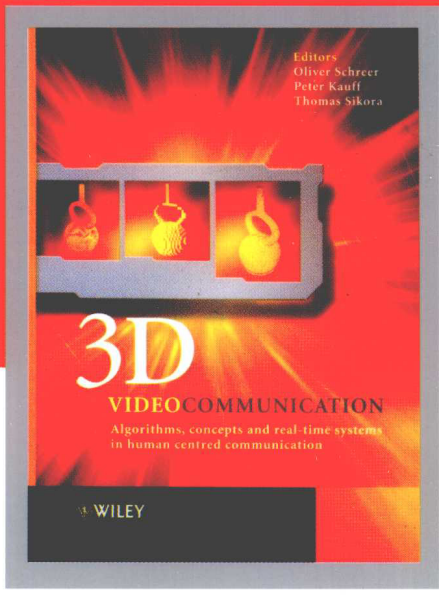
信息技术和电气工程学科国际知名教材中译本系列

WILEY
Publishers Since 1807

3D
Videocommunication

3D视频通信



Oliver Schreer
Peter Kauff
Thomas Sikora
编

戴琼海 曹汛 尔桂花 译

清华大学出版社

信息技术和电气工程学科国际知名教材中译本系列

3D Videocommunication

3D视频通信

Oliver Schreer
Peter Kauff
Thomas Sikora
编

戴琼海 曹汛 尔桂花 译

清华大学出版社
北京

译 者 序

近年来随着多媒体和网络通信技术的飞速发展，基于立体视频通信的应用日益广泛，例如立体视频会议、小组协作系统、虚拟立体电影院等系统在技术上逐步走向成熟，在商业上开始产生效益，并且市场规模还在不断扩大；能够使人们产生“临场”体验的技术不仅应用于航空模拟器、自动虚拟环境系统、网际空间应用以及主题乐园，还开始进入办公室、赛场和家庭。这些应用推动了立体视频的相关研究，包括采集、制作、视图重现、显示、压缩、通信和应用等，成为计算机视觉、计算机图形学、视频处理与通信工程学科的前沿领域和研究热点。而其中的关键是立体采集、重现与显示技术。

本书涵盖了以上所有立体视频相关的研究内容，全面地反映了国际上立体视频通信领域的最新进展和研究成果，深入系统地阐述了三维音视频通信领域中的原理和概念，并从实用角度介绍了各相关技术以及彼此间的联系。因而本书非常适合视觉、图形学、视频处理等领域的研究工作者阅读参考。全书分为四个部分。第 1 部分介绍三维电视、自由视点视频和沉浸式视频会议等重要的三维视频通信应用，以及其中面临的挑战。第 2 部分主要介绍三维音视频处理的理论和概念。第 3 部分重点介绍音视频内容的三维再现。最后一部分讨论与三维数据采集相关的内容。各章最后的参考文献为读者进一步深入钻研提供了方便。原书由国际知名研究机构德国的 Fraunhofer HHI(Heinrich-Hertz-Institute)组织撰写，具有国际权威性。全书体系结构清楚明了，语言流畅，图文并茂。本书的翻译、出版和广泛使用，必将对我国立体视频研究的发展产生深远的影响，对我国在三维音视频处理技术领域跟上国际步伐并占据领先地位起到重要作用。

本书由清华大学戴琼海教授负责并承担主要翻译工作，参与本书翻译的人员还有尔桂花、谢旭东、曹汛、邵航、徐枫、吴城磊、武迪、汪启扉、李冠楠、彭义刚、魏宇平、刘继明、钱俊彦、柯刚凯、路瑶、冯晓端、孙友婷、陆峰、黎政和杨李等。本书涉及较多当前国际最新的名词、概念、方法和技术，由于译者水平有限，译文中可能存在一些不妥之处，敬请专家和广大读者批评指正。

译 者

2010 年 6 月

引 言

Oliver Schreer, Peter Kauff, Thomas Sikora

沉浸式媒体向远程通信应用的发展趋势将越来越受关注。多媒体压缩、显示技术的巨大进步和不断增长的用户带宽为这些服务在未来的应用创造了条件。人们普遍认为目前向着沉浸式媒体发展的趋势将对人们的日常生活产生巨大的影响。

