一直存在争议。目前参观者只能通过图片资料间接了解过去的盛时园林景象。应用增强现实技术可以在完整地保持遗址风貌的基础上,利用虚实场景融合,在现场直观地恢复昔日的盛世景象。从而与遗址的废墟形成鲜明对比,给参观者留下非常深刻的印象。它既可以精致地再现圆明园无与伦比的园林艺术,展现中华民族古典文化的博大精深,又可以使人们了解现代科技的发展和应用,是爱国主义教育的极好手段,具有很高的社会意义和实用价值。

在北京市和海淀区政府部门的支持下,北京理工大学信息科学技术学院承担了基于增强现实的数字圆明园的研究项目。本章就其所开发的第一代、第二代、第三代不同的增强现实定点观景器技术作详细介绍。

8.1 第一代定点增强现实观景器

为实现圆明园的现场重建,定点式AR观景系统可以在固定的地理位置为用户提供现场实时立体增强效果。设备可以水平 360° 、垂直 $\pm 30^{\circ}$ 旋转,用户可使用云台手柄在水平 360° 、垂直 $\pm 30^{\circ}$ 的范围内转动摄像镜筒,在圆明园景区现场观察实时重建在废墟上的大水法、观水法、远瀛观、海晏堂等建筑的原始三维模型,并可以与实际的外界图像进行叠加融合显示。

这一系统具有真实、虚拟、增强三种工作模式,分别显示真实场景环境、完全虚拟的视景效果、虚实融合的增强现实效果。可以通过人工控制,对真实和虚拟场景进行同步变焦操作,实现视景的比例缩放。系统可以根据地理位置和时间自动调节模型的光照方向,同时通过人工控制光照强度,使虚拟物体与真实场景更好地融

合,以实现虚拟物体和真实环境的无缝拼接,达到真假难辨的虚实融合效果。系统的控制面板为用户提供了镜头变焦,调整模型光照强度,真实、虚拟及增强场景模式切换,视频和声音播放等多种人机交互手段。

用于圆明园三维数字现场重建的定点观察增强现实系统为实时单用户旋转镜筒式,它主要由立体增强现实光电立体成像子系统、姿态传感子系统、图像处理子系统和同步转动平台子系统组成,系统结构如图 8.1 所示。图 8.2 是系统的外观以及对应子系统。其中,光电立体成像子系统由双路 CCD、双路显示屏以及立体目视镜组成,功能包括拍摄现场的图像,同时将增强的图像以立体方式显示。姿态传感子系统由系统云台和光电编码器组成,用于探测镜筒的姿态数据。图像处理子系统根据镜筒的姿态数据对圆明园昔日场景的计算机模型进行变换和渲染,与真实的环境图像叠加。实现虚实融合同步转动的机械平台是所有设备的载体,可以使机器整体与用户随动,同时载荷有 UPS 和集电环,实现系统安全供电和避免电缆纠缠。

