

附件 2

5G 高新视频—沉浸式视频技术白皮书（2020）

国家广播电视总局科技司

2020 年 8 月

前 言

当前，移动信息技术飞速发展，5G 技术已经成为国际通信科技巨头竞争的新焦点，世界各国纷纷将 5G 建设视为重要目标。5G 技术具有“大带宽、低时延、广连接”等特点，在 5G 众多应用场景中，视频被公认为是 5G 时代最重要和最早开展的业务。在 5G、超高清、虚拟现实等新兴技术催生下，广电行业传统视听内容的生产和传播，已难以满足新时代人民群众对美好视听的新需求。国家广播电视总局顺应技术革命浪潮，抢抓 5G 发展机遇，深入推进 5G 条件下广播电视供给侧结构性改革，推动构建 5G 视频新业态，提出了“5G 高新视频”概念。

5G 高新视频是指 5G 环境下具有“更高技术格式、更新应用场景、更美视听体验”的视频。其中，“高”是指视频融合 4K/8K、3D、VR/AR/MR、高帧率（HFR）、高动态范围（HDR）、广色域（WCG）等高技术格式；“新”是指具有新奇的影像语言和视觉体验的创新应用场景，能够吸引观众兴趣并促使其产生消费。在 5G 环境下，广电行业将发挥在视音频内容创意、生产、制作和传播方面的丰富资源和天然优势，以文化创意为牵引，以科技创新为支撑，培育孵化 5G 高新视频的新技术、新应用、新产品、新模式、新业态和新消费，为观众带来更美的视听体验，更好地满足人民群众日益增长的精神文化新需求和新期待。

沉浸式视频作为高新视频业态的重要组成部分，是指一种采用裸眼观看方式获得身临其境感受，呈现画面覆盖人眼至少 120°（水平）×70°（垂直）视场角的视频系统及具备三维声的音频系统。沉浸式视频通过播放器、投影幕或 LED 自显示屏、多声道扬声器、播控系统构建出超大视角、超高沉浸感的视听呈现系统，使观众能够同时获得周围多方位的视听信息，带来单一平面视频无法展示出的强大沉浸感，让观众真正有身临其境的感觉，应用场景丰富且形式多样。

为更好地推动沉浸式视频的发展，引导和规范沉浸式视频行业应用，国家广播电视总局科技司下达任务，由中广电广播电影电视设计研究院牵头，组织相关机构、企业等，开展沉浸式视频技术研究工作，编制《5G 高新视频-沉浸式视频技术白皮书（2020 版）》（以下简称白皮书）。

本白皮书分析了沉浸式视频的发展现状，探讨了沉浸式视频的核心要素、关键技术，梳理了沉浸式视频应用场景，总结了沉浸式视频应用面临的挑战，提出了沉浸式视频端到端解决方案。希望本白皮书能为各单位应用沉浸式视频提供参

考和借鉴，为推动广播电视行业转型升级注入新动能、激发新活力，促进广播电视和网络视听行业高质量创新性发展。

本白皮书指导单位：国家广播电视总局科技司

本白皮书牵头单位：国家广播电视总局中广电广播电影电视设计研究院

本白皮书参与单位：国家广播电视总局广播科学研究院、北京广播电视台、北京天文馆、北京邮电大学、四川传媒学院、华为技术有限公司、北京歌华有线电视网络股份有限公司、北京万达文旅规划设计院有限公司、北京未来媒体科技股份有限公司、深圳岚锋创视网络科技有限公司、上海腾斯帝信息科技有限公司、苏州光魔软件科技有限公司、北京澜景科技有限公司、北京汇融视景科技有限公司、杜比实验室、Fraunhofer 研究所、北京东方佳联影视技术有限公司、中科雷欧尼斯（北京）信息技术有限公司、天津凡创科技有限公司、FullDome Pro、安恒集团、利亚德光电股份有限公司、南京洛普股份有限公司、深圳光峰科技股份有限公司、北京鸿合科技股份有限公司、紫科智控（北京）智能科技有限公司、北京光影创意科技有限公司。

本白皮书编写指导：孙苏川、关丽霞

本白皮书牵头人：林长海

本白皮书主要起草人：周耀平、欧阳玥、王建雄、孙松林、周屹、宋宇莹、王家福、林霖、李林、李博、周宏、冷海敏、文殿杰、王倩、蒋昭洁、马云龙、张晓敏、刘洁心、赵珊珊、张健、王虽然、付美霞

目 录

| | |
|---------------------|----|
| 缩略语..... | 5 |
| 1. 概述..... | 9 |
| 1.1 沉浸式视频的定义..... | 9 |
| 1.2 沉浸式视频的发展现状..... | 10 |
| 1.3 沉浸式视频的应用机遇..... | 10 |
| 2. 沉浸式视频核心要素..... | 12 |
| 2.1 画面包围感..... | 12 |
| 2.1.1 球幕..... | 13 |
| 2.1.2 环幕..... | 13 |
| 2.1.3 沉浸屋 CAVE..... | 14 |
| 2.1.4 其他异形呈现形式..... | 14 |
| 2.2 画质..... | 15 |
| 2.2.1 分辨率..... | 15 |
| 2.2.2 动态范围..... | 16 |
| 2.2.3 色域..... | 17 |
| 2.2.4 帧率..... | 18 |
| 2.2.5 色深..... | 18 |
| 2.3 沉浸式声音..... | 19 |
| 2.4 沉浸式视频核心要素表..... | 19 |
| 3. 沉浸式视频关键技术..... | 20 |
| 3.1 全景视频拍摄技术..... | 20 |
| 3.1.1 多机组合拍摄..... | 20 |
| 3.1.2 全景相机拍摄..... | 21 |
| 3.2 视频图像缝合技术..... | 22 |
| 3.2.1 模板缝合..... | 22 |
| 3.2.2 光流缝合..... | 22 |
| 3.3 三维图像映射技术..... | 23 |
| 3.3.1 鱼眼映射..... | 23 |
| 3.3.2 等距圆柱映射..... | 24 |
| 3.3.3 正六面体映射..... | 25 |
| 3.3.4 等角方块映射..... | 25 |
| 3.4 计算机图像制作技术..... | 26 |
| 3.5 三维声技术..... | 27 |
| 3.5.1 基于声道的技术..... | 27 |
| 3.5.2 基于对象的技术..... | 27 |
| 3.5.3 基于场景的技术..... | 28 |
| 3.6 编码及传输技术..... | 29 |
| 3.6.1 沉浸式视频的码流..... | 29 |
| 3.6.2 沉浸式视频的编码..... | 29 |
| 3.6.3 沉浸式视频的传输..... | 30 |
| 3.7 媒体播控技术..... | 31 |