在多媒体应用中，能够使人们产生“沉浸”或“临场”体验的技术不仅用于航空模拟器、自动虚拟环境(CAVE)系统、网际空间应用、主题公园或者 IMAX 剧院等，它还将出现在办公室、比赛场和家庭等场合，并且具有进一步提高人们生活质量的潜力。

最近几年，我们已经从以下几方面感觉到这种趋势：

- 视频会议在各种商业活动中变得越来越有吸引力。今天的视频会议可以增强新兴全球化市场中的分布协作能力，因而也被认为是在做决策过程中能产生高回报的投资行为。目前的高端视频会议系统已经具备远程显示的能力，这使得交流环境变得尽可能的自然。这种商业模式将从沉浸系统未来的发展中受益，而后者能够在场景重建方面提供更好的真实感。
- 团队协作系统的市场发展迅猛。如今最早的同步协作工具正在市场中销售，它们需要适应公司在成本、创新、生产和研发链条中日益激烈的竞争。这些工具中的大部分仍然依靠屏幕共享的方案，因此遇到了在协作成员间缺乏面对面自然交流的问题。新兴的远程沉浸系统采用具有直接交互和交流能力的协同虚拟环境(CVE)，因此能够克服传统团队协作软件的局限性。
- 在娱乐部分，我们开始注意到对在体育场、电影院、演播大厅以及大型群众聚会场所的活动进行现场直播具有可观的经济效益。世界上许多研发部门在预计或调查这方面的应用，例如 e-剧场，d-电影院，家庭剧院和沉浸式电视。在未来几十年，电视、计算机游戏、体育比赛、现场新闻或者我们所熟知的电影都必将要发展为全新的沉浸式应用，以满足消费者的需求。

现如今，在我们讨论的范围之外，很难预测沉浸式媒体未来的发展。但是上面提到的实例已经表明，未来采集、传输和使用媒体信息的一些方式将要改变。由于价格下降和质量提高，宽屏显示、基于音视频的三维(3D)场景再现和直接人机交互将在日常生活中使用，特别是在办公和家庭娱乐中变得越来越重要。沉浸式系统将脱离试验阶段，沉浸式技术将在商业和娱乐业逐渐普及。这对消费者和商业部门及其价值链条的影响将是巨大的。

音视频通信从二维到三维的发展被普遍地认为是未来应用中的关键组成部分。这个领域里的技术挑战是多方面的。它包括从音视频的高质量三维分析和任意视点及声音合成到三维音视频编码的各个方面。为了实现这些应用，理解实时操作、系统结构构成和网络特征是必需的。许多服务的引入要求有新的三维视听数据表示标准和编码标准。由于这些服务将改变人们消费和与媒体应用交互的方式，在研究中考虑人的因素将非常重要。我们最终的目标是实现高质量服务和用户的充分认可。现在研究机构的方向也大都集中在这类以用

户为中心的通信方面。

本书全面讲述了在奇妙的三维音视频通信领域中的原理和概念。从实用角度一步步地介绍了涉及应用和服务发展的概念、器件、技术和遇到的各种挑战。对三维音视频通信领域感兴趣的研究人员和学生会发现这是一本有价值的参考书，它包含了对当前发展水平的全面综述。来自工业界的工程师会发现这本书在构思和建立原创系统时很有价值。

这本书分为四个部分。第 1 部分介绍了三维电视、自由视点视频和沉浸式视频会议这些三维视频通信的重要应用，及其领域中人们面临的挑战。第 2 部分包括三维音视频处理的理论部分。沿着普通的信号处理逻辑思路。第 3 部分讲述了音视频内容的三维再现。最后一部分，讨论了三维数据传感器的若干方面。