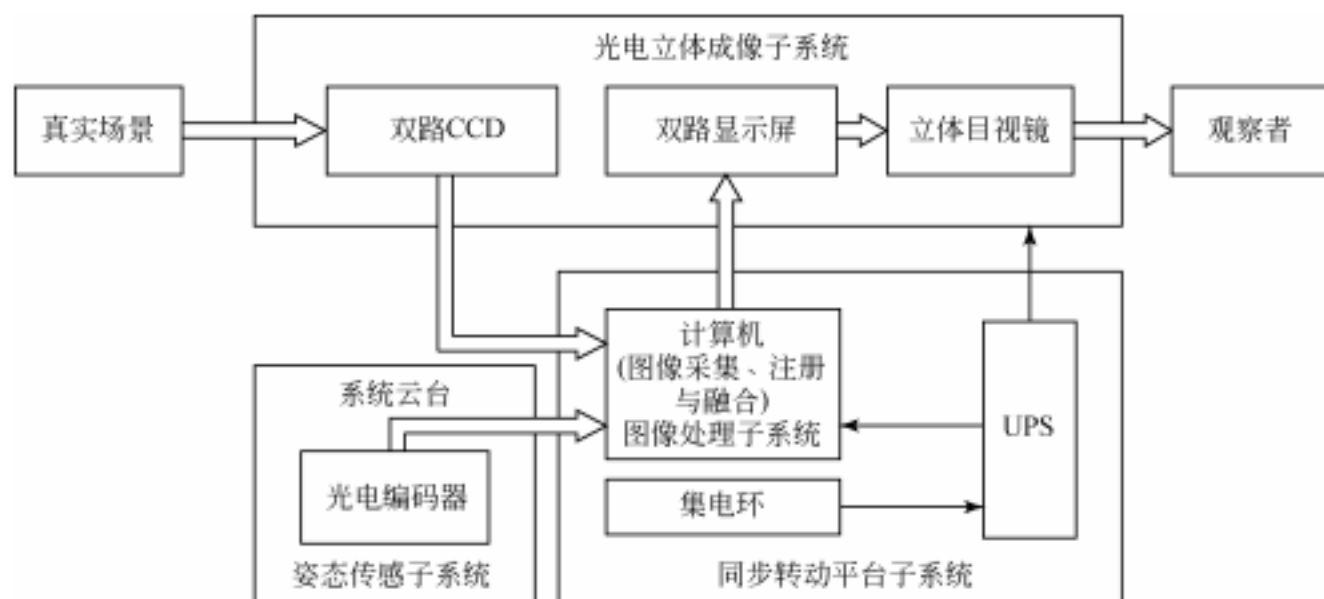


图 8.1 增强现实定点观察系统的结构组成

(1) 光电立体成像子系统。光电立体成像子系统的功能是获取真实环境图像和显示增强的图像。为了表现真实的、身临其境的感觉,图形采集和显示部分采取立体视觉通道的设计方式,即左右两路摄像机分别对应人的左右眼,显示系统使用双 LCD 和双通道目视镜。双路摄像机拍摄的带有视差的真实场景图像与经过注册的虚拟物体图像融合以后,在左右显示器上同步显示,经过双通道目视镜合成,可以使用户观看到具有立体视觉效果的增强场景,体验到身临其境的浸没感。

(2) 姿态传感子系统。姿态传感子系统的功能是获取注册定位信息。因为定



图 8.2 增强现实定点观察系统的外观及对应子系统

点系统被放置在固定的观察位置,所以通常虚拟模型注册定位所需要的六自由度信息减至二自由度,即只需要实时获取放置在云台上的镜筒的俯仰角和偏航角,降低了系统实现的复杂度。俯仰角和偏航角是通过与云台系统连接的高精度光电编码器获得的。光电编码器采用圆光栅莫尔条纹和光电转换技术,将机械轴转动的角度量转换成数字电信号输出给单片机,进行角度的测量。光电编码器具有体积小、重量轻、力矩小、可靠性高、寿命长的特点,其测量精度可达到 0.07° ,使得定点观察系统满足虚实图像融合的精度要求。因此,定点观景器的姿态传感子系统由系统云台、两个光电编码器及齿轮传动机构构成。两个光电编码器通过固定架与云台体固定,其转轴分别通过齿轮传动机构与云台体的水平方向和俯仰方向联动实现姿态角的跟踪。用户操纵云台手柄旋转镜筒,带动光电编码器测角机构联动,并将其输出角度信息传输至计算机,实时完成视点方位的跟踪注册。

(3) 图像处理子系统。图像处理子系统的主要功能是实现虚实场景融合,其主体是工控机和一个基于 DirectX 的实时漫游引擎。可以根据姿态传感子系统所测得的角度信息,计算出昔日场景的三维数字模型在真实场景中的映射坐标,经过投影变换和渲染得到相应的投影图像,再利用图像融合技术将其叠加在真实环境的图像之中。必要时可添加相关文字、声音等增强信息,分左右视点取样图像,提供给光电立体成像子系统,形成具有沉浸感的立体增强现实图像。

(4) 同步转动平台子系统。同步转动平台子系统是一个和谐的随动机构,其作用是载荷工控机、不间断电源(UPS)和其他设备,实现系统与用户同步转动,防止电缆纠缠,简化外部设计。它安装在系统底盘上,含双层套筒、同步转动平台、轴承、集电环、数据及电气接口双层套筒的内层与系统底盘固定,外层与云台相连接,可以转动。工控机、UPS等装置安装在与外层套筒联装的同步旋转平台上,集电环向同步旋转平台上的UPS提供动力电源。用户通过云台联动杆驱动云台和双层套筒的外层套筒旋转,进而带动同步旋转平台以及其上的工控机、UPS等进行同步旋转。在实现系统在水平面上 360° 全向旋转的同时,避免了电源线及信号线在使用过程中的缠绕现象。通过不间断电源保证了室外环境下对整个系统安全可靠地持续供电。图8.3所示为用户从AR-View中观看到的遗址场景增强显示效果。



图8.3 增强现实效果对比

8.2 第二代增强现实定点观景器

8.2.1 第二代增强现实定点系统的优点

从外形看,第二代系统与第一代系统较为相似。但是,第二代系统相对于第一

代系统作了很大的改进,以达到更加适应工程使用的效果。具体的改进如下。

(1) 将第一代定点观景器中的庞大光学系统大幅加以简化,用双摄像机获取真实图像,通过软件生成具有视差的双眼视差图像,最终分别显示在左右两眼的微型显示器上。这个改进使得光学立体成像子系统的体积更小,重量更轻,更便于观察者轻松地改变所观察的视角。

(2) 第二代样机有效地精简了第一代设备中庞杂的转动机构,齿轮传动等装置也大大简化。这样的改进减小了整个样机的复杂程度,提高了可加工性。观察者可自己用手控制观察方向,该改进是一种更加直接和方便的交互方式,同时也便于整个系统的维护,提高了系统的稳定性。

(3) 在第二代样机中增加了立体声效果,使大水法中水的声音也变得立体了,用户可通过佩戴蓝牙耳机,在转动装置的同时也感受到声音的立体变化。当用户面向大水法时,会感到水声是迎面扑来的,但是如果用户把装置转到右边,便会感觉声音从左边传来,同理,当装置转向左边,便会觉得声音从右边传来。这样,能够给用户增加更强的浸没感,大幅提升用户的体验效果。