| | |
|------------------------|----|
| 3.7.1 视频解码播放..... | 31 |
| 3.7.2 投影拼接融合..... | 31 |
| 3.7.3 自动校准技术..... | 32 |
| 3.7.4 透视校正技术..... | 32 |
| 3.7.5 环境特效控制..... | 32 |
| 3.8 投影呈现技术..... | 33 |
| 3.8.1 投影几何校正技术..... | 33 |
| 3.8.2 异形边沿融合技术..... | 34 |
| 3.8.3 球面投影亮度均一化技术..... | 34 |
| 3.9 LED 屏呈现技术..... | 35 |
| 3.9.1 异形无缝拼接技术..... | 35 |
| 3.9.2 透声及吸声技术..... | 36 |
| 3.9.3 图像球形矢量变换技术..... | 37 |
| 4. 沉浸式视频端到端解决方案..... | 38 |
| 4.1 总体工艺架构..... | 38 |
| 4.1.1 直播工艺流程..... | 39 |
| 4.1.2 点播工艺流程..... | 40 |
| 4.2 沉浸式视频的拍摄..... | 41 |
| 4.2.1 拍摄设备的选择..... | 41 |
| 4.2.2 与传统视频拍摄的差异..... | 42 |
| 4.2.3 常用拍摄方式..... | 44 |
| 4.3 沉浸式视频的制作..... | 45 |
| 4.3.1 实拍类视频的制作..... | 45 |
| 4.3.2 计算机图像制作..... | 47 |
| 4.4 沉浸式视频的传输..... | 47 |
| 4.4.1 有线电视网络..... | 47 |
| 4.4.2 5G 无线网..... | 49 |
| 4.4.3 互联网..... | 50 |
| 4.5 沉浸式视频呈现..... | 51 |
| 4.5.1 视频播放系统..... | 51 |
| 4.5.2 投影系统..... | 52 |
| 4.5.3 动感系统..... | 53 |
| 4.5.4 周边系统..... | 53 |
| 4.6 沉浸式视频的声音方案..... | 54 |
| 4.6.1 基于对象的三维声方案..... | 55 |
| 4.6.2 基于场景的三维声方案..... | 56 |
| 4.7 呈现空间的音视频指标要求..... | 57 |
| 4.7.1 亮度..... | 57 |
| 4.7.2 对比度..... | 57 |
| 4.7.3 声道数量..... | 57 |
| 4.7.4 扬声器布置..... | 58 |
| 4.7.5 声压级..... | 58 |
| 4.7.6 频率响应..... | 58 |
| 4.8 呈现空间的环境指标要求..... | 58 |

| | |
|-----------------------|----|
| 4.8.1 声学环境要求..... | 58 |
| 4.8.2 照明技术要求..... | 60 |
| 4.8.3 暖通技术要求..... | 61 |
| 4.9 沉浸式视频系统小结..... | 61 |
| 4.9.1 拍摄制作..... | 61 |
| 4.9.2 编码..... | 62 |
| 4.9.3 传输..... | 62 |
| 4.9.4 终端呈现..... | 63 |
| 5. 沉浸式视频的应用场景..... | 65 |
| 5.1 重大活动应用..... | 66 |
| 5.1.1 项目概述..... | 66 |
| 5.1.2 整体架构..... | 66 |
| 5.1.3 部署方案..... | 67 |
| 5.1.4 应用效果..... | 67 |
| 5.2 游艺场馆应用..... | 68 |
| 5.2.1 项目概述..... | 68 |
| 5.2.2 整体架构..... | 68 |
| 5.2.3 部署方案..... | 68 |
| 5.2.4 应用效果..... | 69 |
| 5.3 展陈场馆应用..... | 69 |
| 5.3.1 项目概述..... | 69 |
| 5.3.2 整体架构..... | 70 |
| 5.3.3 部署方案..... | 70 |
| 5.3.4 应用效果..... | 71 |
| 5.4 教育培训应用..... | 71 |
| 5.4.1 项目概况..... | 71 |
| 5.4.2 整体架构..... | 72 |
| 5.4.3 部署方案..... | 72 |
| 5.4.4 应用效果..... | 73 |
| 5.5 家庭终端应用..... | 73 |
| 5.5.1 项目概述..... | 73 |
| 5.5.2 整体架构..... | 74 |
| 5.5.3 部署方案..... | 74 |
| 5.5.4 应用效果..... | 75 |
| 5.6 虚拟拍摄应用..... | 75 |
| 5.6.1 项目概述..... | 75 |
| 5.6.2 整体架构..... | 75 |
| 5.6.3 部署方案..... | 76 |
| 5.6.4 应用效果..... | 77 |
| 6. 沉浸式视频应用的挑战与前景..... | 78 |
| 6.1 存在的问题..... | 78 |
| 6.1.1 内容制作域..... | 78 |
| 6.1.2 传输分发域..... | 79 |
| 6.1.3 终端显示域..... | 79 |

| | |
|-----------------|----|
| 6.2 发展前景展望..... | 80 |
| 参考文献..... | 81 |

缩略语

| 缩略语 | 英文名称 | 中文全称 |
|--------|--|----------------|
| AR | Augmented Reality | 增强现实 |
| ARQ | Automatic Repeat-reQuest | 自动重传请求 |
| AVS | Audio Video Coding Standard | 音视频编码标准 |
| CAVE | Cave Automatic Virtual Environment | 沉浸屋（洞穴状自动虚拟系统） |
| CBA | Channel-based Audio | 基于声道的音频 |
| CDN | Content Delivery Network | 内容分发网络 |
| CMTS | Cable Modem Terminal Systems | 电缆调制解调器终端系统 |
| CMS | Content Management System | 媒资管理系统 |
| CG | Computer Graphics | 计算机图形学 |
| CMP | Cube Map Projection | 正六面体映射 |
| CUDA | Compute Unified Device Architecture | 通用并行计算架构 |
| DD+ | Dolby Digital Plus | 杜比数字+ |
| DTS | Digital Theater System | 数字影院系统（公司名） |
| DCI | Digital Cinema Initiatives | 数字电影倡导联盟 |
| DIGSS | Digital Immersive Giant Screen Specifications | 沉浸式数字影院规范 |
| DLP | Digital Light Processing | 数字光处理技术 |
| DOCSIS | Data Over Cable Service Interface Specifications | 有线电视双向数据传输系统 |
| DMX512 | Digital Multiple X 512 | 一种数据调光协议 |
| EAC | Equi-Angular Cubemap Projection | 等角方块映射 |
| eMBB | Enhanced Mobile Broadband | 增强移动宽带 |
| eMTC | Enhanced Machine Type Communications | 增强机器类通信 |
| E2E | End-to-End | 端到端 |
| ERP | Equi-Rectangular Projection | 等距圆柱映射 |