第 1 部分“3D(立体)视频通信的应用”讲述了三维视频通信的发展水平、挑战及潜力。这部分从 W. A. Ijsselsteijn 写的“远程呈现的历史”开始，阐述了这个新领域研究和发展的基础及其产生的原因，历史性地回顾了远程呈现是如何兴起的。接下来的一章是 C. Fehn 写的“三维电视广播”，它详细地描述了这一在三维视频通信领域中的关键应用。本章描述了电视的历史和下一代电视系统的概念，详述了点对点的立体视频结构链。这些新兴的技术也对内容生成和后期制作产生重大影响。O. Gau 写的“内容创造和后期制作中的三维技术”讨论了这个领域新的趋势和方向。三维音视频通信的最终目标是提供给用户一个自由视点，由 M. Tanimoto 写的“自由视点系统”阐述了这个领域的挑战并展示了实验系统的初步结果。沉浸式视频会议引入了双向三维视频通信的概念。P. Kauff 和 O. Schreer 写的“沉浸式视频会议”描述了当前远程呈现系统的发展水平和历史，概述和讨论了应用在沉浸式视频会议系统原型中的设想方法和沉浸式技术。

本书的第 2 部分讲述了如何显示和处理三维音视频数据的问题。主要目的是提供给读者一个完整的涉及音视频处理所有方面的综述。S. Ivekovic，A. Fusiello 和 E. Trucco 写的“多视图几何基础”概述了从一个三维场景向单个摄像机成像过程的相关理论。这章主要讨论针孔摄像机模型，解释使用对极几何的立体摄像机系统和基于三焦张量的三视图几何。其中也包含校准和重建的若干方面。这章是下一章的基础。

立体显示分析可以从同一场景的若干相机图像得到隐含的深度信息。由 N. Atzpadin 和 J. Mulligan 写的“立体分析”一章详述了目前在立体视频处理中用到的基于两个或三个摄像机的方法。讨论了视差分析面临的基本问题，对不同算法进行了综述。从一个场景或实物的多个视图我们可以得到更为精细的三维模型，P. Eisert 写的“三维模型体的重建”主要是介绍这个领域的相关算法，这对许多新的三维多媒体服务非常重要。

三维处理中下一个重要的步骤涉及渲染新的视图。第 9 章是由 R. Koch 和 J.-F. Evers-Senne 写的“视图合成和渲染的方法”，它对现有绘制算法进行了分类。在这个新的研究领域中针对部分无几何信息的算法和具有几何信息(包括具有隐含的几何信息)的算法进行了广泛的讨论。

M. Schwab 和 P. Noll 写的“三维音频采集和分析”一章主要内容是介绍人类语音三维的获取过程。这包括回声控制，降噪和三维音频的声源定位，这些都有力地支持音视频场景内容的渲染。

标准化研究对来自不同厂商的系统的兼容是一个重要的课题。A. Smolic 和 T. Sikora 写的第 11 章由“编码及标准化”概述了频繁用于音频、图像和视频存储及传输的基本编码

方案，也讨论了一些国际编码标准，如 ITU、MPEG-2/4 和 MP3。目前有关三维视频通信标准的最新进展是 MPEG-4 AdHoc 工作组针对三维音视频的工作。这一章，给读者提供了这些活动的详细综述。

第 3 部分包含了音视频内容三维再现的各个方面。由 W. A. Ijsselsteijn、P. J. H. Seuntiens 和 L. M. J. Meesters 写的“三维显示中人的因素”开始，作者讨论了立体视频的若干方面。对开发受人欢迎的三维视频系统来说，人的因素是非常重要的。此章介绍了人类深度感知的基本知识，这个领域的知识是立体视频的基础，同时讨论了立体感觉再现原理及其对三维视频图像质量的影响。S. Pastoor 写的“三维显示”讨论了现有三维显示的核心技术。在现有的显示系统和产品的背景下描述和讨论了半自动及全自动三维视频。

结合混合实境应用，头盔显示器(HMD)对真实场景中显示虚拟内容起着重要的作用。这个领域的发展在由 S. Pastoor 和 C. Conomis 写的第 14 章“混合实境显示”中讲述。在大体上描述了混合实境显示所遇到的挑战和在这个领域中人类空间视觉的几个特征之后，本章给出了对不同技术和系统的全面叙述。