8.2.2 第二代系统的构建

实验设备如图 8.4 所示,分为服务器端和用户端。服务器端包含用于处理数据的计算机、数据处理盒、不间断电源,用户端包含显示采集单元、跟踪器和投币器。



图 8.4 第二代增强现实定点观察系统

跟踪实验系统如图 8.5 所示,搭建一个可水平旋转的平台,平台为一个足够重量和大小的圆盘,用以固定其上所有装置。水平旋转编码器安装在转轴中,由一个轴承与支撑柱相连接,因此可获得平滑的支撑柱水平旋转角度。支撑柱与底盘保持垂直,在支撑柱顶部的平台上安装实景摄像机。平台底部俯仰旋转轴与俯仰旋转编码器相连,因此可获得旋转平台的

俯仰角度。旋转编码器均采用 Koyo TRD 增量式编码器。如图 8.6 所示的控制设备用于控制双路角度跟踪,处理后由串口输出给计算机。

双目显示单元采用一对独立的凸透镜组成,配合 3.5in LCD。显示视场角大约

为 42° , 出瞳距离大约为 18mm。整个显示单元尺寸为 $20\text{mm} \times 17\text{mm} \times 8\text{mm}$ 。实景采集摄像机由双路 Honeywell GC-655P-G 摄像机组成。虚实融合在三维渲染引擎 Virtools 中完成, 在实现大水法建筑虚实叠加的同时, 由粒子效果实现仿真喷水效果, 并配合二维立体声效果。为实现自助普适娱乐应用, 设备增加了投币器功能。系统开始呈现欢迎界面, 由投币启动激活观看程序, 在一定时间后结束单次观看, 回到欢迎界面。



图 8.5 实验样机的跟踪系统



图 8.6 服务器端控制设备

8.2.3 增强现实定点观景器的用户使用情况调查

在圆明园管理处的协助下, 在圆明园大水法侧前方搭建临时观景台, 将定点增强现实观景设备安放在观景台上, 以获得更好的观赏角度。在 2008 年国庆节假期的 3 天中, 有超过 400 人使用本设备, 如图 8.7 所示。其中 100 人填写了调查问卷, 问卷针对设备印象、使用方式、功能和价值等几方面设计问题, 如表 8.1 所示。



图 8.7 游客争相体验第二代增强现实定点观察系统

表 8.1 调查问卷及意义

编号	问题	目的
1	你认为设备外观是否很好?	设备印象
2	你认为大水法的虚拟模型是否精美?	设备印象
3	你认为设备是否简单易用?	使用方式
4	你认为是否能迅速调整视差,以适应立体显示效果?	使用方式
5	你认为投币器是否好用?	使用方式
6	你认为是否能直接使用设备而不需要设备介绍与指导?	使用方式
7	你认为设备是否有助于了解历史遗迹?	功能
8	你认为设备是否适合大众使用?	价值
9	如果收费是否使用该设备?	价值
10	你认为设备是否能应用于其他同类遗址?	价值

针对问题设计李克特量表四维评分形式设计答案为:“非常同意”、“同意”、“不一定”、“不同意”。由 100 份问卷答案统计,如表 8.2 和图 8.8 所示,填写问卷人年龄为 9~68 岁。

表 8.2 问卷评分结果 (%)

问题编号	非常同意	同意	不一定	不同意
1	16	41	29	14
2	23	51	24	2
3	27	52	18	3
4	32	34	27	7
5	40	38	22	0
6	36	33	20	11
7	22	44	20	14
8	49	45	6	0
9	40	48	11	1
10	47	39	12	2

从问题 1 和问题 2 来看,有超过一半的人表示能接受设备外形和虚拟模型效果,但结果并不乐观。仍有较多的人认为外形和虚拟模型渲染效果有待提高。虽然对于虚拟模型效果的认可,不同人的评价存在较大的主观性,但设备印象方面仍是下一步需要改进的重点。