| | | |
|--------|--|--------------------------------|
| FOV | Field of View | 视野 |
| FPS | Frames Per Second | 每秒的帧数 |
| FTTx | Fiber To The x | 各种光纤通讯网络的总称 |
| GPU | Graphics Processing Unit | 图形处理器 |
| HDR | High Dynamic Range | 高动态范围 |
| HFR | High Frame Rate | 高帧率 |
| HINOC | High-performance Network over Coax | 高性能同轴网络 |
| HLG | Hybrid Log Gamma | 混合对数伽马 |
| HLS | HTTP Live Streaming | 基于 HTTP 的流媒体网络传输协议 (苹果公司技术) |
| HOA | Higher Order Ambisonic | 高阶声场信号 |
| HFC | Hybrid Fiber-Coaxial | 混合光纤同轴网络 |
| IMAX | Image Maximum | 一种巨幕电影放映系统 |
| IPS | International Planetarium Society | 国际天文馆协会 |
| ITU | International Telecommunication Union | 国际电信联盟 |
| I/O | Input/Output | 输入/输出 |
| IPTV | IPTV | 交互式网络电视 |
| lx | Lux | 勒克斯(照度单位) |
| JOC | Joint Object Coding | 传输对象音频的编码方式(杜比公司技术) |
| LED | Light Emitting Diode | 发光二极管 |
| MEC | Mobile Edge Computing | 移动边缘计算 |
| MPEG | Moving Picture Experts Group | 动态图像专家组 |
| MP4 | Moving Picture Experts Group 4 | 一种用于音频、视频信息的压缩编码标准 |
| MPEG-H | Moving Picture Experts Group-H | 动态图像专家组所开发的一组标准 |
| MR | Mixed Reality | 混合现实 |
| MVC | Multiview Video Coding | 多视角视频编码 |

| | | |
|-------|--|-----------------------|
| NTSC | National Television Standards Committee | 美国国家电视标准委员会 |
| NR | Noise Rating Number | 噪声评价曲线 |
| OBA | Object-based Audio | 基于对象的音频 |
| ODN | Optical Distribution Network | 光分配网 |
| OTN | Optical Transport Network | 光传送网 |
| OTT | Over The Top | 通过公共网络向用户提供内容分发业务（电视） |
| P2MP | Point-to-Multipoint Connection | 点到多点连接 |
| PQ | Perceptual Quantization | 感知量化 |
| PCB | Printed Circuit Board | 印制电路板 |
| PON | Passive Optical Network | 无源光纤网络 |
| QoS | Quality of Service | 服务质量 |
| RIST | Reliable Internet Stream Transport | 可靠的互联网流传输协议 |
| RTMP | Real Time Messaging Protocol | 实时消息传输协议 |
| RTT | Render to Texture | 渲染到纹理 |
| SDR | Standard Dynamic Range | 标准动态范围 |
| SBA | Scene-based Audio | 基于场景的音频 |
| SLA | Service Level Agreement | 服务级别协议 |
| SMPTE | Society of Motion Picture and Television Engineers | 电影和电视工程师协会 |
| SRT | Secure Reliable Transport | 安全可靠传输协议 |
| TS | Transport Stream | 用于传输的数据流 |
| UDP | User Datagram Protocol | 用户数据报协议 |
| URLLC | Ultra-reliable Low Latency Communications | 超高可靠超低时延通信 |
| VR | Virtual Reality | 虚拟现实 |
| VOD | Video on Demand | 视频点播系统 |
| WCG | Wide Color Gamut | 广色域 |

| | | |
|-----|----------------------------------|------|
| WDM | Wavelength Division Multiplexing | 波分复用 |
| XR | Extended Reality | 扩展现实 |

1. 概述

人的视觉系统是一个高度精密的光学系统，双眼立体视觉的重要基础是空间深度感。达芬奇在文艺复兴时期最早指出，相比于平面透视，采用曲面透视方法，图像视点不会出现扭曲，如图 1-1 所示。根据这一原理，影视制作人通过拍摄或渲染与曲面屏幕几何图形相匹配的透视影像，创造出一种沉浸式的视觉体验，模仿人们在现实世界中的视觉体验。曲面屏幕能够真实地再现现实世界中的视觉体验，现今技术发展阶段，最符合这种视觉特点的表现形式是球幕，沉浸式视频利用了这一原理。

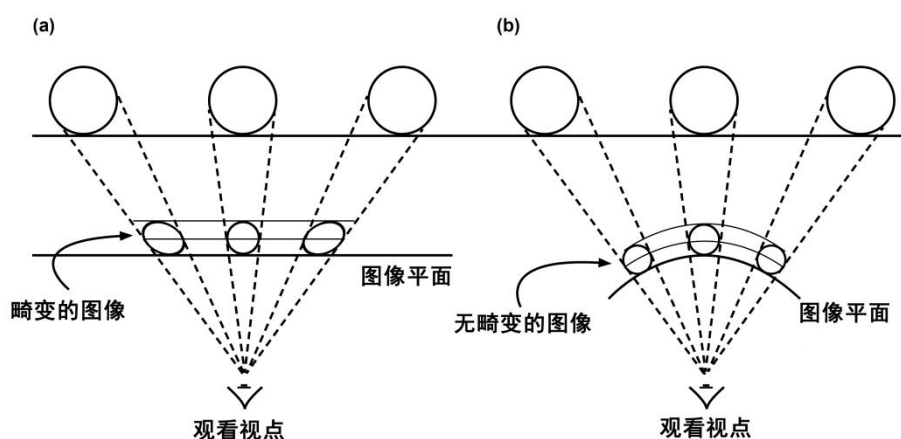


图 1-1 (a) 平面透视

(b) 曲面透视

1.1 沉浸式视频的定义

沉浸式视频指的是一种采用裸眼观看方式获得身临其境感受，呈现画面覆盖人眼至少 120° （水平） $\times 70^\circ$ （垂直）视场角的视频系统及具备三维声的音频系统。

沉浸式视频通过视频、音频及特效系统，构建具备大视角、高画质、三维声特性，具备画面包围感和沉浸式声音主观感受特征的视听环境，使观众能够在所处位置同时获得周围多方位的视听信息，让观众体验到单一平面视频无法实现的高度沉浸感，呈现形式包括但不限于球幕、环幕、沉浸屋 CAVE 等异形显示空间。

1.2 沉浸式视频的发展现状

沉浸式视频源于特种电影领域。心理学研究表明，就虚拟场景而言，当观察者的余光被场景全部包围后，观察者会自然地产生一种身临其境的感觉。电影制作人利用人眼视觉这一特点，采用大尺寸屏幕以及播放处理设备的图形建构能力，将大视场视景显示在观众面前，从而能够使观众完全沉浸在虚拟场景之中。随着电影技术不断创新，出现了 3D 电影、动感电影、特效电影、球幕电影等不同类型的电影形式，呈现形式也由单一的平面幕发展成弧幕、环幕、球幕等。