即使在视觉占主导地位的情况下，听觉也可以帮助分析环境和创造沉浸感。正确的或者至少大体正确的空间音频再现成为一个重要的主题。在第 15 章，T. Sporer 和 S. Brix 讲述了“空间音频和三维音频渲染”的基础。

第 4 部分介绍了三维数据传感器领域。现行的技术能够创造具有高精确度的三维精细模型。在第 16 章，J. G. M. Goncalves 和 V. Sequeira 写的“基于传感器的深度检测”中，他们讲述了现有的各类图像获取技术，讨论了这些方法的局限性、精度和校准方法。

本书以 Y. Abdeljaoued、D. Marimonisanjvan 和 T. Ebrahimi 写的第 17 章“混合实境的跟踪和用户界面”结束。作者讨论了目标跟踪问题，目的是使虚拟场景与真实场景的目标精确匹配。此外，本章强调了交互技术的重要性，这种技术可用于产生令人信服的混合实境系统。

目　录

第 1 部分　3D(立体)视频通信的应用

第 4 部分 三维数据传感器

第 1 部分
3D(立体)视频通信的应用

1 远程呈现的历史

Wijnand A. Ijsselsteijn
Eindhoven University of Technology, Eindhoven, The Netherlands

1.1 绪论

远程呈现(telepresence),这个词第一次以远程操作的背景在 1979 年 Marvin Minsky (在他的朋友 Pat Gunkel 建议下)提出的一个大胆的基金研究计划中使用,这个计划是关于远程人工操作能源和生产系统的,其内容是他在 1980 年发表的一篇经典论文的核心。远程呈现是指通过人和系统进行人机交互达到模拟一个人在现场进行操作的感觉,也就是说,用户凭借系统的人机交互界面,通过个人的操作获得相应的感观上的反馈,从而得到逼真的现场感觉。

呈现的概念在较早的戏剧表演文献中曾经讨论过,因为在演出中,演员需要具有"舞台表现力"(用来衡量演员舞台表演的表现力和可信度)。Bazin(1967) 也讨论过这种与摄影和电影相关的呈现。他写道:

> 呈现,很自然地,是根据时间和空间定义的。"呈现某人"一般是指我们认为他和我们同时存在,而且这种存在能够在我们感知的范围之内——在电影院通过我们的视觉感知,而在广播中就要通过我们的听觉感知。在摄影技术以及后来的电影技术被发明之前,造型艺术(尤其是肖像画)是唯一处于真实的物理上的存在和不存在之间的中间物。(引自 Bazin(1967),96 页,最初于 1951 年在 Esprit 上发表)

Bazin 注意到在剧院,演员和观众有一个相互呼应的关系,能够彼此在同一时间和空间内对对方的行为作出反应。但是对于电视以及其他一切广播媒介来说,这种交互性在一个方向上是不完全的,这在现场和非现场之间增加了一种新的形式,被称为"伪现场"。Bazin 叙述到:

> 观众看见了,但是没有被看见。这里没有任何的反馈……然而,这种不在现场不是真正意义上的不在现场。通过摄像机,电视演员有一种被成千上万双眼睛关注和成千上万只耳朵聆听的感觉。(引自 Bazin(1967)97 页脚注)

在一个真实的物理空间中和其他人一起存在并交流的概念可以追溯到 Goffman(1963) 的工作,他使用了协同呈现(co-presence)的概念来解释个人对于其他人的感知。

> 协同呈现的条件在较少变化的环境中才能满足。人们必须充分接近,以便能够清楚地感受到他们在做什么,包括各自对于其他人的体验,以及对方能够明白自己的感受。(引自 Goffman(1963),17 页)

伴随这种交互和反馈而来的是个人如何向其他人呈现他们自己这一问题。然而,这里我们要注意,Goffman 仅仅将协同呈现的概念用在"实际"物理空间中的社会交互中。在现