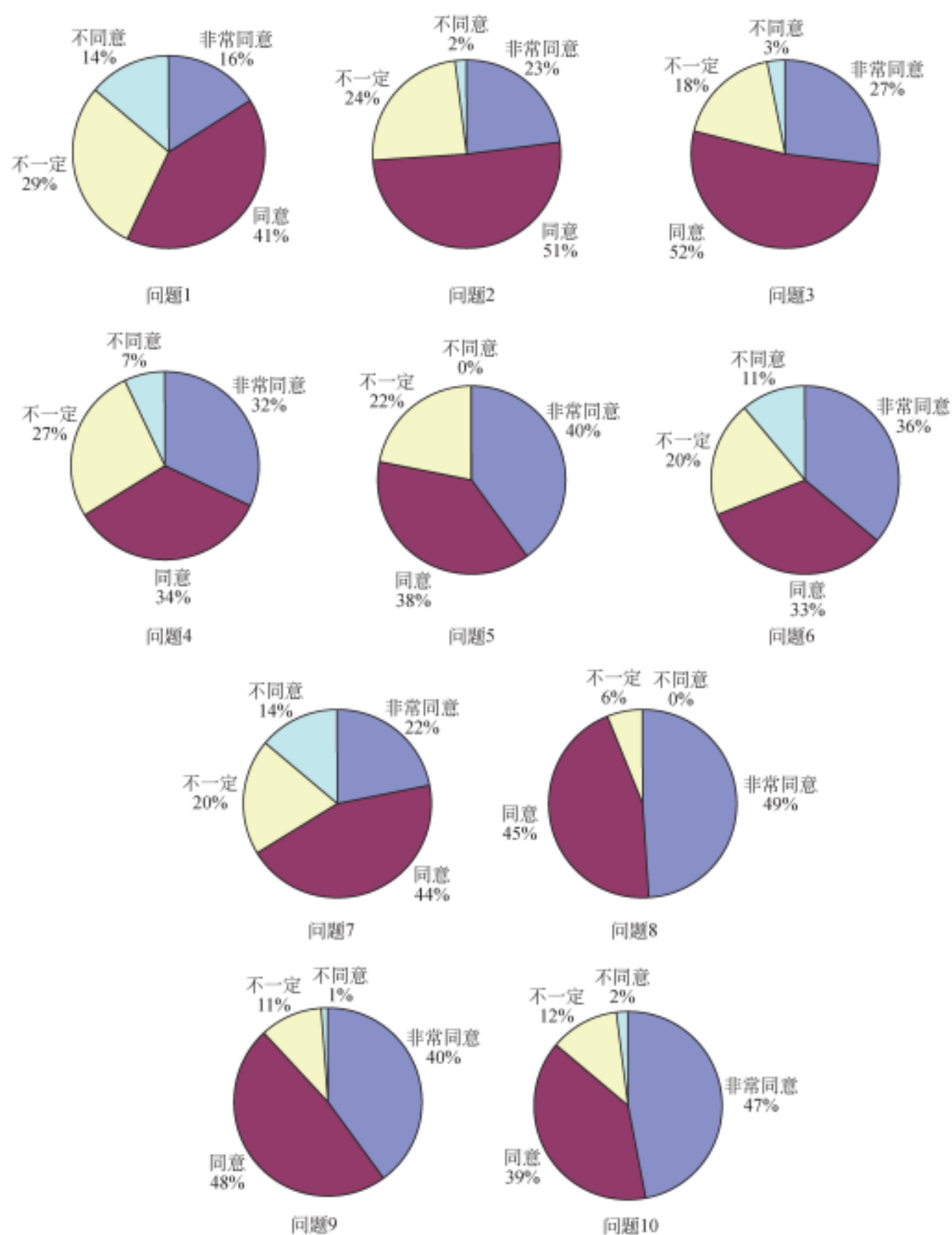


图 8.8 调查结果饼状图

从问题 3 来看,有 27%+52% 的人认为设备使用简易,但由于不少使用者是老人和小孩,他们抱怨设备设计的高度较高,在使用时会出够不着观景器的问题。从问题 4 来看,共有 34% 的人认为不那么容易能顺利观看立体图像,证明立体成像单元还有待考量。从问题 5 来看,40% 的人非常同意,38% 的人同意,看来投币器为普通用户所熟悉,其加入得到了他们的认可。从问题 6 来看,69% 的人回答同意。在使用过程中,在一段时间内只给用户代币,不提供如何使用设备的指导。在观察中发现,大部分用户会熟悉的投币,然后凑近观看设备。当然,完全没有监管也出现了其他问题,由于设备只能旋转 270°,否则会导致连接线缠绕问题,这也是将用户端和设备端分离的一个缺点。

从问题 7 和问题 8 的回答结果来看,绝大部分人认为使用设备对理解和认识圆明园遗址有帮助。从问题 9 来看,大部分人认为设备在收费后仍然具备吸引力,同时在其他地方有可推广性,不过他们表示价格不应过高。

从问卷调查来看,虽然一些问题的答案带有较强的主观性,不同的人可能有不同的结果。但是从比例来看,还是可以总结结果。正面的结果是新设备虽然还处于原型机阶段,但经受住了普通观众的使用,同时从设备观看效果来看,普通观众能够接受并使用顺畅。增加的投币功能是下一步商业化大范围使用的基础。从另一方面来看,立体显示单元仅用一对凸透镜难以适应不同用户的需要,而设备外观与适合公众使用的人性化设计都将是急需改进的。虽然客户端和服务端分离减少了客户端设备的复杂性和重量,但会带来电线缠绕等使设备在独立无监管运行中出现的新问题。

8.3 AR-View 增强现实定点观景器

根据第二代原型机的调查经验,如何让设备更加适宜大众使用成为主要改进方向,更完善的功能、人体工程学和工业设计都应包含在新的产品设计之中。下面的新一代定点观景器被命名为 AR-View,如图 8.9 所示。AR-View 的功能特点如下。

- (1) 舒适大角度的显示单元。
- (2) 一体化集成设计。
- (3) 时尚工业化外形与人体工程学结构。
- (4) 可水平旋转 360°,俯仰 135°。
- (5) 总质量为 170kg。