数字技术时代，电影与电视行业出现融合，沉浸式视频得到进一步发展。随着超高清 4K/8K 技术日趋成熟以及三维声的推广应用，沉浸式视频也进入了一个新阶段。相对于 2K 高清视频标准，超高清视频在高分辨率、高帧率、高色深、广色域、高动态范围上实现了突破：4K/8K 高分辨率为观众提供了更为丰富的画面层次和更为精致的画面细节；高帧率技术能够提升影像的细腻度和流畅感；高色深、广色域提升了画面颜色的丰富程度；高动态范围技术的使用大大地提高了画面的对比度，能更好地展现亮部和暗部的细节；技术标准上的全面提升将革新观众的视听体验，呈现更加逼真的视听场景。在三维声方面，杜比全景声、MPEG-H、AVS2-P3 和 DTS-X 等技术的应用，将音频扩展到三维空间，进一步强化沉浸式视频的空间表现力和临场感，为用户带来了全新的音频体验。

在沉浸式视频呈现方面，投影融合技术已趋于成熟，高分辨率、高亮度、广色域的工程投影机已大规模使用在球幕影院、展览展示、大型演出及光影秀中；各种形式的 LED 自显示屏也越来越多地应用到影院和游乐场所的环幕、球幕中。

沉浸式视频通过多媒体艺术、装置艺术、人机交互、VR/AR/MR 等技术融合，将实体空间营造成沉浸式场景，在游艺、消费领域的应用越来越广泛；同时，沉浸式视频还用于一些大型的虚拟仿真项目，如虚拟战场仿真、数字城市规划、三维地理信息系统等大型场景仿真环境。

1.3 沉浸式视频的应用机遇

随着用户收看体验日趋个性化和多样化，特别是内容丰富化的需求，沉浸式视频在家庭端的应用将出现更为广阔的空间，催生更具沉浸感的家庭影院终端。

基于超大拼接屏、球幕屏或环幕屏的沉浸式视频呈现设备或装置，将呈现更高清晰度和更大的画面；结合三维声效果，可给予观众身临其境的视听感受。快速产业化应用的 5G 技术与沉浸式视频的融合，更会衍生出新的使用模式和场景。

沉浸式视频对传输带宽提出更高的要求。在某些存在实时性、移动性要求的应用场景中，如大型赛事/活动/事件直播、远程现场实时展示等，4G 网络已无法满足 4K、8K 视频传输对网络流量、回传时延等技术指标的要求，5G 网络以其良好的承载力正成为解决该场景需求的有效手段。利用 5G 大带宽、低时延特性可以对高质量的视音频数据进行实时采集，实时传输至后端信息处理中心，通过服务端的信息处理分析，最后分发给用户，高保真还原每一个细节，让用户在沉浸式的呈现终端得到身临其境的视听体验。

此外，沉浸式视频与高新视频的其他应用，如 VR 视频，既存在一定的技术共通性，也存在一定区别：沉浸式视频更偏向于对电影和影院观影方式的扩展，而 VR 更偏向于个人独立观影体验的扩展。部分题材的 VR/360° 全景视频可以直接或通过截取、变换等手段供沉浸式视频使用；针对全球幕摄制的沉浸式视频理论上也可以直接在 VR 体验场景中播放；视频制作的相关技术手段也是相通的。但是，沉浸式视频更强调利用屏或幕构成的沉浸式空间，而 VR 更强调的是面向个体的虚拟全景视野，这使得二者在视频题材、剧本、拍摄手段等方面并不能完全通用，画面和声音品质要求不同，且在编解码和传输处理等具体技术应用方面也有明显差异。

2. 沉浸式视频核心要素

沉浸式视频的核心要素包括以下三个方面：画面包围感、画质、沉浸式声音。

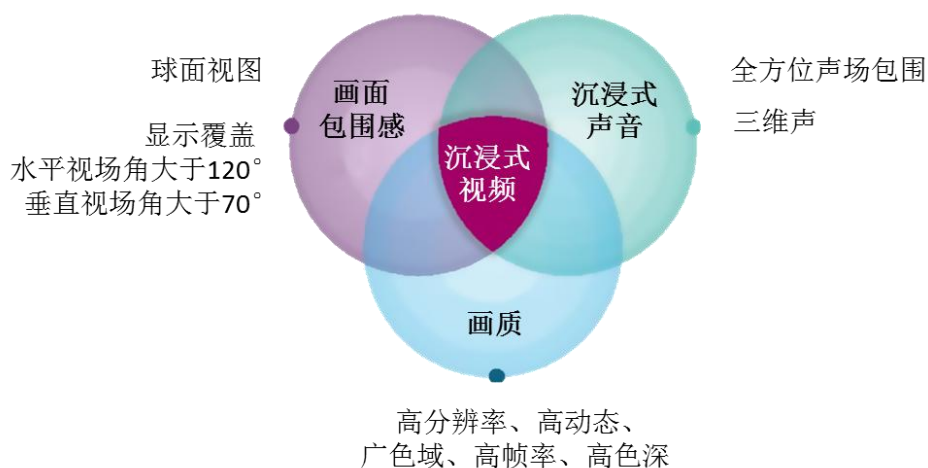


图 2-1 沉浸式视频核心要素

2.1 画面包围感

画面包围感是沉浸式视频的核心要素。根据人眼视觉特性，当影像能更多地覆盖人的视野，形成画面在人眼中的水平视场角大于 120° ，垂直视场角大于 70° ，就可以使人感受到一种身临其境的沉浸式效果。

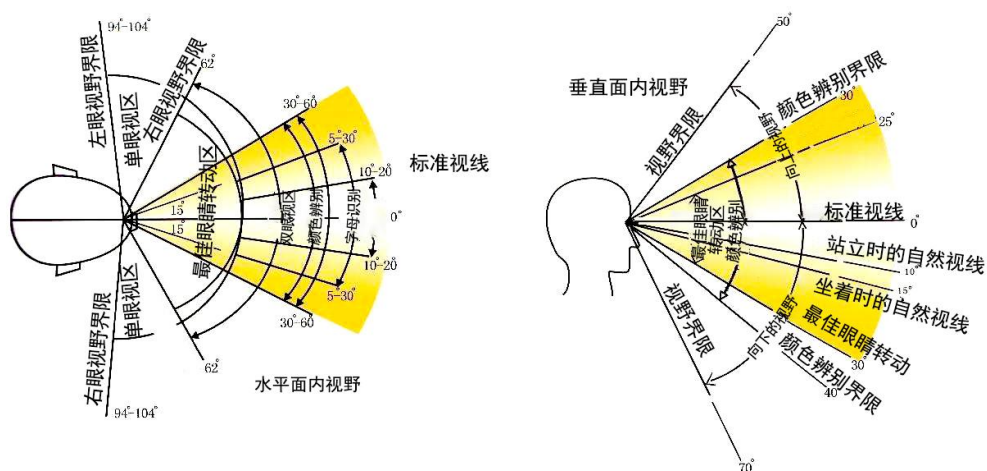


图 2-2 双眼水平视野、垂直视野示意图

为了实现上述的画面包围感，沉浸式视频可采用的呈现形式包括球幕、环幕、沉浸屋 CAVE 等异形显示空间，显示方式可采用投影+反射式幕或 LED 等自显示屏。

2.1.1 球幕

球幕是以球形恒曲率屏幕全方位包围观众，达到沉浸体验感的一种呈现形式。球幕可以分为全球幕和半球幕，如图 2-3 所示。根据观看方式的不同，球幕又可以分为穹顶式、倾斜式和垂直式。穹顶式球幕，观众一般仰躺着观看，如天文馆星象厅里的球幕；倾斜式球幕，观众一般坐在阶梯座椅上观看，如 IMAX 球幕影院；垂直式球幕，观众一般坐在悬挂座椅上或站立在观景平台上观看，如飞行影院。