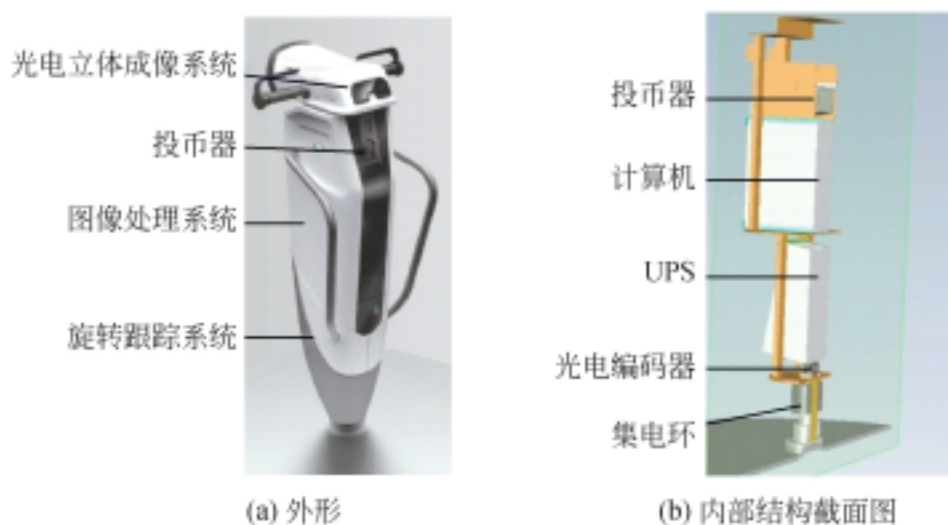


图 8.9 AR-View 系统的新设计

(1)大角度显示。与立体显示相比,大角度显示更容易被普通公众接受,且显示效果沉浸感好。重新设计光学显示单元,其出瞳直径达到 8mm,出瞳距离为 19mm,视场角为 52° (注:显示单元光学设计非本书研究内容,由其他合作研究人员完成)。配合 0.61in OLED(organic light-emitting diode)显示部件,分辨率达到 800×600 像素。如图 8.10 所示,显示设备仍由双目光学显示单元和双 OLED 显示部件构成,仍可显示双目立体图像。但考虑到普通用户对立体显示适应的问题,在新一代定点观景设备中选择了双目非立体显示,并由一路高分辨率摄像机采集真实场景。

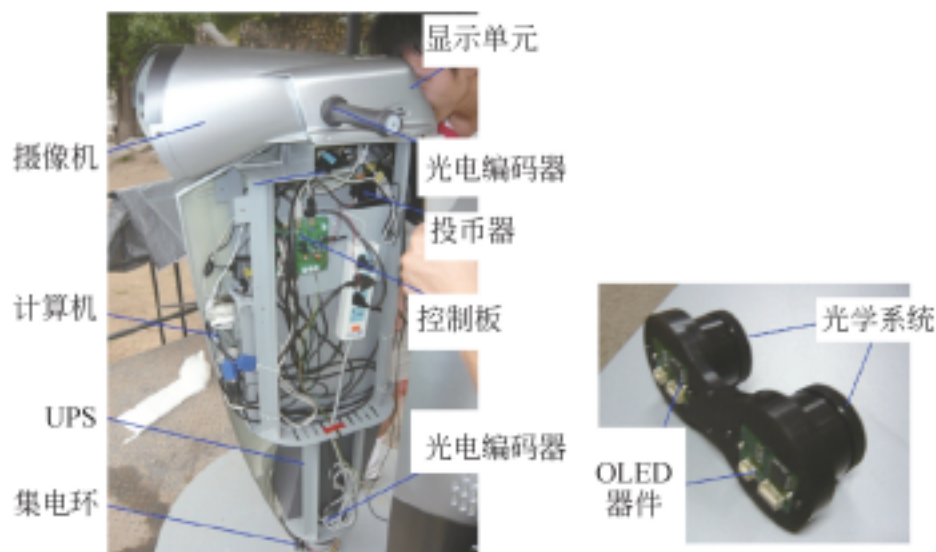


图 8.10 AR-View 的系统结构图与其显示部分

(2)一体化设计。为解决分布式带来的问题,同时必须避免一体化使设备变

得笨重,设备选型变得尤为重要。重量轻、体积小和设备性能需要同时考虑。选用惠普 DC7900 型计算机作为图像处理系统, Sony HDR-HC52E 摄像机为实景采集摄像机, 仍然使用 Koyo 光电编码器作为角度跟踪器。重新设计设备外形, 计算机、UPS、投币器、集电环都集中在设备腹部; 显示与采集部件都在设备顶部。设备结构如图 8.10 所示, 其中只有中心支撑结构是金属部件, 其他部分是工程塑料, 尽量减小设备的重量。由于一体设计, 系统可以自由地实现 360° 旋转。

(3) 实用人性化设计。如图 8.11 所示, 设备外观更加时尚。同时为适应普适娱乐需要, 设备高度降至 1550mm, 观看目镜位置低于 1500mm。同时增加两个把手, 方便小孩轻松旋转设备。对于个子高的人, 特别设计目镜表面为倾斜面, 以便贴近舒适。在圆明园大水法景区, 加固先前的临时平台, 2009 年 9 月, AR-View 正式由园区管理进入对公众开放的试运行阶段。



图 8.11 圆明园实际应用中的 AR-View

AR-View 不仅在圆明园遗址公园大水法景区面向公众开放, 同时在中国科技馆新馆中也作为普及增强现实技术, 应用于“昔日重现”展区, 每天吸引大量观众使用。在北京市规划馆 2010 年改造中, AR-View 也作为普适娱乐设备被中外游客使用。

8.4 随动式增强现实定点观景器

为了进一步提高系统的跟踪精度, 同时减少客户端体积和重量, 让系统更加实用, 系统构架采用基于反馈式的随动控制跟踪系统 (feedback follow-up tracking

system, FFTS)。传统增强现实实时跟踪注册技术的研究方向和重点一直是不断提高各种跟踪方式自身的跟踪精度,并考虑将多种跟踪器的跟踪数据进行数据信息融合,从而以取长补短的方式获取更加精确的跟踪数据。此研究方向的确取得了很大的进步,但进一步提高跟踪精度的要求,单一地依靠跟踪器本身的精度提高已很难满足需求。于是,系统参考了同样以高性能、高精度的最优控制问题为主要研究方向的自动控制原理,考虑将该领域解决此类问题的思想用在需要解决的问题中。