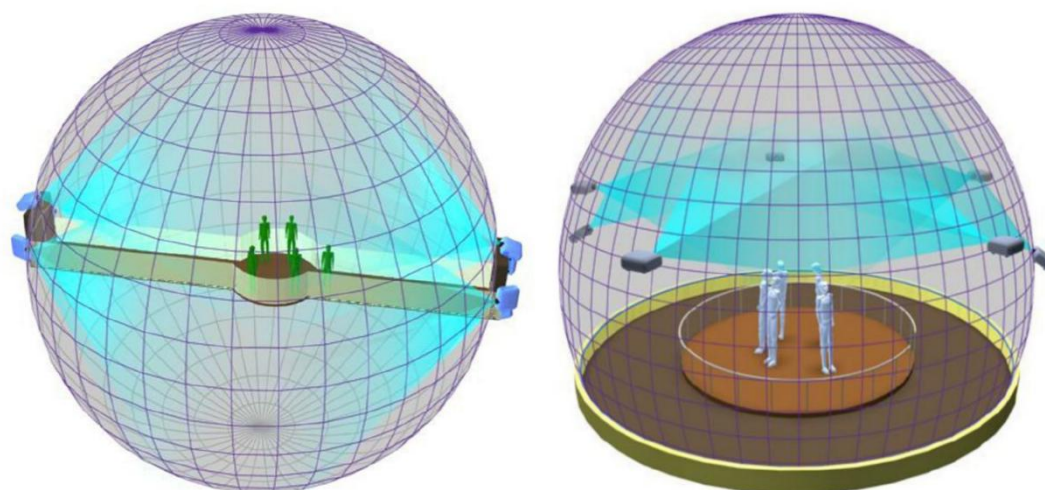


图 2-3 (a) 全球幕模型

(b) 半球幕模型

2.1.2 环幕

环幕是以柱面环形屏幕在水平方向上包围观众，实现沉浸式体验的一种呈现形式。为了满足沉浸式视频所需的包围感，环幕的弧度可以从 120° 到 360° ，观众站立或坐在环内侧观看，如图 2-4 所示。

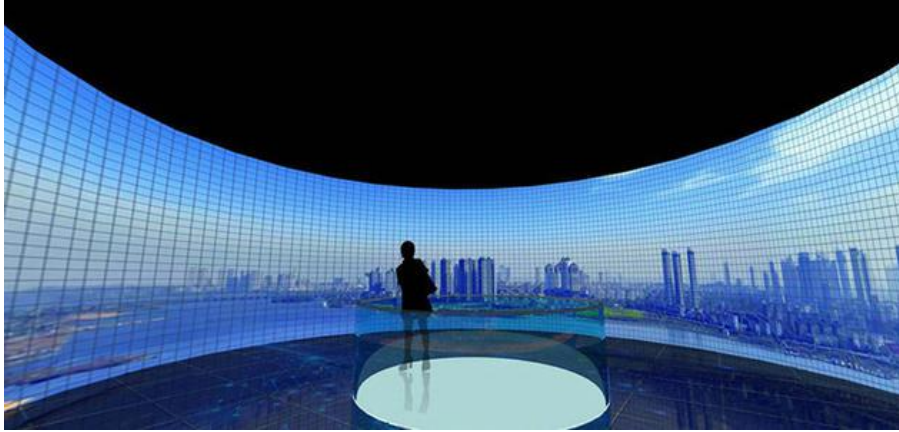


图 2-4 环幕案例

2.1.3 沉浸屋 CAVE

沉浸屋 CAVE 是以多面屏幕（或墙面作为影像显示介质）组成的像屋子（洞穴）一样，具有高度沉浸感的呈现形式。沉浸屋通常由三至六个显示面组成，如图 2-5 所示。除了直角的 CAVE 外，还有圆角的 CAVE（将显示面相交处做弧状处理），使其视觉上更柔和，有利于衔接处画面的连贯和完整。



图 2-5 沉浸屋案例

2.1.4 其他异形呈现形式

其他异形呈现形式包括折幕、碗幕、胶囊影院等形式。碗幕多为双曲面屏幕，观众坐在悬空座椅里或者站立在平台上进行观看，模拟飞行状态俯瞰大地的感觉，视点着重于斜下方向，观众沉浸在大范围的影像中，再配合座椅的运动，营

造出独具特色的沉浸式体验。近年来在主题公园里兴起的飞行影院就属于这一类，如图 2-6 所示。

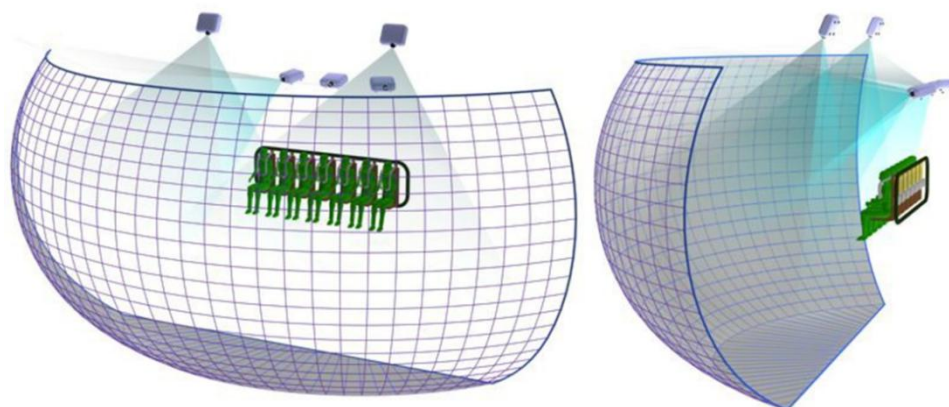


图 2-6 飞行影院模型示意

2.2 画质

沉浸式视频追求画面的高质量。分辨率、对比度、色域、帧率、色深等多方面是体现画质的重要指标。分辨率决定了图像的清晰度，动态范围决定了图像明暗对比度，色域决定了颜色的丰富度，帧率决定了图像的流畅度，色深决定了颜色渐变的平滑度。

2.2.1 分辨率

分辨率是决定沉浸式视频技术质量的一个重要参数，由画面内容及呈现端系统的像素数决定。高分辨率带来了更为丰富的画面层次和更精致的画面细节，超高清视频 UHD 定义了 4K 和 8K 两种分辨率。而沉浸式视频作为一个系统，在分辨率选择上，需综合画面分辨率与人眼的水平视场角、人眼视敏角三者的关系确定。

基于人眼的生理特性，2004 年，国际天文馆协会（IPS）对球幕的分辨率给出了建议，指出观众观看天象的要求应达到介于“模糊感知”和“精确辨别”之间的“辩知”水平，此时银幕面上相邻两个像素与视点连线夹角应不低于 3 弧分（Arcminutes）；若要达到“精确辨别”，这个指标对应的像素夹角为 1.5 弧分。

基于以上，球幕呈现形式下需要满足的水平方向最低分辨率计算公式如式（2-1）所示，计算用表如表 2-1 所示。假定观看点在球心，式（2-1）中 Y 为水

平方向最低分辨率， θ 为水平视场角， α 为银幕面上相邻两个像素与视点连线夹角。

$$Y \geq \frac{\pi\theta}{360 \sin \frac{\alpha}{2}} \quad \text{式 (2-1)}$$

表 2-1 球幕呈现方式下水平方向最低分辨率要求表

| 最低分辨率 Y 水平视场角 θ | 视点夹角 α | |
|---------------------------|-------------------|-----------------|
| | $\alpha = 1.5$ 弧分 | $\alpha = 3$ 弧分 |
| 120° | 4800 | 2400 |
| 180° | 7200 | 3600 |
| 360° | 14400 | 7200 |

如上述表中显示：以满足水平视场角为 180° 的球幕为例（一般采用半球球冠的球幕来实现），若要达到精确辨别的需求，分辨率对应的像素数为 7200，相当于 8K；若要达到模糊辨别的需求，分辨率对应的像素数为 3600，相当于 4K。

对于相同的水平视场角，环幕的弧长均比球幕要短，所以可以认为，只要满足了球幕的分辨率要求，就可以满足环幕的分辨率要求。而对于沉浸屋需要的分辨率，则需针对具体的项目分别计算。

与超高清视频 UHD 采用的画幅比 16:9 不同，沉浸式视频节目具有特殊画幅比（Aspect Ratio）。如半球幕节目的画幅比是 1:1，环幕和沉浸屋的画幅比需要根据幕的比例或房间的尺寸来确定。

2.2.2 动态范围

传统电视中所使用的标准动态范围（SDR）的亮度范围比较有限，无法有效还原画面中光线和颜色的渐变和层次，大量细节丢失。高动态范围（HDR）是一种能够显著提升图像对比度的技术，其核心在于提高高光画面的亮度，控制暗部画面的亮度，以此强化对比度，表现出画面中的光线和颜色的渐变和层次细节。

目前主流的 HDR 技术有 HDR10、HDR10+、SL-HDR1、杜比视界和 CQUA-HDR 等。

HDR 有两种转换曲线，分别是 PQ 曲线和 HLG 曲线，由于曲线特性的差异，它们对显示兼容和适配的支持也不一样。PQ 曲线在所有的回放端都需要做映射

处理,以便适配不同显示设备的显示能力,这种映射一般需要借助元数据来实现。PQ 在制作、传输和回放环节均需引入元数据,根据不同的终端实现相对准确的调整,适配效果较好。而 HLG 曲线是一条相对曲线,如果回放设备亮度等于制作监看亮度,则无需在回放端进行映射,即可正确回放;如果回放设备亮度不等于制作监看亮度,则只需要传输节目制作时的实际峰值亮度,就可以在回放端根据伽马值进行调整,以实现显示兼容和适配,但显示屏的性能将对显示效果产生较大的影响。如果希望在不同显示屏上获得一致的效果,仍然需要利用动态元数据进行映射。

HDR 易于在自发光类显示屏技术上获得实现,若显示屏表面经过吸光处理,HDR 更易实现;目前,大多数采用内弧形式的反射式银幕,很难实现“高动态”要求。**建议沉浸式视频支持 HDR,但投影呈现的内弧形球幕终端不适用。**

2.2.3 色域

色域 (Color Gamut),是指屏幕显示设备所能表达的颜色数量构成的范围区域。广色域 (Wide Color Gamut) 是一种更大范围的色彩标准,国际标准规定色彩覆盖率能达到 NTSC 的 92%,即为广色域。图 2-7 表示了不同色域所覆盖的色彩范围。全高清电视对应的色域为 ITU-R BT. 709,DCI 规定了数字电影对应的色域为 P3,超高清电视对应的色域为 ITU-R BT. 2020。ITU-R BT. 2020 是目前对显示设备提出的最大的色彩范围要求。**建议沉浸式视频的色域支持 ITU-R BT. 2020,条件暂不具备的可放宽至 DCI-P3。**

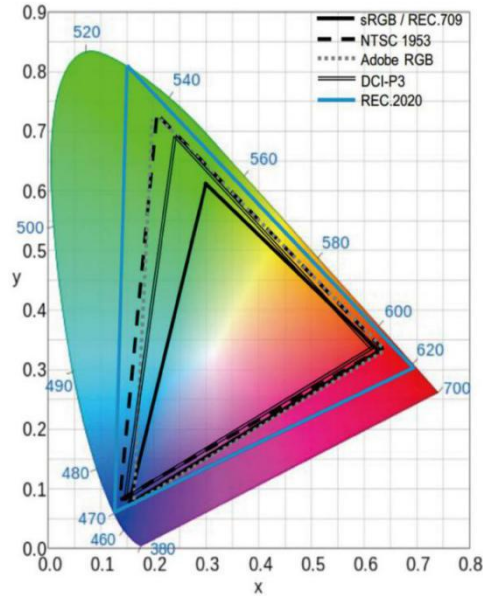


图 2-7 不同色域的色坐标覆盖范围

2.2.4 帧率

帧率 (Frame Rate) 是以帧为单位的位图图像连续出现在显示端上的频率，高帧率可以得到更流畅的影像。一般来说每秒 30 帧画面是人眼可以接受的，将帧率提升至 60 帧，可以让动态的图像减少“拖影”，增强画面的流畅感，提高人眼观看的舒适度。沉浸式视频的帧率最低每秒 50/60 帧，未来需要达到每秒 100/120 帧。