在成熟的自动控制理论中,可以采用不同的原理和方式对被控对象进行控制,但最基本的一种是基于反馈控制原理的反馈控制系统。早在1930年,Wiener就将生物与机器的控制归纳为反馈控制。在反馈控制系统中,控制装置对被控装置施加的控制作用取自被控量的反馈信息,用来不断修正被控量和控制量之间的偏差,从而实现对被控量进行控制的任务,这就是反馈控制的原理。因此,自动控制系统具有两种最基本的形式:开环控制与闭环控制。

(1) 开环控制是一种最简单的控制系统,只有输入量影响输出量,对输出量没有监测,当出现扰动时输出量将偏离给定量的数值。显然开环控制不具备抗干扰能力,多用于干扰影响不大、控制精度要求不高的场合。由此分析,常见的增强现实跟踪系统就属于此类,如图8.12所示,由跟踪器获得的用户实时三维位姿传输给虚拟场景,控制虚拟摄像机完成虚实配准。

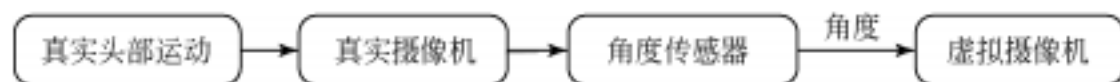


图 8.12 传统增强现实角度跟踪方式

(2) 闭环控制是在开环控制的基础上引入人工干预过程演变而来的,除了输入量参与控制外,输出量也参与控制。闭环系统中检测输出量,并将其转换为与输入量相同的物理量参与控制。由于有了反馈,系统具有更好的抗干扰能力。从结构上看,闭环控制系统除了有正向通道,还有从输出端到输入端的反馈通道。为了在增强现实跟踪系统中引入反馈通道,可将 FFTS 用于增强现实跟踪,其基本结构如图 8.13 所示。角度传感器实时测量真实用户头部角度,由该角度控制电机和机械结构组成执行单元,执行单元带动被控单元——真实场景采集摄像机;光电编码器测量被控单元的真实角度值,作为反馈进一步控制执行单元运动;最后将被控单元的真实角度传输给虚拟摄像机,完成反馈回路,保证真实摄像机与虚拟摄像机的角度误差在很小的范围内,精确虚实配准。

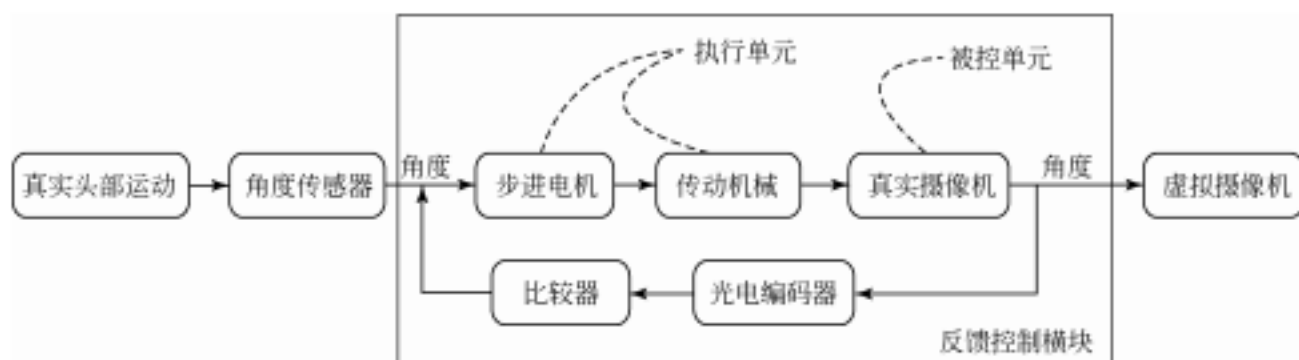


图 8.13 带反馈的增强现实跟踪方式

针对古迹复原的应用要求,除了跟踪精度的提高,系统在构架时需要充分考虑增强现实系统在古迹复原应用中的轻便性与实用性要求。第三代系统采用客户端/服务器模式,由服务器完成所有数据处理和图像渲染功能,客户端只完成终端显示和数据采集的任务,形成瘦客户端。

如图 8.14 所示,服务器端由随动观测器和控制计算机构成,其中随动观测器包含高速机械随动转台、角度测量器、实景图像采集摄像机是随动控制中反馈控制模块的主要构成部分。控制计算机中包含高速转台的电机控制和随动跟踪控制部分,同时虚实融合场景管理和渲染都由控制计算机完成。客户端由虚拟现实沉浸式头盔显示器作为增强现实的显示单元,在头盔显示器上绑定用于头部角度跟踪的罗盘角度跟踪器。对于客户端而言,带有角度跟踪器的头盔显示器可最大程度