2.2.5 色深

色深 (Color Depth) 称为颜色的量化精度，是表示数码影像色彩数目的术语，单位是比特 (bit)，数值为 2 的幂指数。bit 数越高，色深值便越高，影像所能表现的色彩也越多。4K 超高清标准中规定色深为 10bit，即能够显示 10.7 亿种颜色 (每种颜色具有 2^{10} 个灰阶，三种颜色就是 $2^{10} \times 2^{10} \times 2^{10} = 10.7$ 亿)，而标清标准对应的是 8bit 色深，只有 1677 万种颜色 ($2^8 \times 2^8 \times 2^8 = 1677$ 万)，差距悬殊。8K 超高清对应 12bit~14bit 色深，能进一步提升颜色的细腻程度。同时，10bit 以上色深也是实现 HDR 的必要条件，量化精度的提高可以保证高动态范围内的色彩过渡平滑柔和。建议沉浸式视频色深使用不低于 10bit，推荐使用 12bit。

2.3 沉浸式声音

真实世界的声音来自环境的四面八方，人耳可以通过声波的时间差、强度差、相位差、频率差等辨别声音的方位。现有的立体声和 5.1 环绕声只能重放部分方向来的声音信息，而想要获得沉浸感，需要尽可能全方位地再现真实世界的声音。因此，沉浸式声音是沉浸式视频系统的一个核心要素。

形成沉浸式声音，要求呈现空间的还音系统的声辐射至少能覆盖观众的前、后、左、右、上五个方向，能真实地营造出声场的水平纵深和垂直高度，能够正确地定位声音的方向和位置。

沉浸式声音主要通过三维声技术来实现。三维声技术实现主要分为三大类：基于声道的音频（CBA）、基于对象的音频（OBA）和基于场景的音频（SBA）。

基于对象的技术是目前主流的三维声技术。在电影领域，三维声技术已广泛应用，如 Dolby Atmos 全景声、DTS-X、HOLOSOUND、WANOS 全景声、中国多维声、Auro-MAX 等。在广电领域，三维声技术主要有 Dolby Atmos 和 MPEG-H、AVS2-P3。

沉浸式视频的声音系统推荐使用基于对象或场景的三维声技术，需要实现不低于 5.1.4 声道的声音系统，即在传统 5.1 环绕声基础上增加 4 个顶部声道。

2.4 沉浸式视频核心要素表

用于沉浸式视频制作的核心要素见表 2-2。

表 2-2 沉浸式视频核心要素表

| 序号 | 参数 | 推荐数值 |
|----|---------|------------------------------|
| 1 | 水平视场角 | 大于等于 120° |
| 2 | 垂直视场角 | 大于等于 70° |
| 3 | 呈现形式 | 球幕、环幕、沉浸屋 Cave 等异形显示空间 |
| 4 | 分辨率 | 4K/8K |
| 5 | 动态范围 | HDR（投影呈现的内弧形球幕终端不适用） |
| 6 | 色域 | 不低于 DCI-P3，推荐 ITU-R BT. 2020 |
| 7 | 帧率（fps） | 不低于 50/60，推荐 100/120 |
| 8 | 色深（bit） | 不低于 10，推荐 12 |
| 9 | 声音系统 | 不低于 5.1.4 声道 |

3. 沉浸式视频关键技术

沉浸式视频系统所涉及的关键技术包括全景视频拍摄技术、视频图像缝合技术、三维图像映射技术、计算机建模渲染技术、三维声技术、编码及传输技术、媒体播控技术、投影呈现技术、LED 屏呈现技术等。

3.1 全景视频拍摄技术

由于普通摄影机的成像方式和画面摄取角无法覆盖水平与垂直 360° 全景范围，所以沉浸式视频基本的拍摄制作是采用多个相机、多个镜头，在同一时间、同一位置分别拍摄不同角度有部分重叠的画面，然后将采集到的多角度图像进行拼接以得到完整的 360° 全景画面。

3.1.1 多机组合拍摄

全景视频拍摄理论上可以采用各种相机+镜头的组合，但受光学原理、拍摄环境、相机体积以及多机同步等各种问题的限制，行业内通过多年实践已形成了一些常用的组合方式，如下表所示。

表 3-1 多机组合方案比较

| 相机组合方案 | 设备举例 | 特点 |
|---------|----------------------|-----------------------------|
| 运动相机组合 | GoPro Omni 或其他各种支架组合 | 画质一般，体积小 |
| 专业摄影机组合 | RED Dragon 电影机组合 | 全画幅或更大的传感器尺寸，画质优异，体积大，系统造价高 |
| 无反相机组合 | SONY A7s 系列相机 | 专业摄影机组合的折中方案 |

为了达到更好的观看效果，多人观看的沉浸式视频除了在分辨率需要能够达到 4K、8K 甚至更高，画面色彩和质量相对于 VR 视频也有更高的要求。多机同步拍摄的优势在于使用了多个具备较高参数的摄像机采集图像，传感器尺寸可达到

Super35 或更高，获得的图像具有较高的分辨率、动态范围和彩色量化深度，多用于对图像质量有很高要求、能够承受后期复杂制作流程的应用场景。

多机同步拍摄的主要问题是需要使用比较复杂笨重的多机支架；实际拍摄中对于场地、移动要求较高；拍摄前需要对各个摄像机、镜头等进行匹配调整；拍摄完成后各机位采集的素材需要进行时间码对齐、色彩统一调整、图像拼接缝合等处理才能形成全景图像，因此不适合于直播应用。国外有电影拍摄的镜头组合成相机阵列的应用，在一定程度上解决了支架、多机同步等问题，但造价昂贵。



图 3-1 两种用于球幕拍摄的相机阵列 Shot over Hydra 和 Pictor Vision Eclipse

3.1.2 全景相机拍摄

一体化全景相机一般由多个内置的超广角/鱼镜头组成，是目前沉浸式视频/VR/全景视频常用的拍摄设备，可视为上述全景拍摄系统中运动相机组合的集成化产品。

一体化全景相机的优点是系统整合度高，便于现场操作，易于后期制作及直播；机内各镜头和图像传感器的参数及帧同步等工作在机内完成；现场拍摄时只需全局配置拍摄参数；统一供电，统一 I/O；体积较小，重量较轻，可配合各种轨道、吊索、无人机等进行拍摄；大部分一体化全景相机可以进行机内实时拼接、实时监控，可用于不要求极致画面质量的直播等场景；相机厂家一般会提供匹配相机参数的定制化软件，拼接效率较高。

一体化全景相机的缺点是受集成度限制，图像传感器小，成像质量一般，对环境光线要求高。目前性能最好的专业级一体化全景相机可拍摄(拼接后能达到)最高 8K 10bit 30fps、8K 8bit 60fps 或 11K 8bit 30fps 的全景视频。