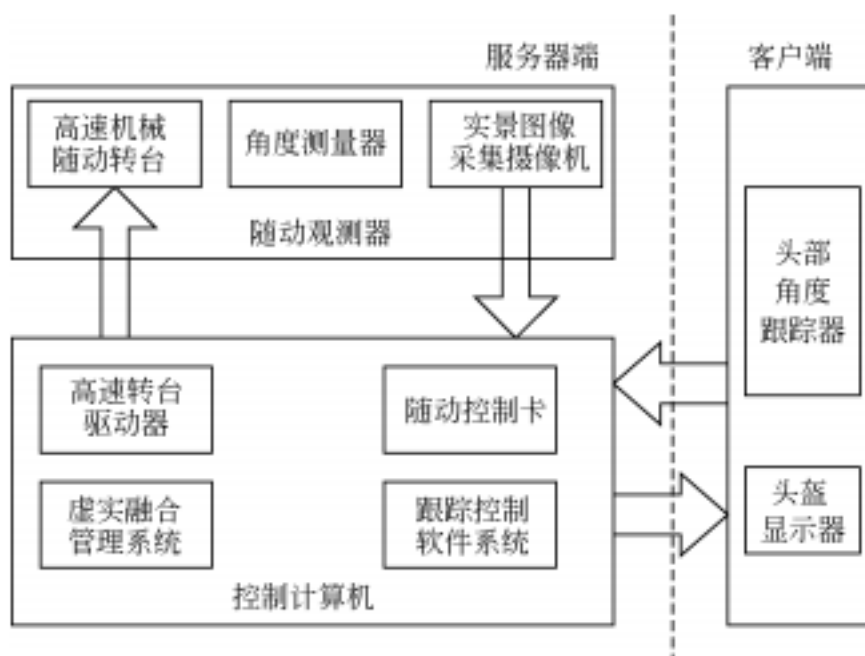


图 8.14 系统构架图

地降低操作负担,提高操作实用性。角度跟踪器将角度测量值传输给计算机,控制计算机将配准的虚实融合图像传输给显示器,用户只需佩戴轻便的头盔显示器就可以自由观看。

8.4.1 系统构成

服务器端和客户端配合完成的反馈式随动跟踪部分的系统流程图如图 8.15 所示。服务器端硬件设备连接图如图 8.15(a)所示。用户佩戴浸没式头盔自由移动时,角度跟踪器实时跟踪用户头部的三自由度角度信息。由于在定点观察系统中,人头部移动观看时主要是航向角和俯仰角的变化。同时考虑到三维转台在结构上的复杂性,所以只将三自由度角度信息中的航向角和俯仰角作为角度跟踪信息参与处理和控制在。在利用角度控制高速转台前,需要检测角度信息是否超过转动阈值,因为俯仰角在 $0^{\circ} \sim 180^{\circ}$ 范围内变化,二维高速转台的俯仰角在 $0^{\circ} \sim 90^{\circ}$ 范围内,需要在俯仰角超过电机能够旋转的角度时限制电机的旋转。经由判断的角度信息可传输给二维高速转台并控制其运转,角度检测装置实时检测当前转台运转角度形成反馈,控制二维转台旋转至指定角度,最后由转台真实旋转的角度值实时控制虚拟摄像机,形成虚实融合的图像。

在服务器端,步进电机和机械结构构成的二维高速机械转搭载模拟摄像机由驱动盒 BY-2HB03M 和 PCI 电机控制卡驱动控制。在电机旋转部分加装旋转编码器 TRD-NH2500 测量旋转的角度。搭载的模拟摄像机是 1/3in 480 线 Sony CCD 摄像机,微视图像公司 Microview V200 作为 PCI 图像采集卡实时采集真实场景图像。水平与俯仰二维旋转的随动电机控制示意图如图 8.15(b)所示。在客户端采用虚拟现实浸没式头盔显示器 eMagin Z800,图像显示分辨率可达 800×600 像素,绑定罗盘跟踪器实时测量头部旋转角度。

8.4.2 样机设备与操作使用

为提高系统的实用性,实验样机将各功能模块化,使整个系统一体化。在系统安装结构、系统重量和操作实用性上都有所考虑,样机设备如图 8.16 所示。

样机的服务器端重量低于 7kg,客户端重量低于 0.5kg。其中服务器端包括计算机(内置随动控制卡)、控制盒(内置高速转台驱动器)、随动观测器(包括高速转台、角度测量器与图像采集摄像机)。为提高稳定性,控制盒与随动观测器间通过航空插座连接电源、视频传输、电机控制线等多种连线。客户端只有重 227g 的头盔显示器与不足 100g 的角度测量器,形成了真正的瘦客户端。使用过程中,用户

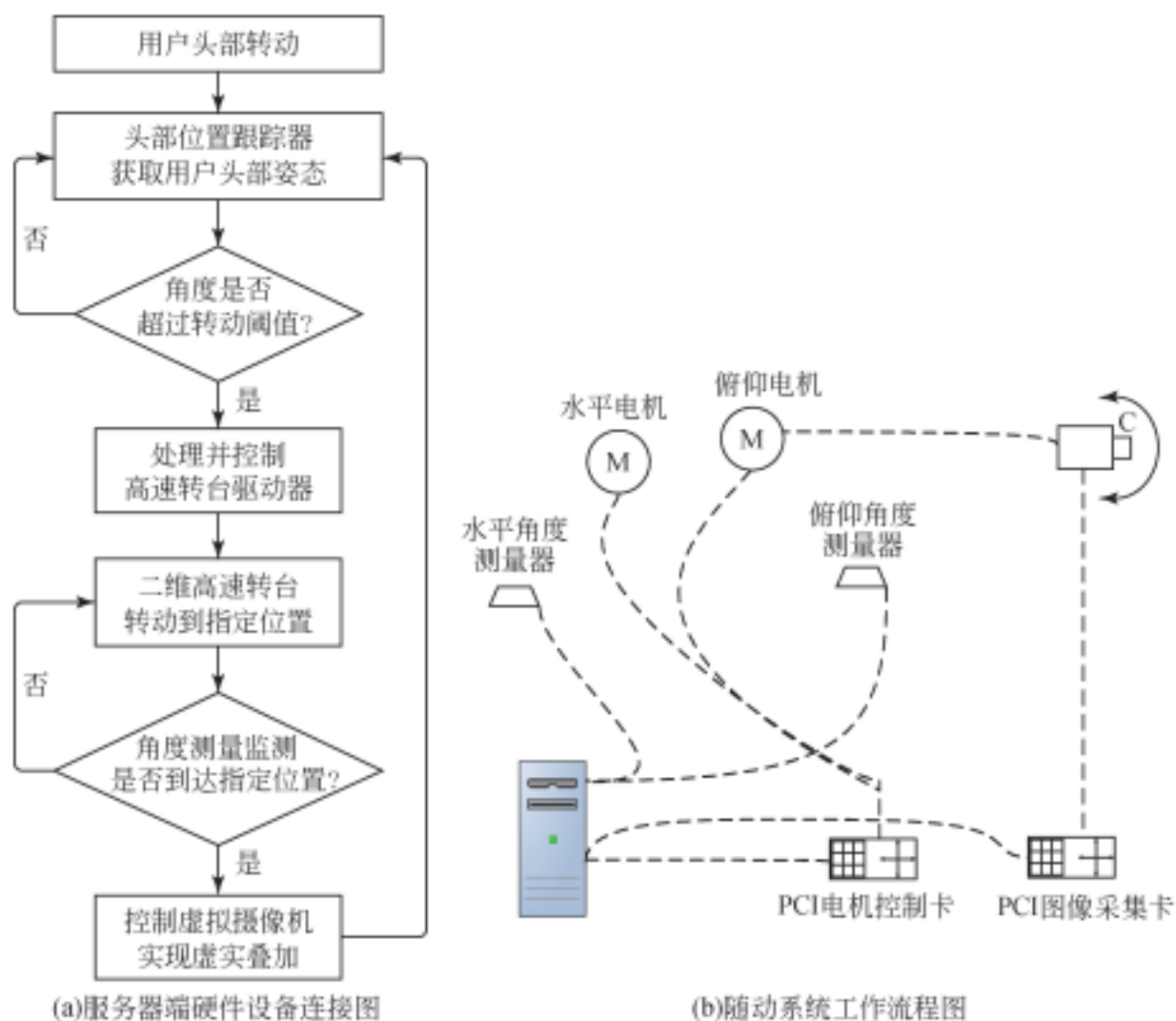


图 8.15 随动式增强现实定点观景器的系统流程图

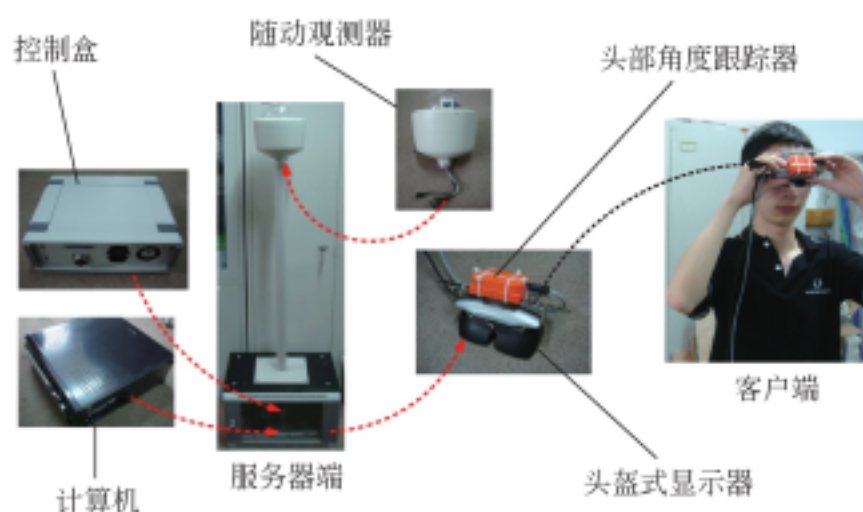


图 8.16 实验样机图

只需手持轻便的用户端设备,原地自由观看四周的景物,跟踪系统即可完成定点增

强现实系统中的角度跟踪注册。系统安装方便,操作简单,说明系统可以满足室外增强现实中对跟踪系统的轻便性与实用性要求。

实验样机在实验室搭建完毕后,于圆明园遗址公园西洋楼景区的大水法遗址处进行实地实验。由于系统自身的设计对于户外常见干扰都有较好的抗干扰性,所以户外实验中系统的各项参数与室内使用时的各项参数无异,没有明显的受干扰情况。图 8.17 是圆明园增强现实三维注册效果。



(a) 圆明园大水法遗址

(b) 圆明园大水法遗址增强现实显示效果

图 8.17 随动式增强现实定点观景器的融合显示效果

(TP-6967.01)

www.sciencep.com

ISBN 978-7-03-044025-9



9 787030 440259 >

销售分类建议：光学工程

定 价：128.00 元