3.2 视频图像缝合技术

来源上，沉浸式视频可采取实景拍摄和计算机建模两种方式获得。实景拍摄一般采用多机组合拍摄的方法，需要进行视频图像的缝合才能形成一个完整的沉浸式视频画面。

目前应用于全景视频的缝合方法主要有两类：基于模板的缝合和基于光流的缝合。

3.2.1 模板缝合

模板缝合就是将相邻的两张图像直接拼到一起并进行融合渲染。模板缝合有两种常见的方式，第一种是通过设立一组参考帧，通过参考帧中相邻图像的像素进行匹配计算，得到模板参数，然后将此参数应用到视频中的其他图像中。第二种是在相机出厂之前进行拼接标定来确定相机的内外参数，确定好相机镜头之间的相对位置，然后将此参数直接用于视频或者图像的拼接。由于模板缝合存在各帧之间的差异，可能在融合区域出现重影，画面质量相对较差。模板缝合的拼接速度比光流缝合快，对处理设备的性能要求也没有光流那么高。

3.2.2 光流缝合

所谓光流，是通过一组图像序列的像素点来判断图中的物体位置。光流理论基于同一物体在相邻的两帧图像之间像素灰度相同的假设，根据像素匹配得到图像 HSV 颜色模型的色调（H）、饱和度（S）、明度（V）参数值，经过双边滤波等处理方法得到最终包含所有拼缝处像素的插值表。用插值表来校正像素的位置，将相似像素从一幅图像“扭曲”或“流动”至另一幅图像然后拼接融合起来。常见的插值算法有多项式插值和分段插值。光流拼接的拼接效果优于模板拼接，但是速度也会相应的慢一些，对运算性能的要求也相对更高。

3.3 三维图像映射技术

三维图像映射是指将三维（球面）的图像转换为适于压缩编码的二维平面图像。

3.3.1 鱼眼映射

鱼眼映射指使用鱼眼镜头成像的原理将三维立体空间里的目标点与平面的成像点对应起来，通常有正交鱼眼映射和等距鱼眼映射两种方式。

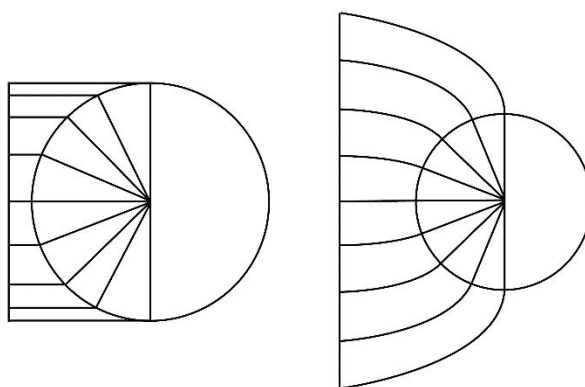


图 3-2 正交鱼眼和等距鱼眼映射原理示意图

(1) 正交鱼眼映射

正交鱼眼映射是把半球面上的点平行投射到平面上，在图像平面上得到一个圆形的映射图像。图 3-2（左）描述了正交鱼眼映射。半球映射视角范围最大到 180° 。在 0° 附近，映射图像几乎没有变形，但是在平面的最外边缘（接近 90° ），图像变形很大，这个地方的图像看起来显得比较拥挤。由于正交鱼眼映射的这种放射性变形，球幕边缘的信息丢失严重。因此，在实际应用中，相比于等距鱼眼映射，正交鱼眼映射的应用较少。

(2) 等距鱼眼映射

等距鱼眼映射是让映射图像上的点到图像中心的距离，和球面上对应点与球心连线到照相机光轴的方向角成比例。在等距鱼眼映射图像上，整个图像分布几乎是均匀的。等距鱼眼映射视角范围最大可以到 360° 。等距鱼眼映射是目前球幕视频、全景视频（主要是全景运动相机）使用比较广泛的一种映射方式。

3.3.2 等距圆柱映射

ERP 模型 (Equi-Rectangular Projection) 等距圆柱映射, 通常也被称为经纬映射。目前在 360° 全景/VR 视频应用中最为常用。等距圆柱映射的实现是用相同数量的采样点保存每条纬线上的数据, 从而得到对应的二维平面上的矩形视频。映射后的二维平面图像水平方向为球面经度 $-180^{\circ} \sim +180^{\circ}$, 垂直方向为球面纬度 $-90^{\circ} \sim +90^{\circ}$ 。360° 全景视频映射后的原始平面图像纵横比为 2:1。

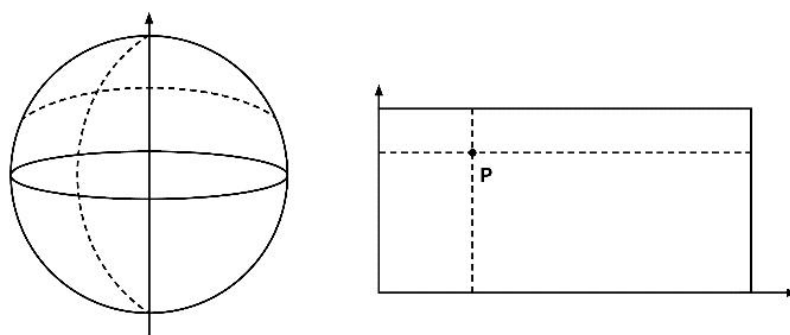


图 3-3 等距圆柱映射原理示意图

图 3-3 所示的是单位球面上的 P 点和在等距圆柱映射平面上的对应点的关系。常见的矩形世界地图就是等距圆柱映射图像。

这种映射方式的映射关系简单, 即使是映射后的二维平面视频也很直观, 方便用户观察。其直观的特性便于后期制作人员对视频的编辑, 使其成为流行的映射算法, 主流的视频后期制作工作站都支持这种映射方式, 是目前 VR/全景视频领域最常用的方式。

ERP 映射存在的问题是, 离赤道越远的像素拉伸越严重, 图像变形越大。由于 ERP 对球形视频每条纬线上都用相同数量的采样点, 导致越靠近两极冗余采样点数量就越多, 而主要视野范围的赤道附近却只分配了少量的像素。在两极处, 本来只需要一个采样点, 但却用了和赤道处同样多的采样点, 冗余数据量最大。一个球的等距圆柱映射面积比球的表面积多 57%, 严重影响了视频编码码率和球面还原后的画质呈现。

3.3.3 正六面体映射

CMP 映射（Cube Map Projection，正六面体映射）是将球面映射到一个正六面体（立方体）的六个正方形面上，立方体的上下两个面分别对应两极区域，中间的四个面对应赤道区域，每个面上的像素是从球心处以 90° 的 FOV 看到球上的像素在平面上的映射。这种映射方式在计算机图形领域已有较长时间的应用。

在球面映射到几何体表面的方法中，CMP 映射可以改善 ERP 中像素分布极度不均匀及图像严重拉伸等问题。但由于立方体每个面的中心位置距离球体最近，越靠近边角的地方距离球体越远，因而导致球面上相同数量的像素点映射到立方体边缘区域所分配到的采样像素数量会多于映射到中心区域时所分配的采样像素数量，即边缘区域稀疏，中心区域稠密，所以像素分布依然是不均匀的。

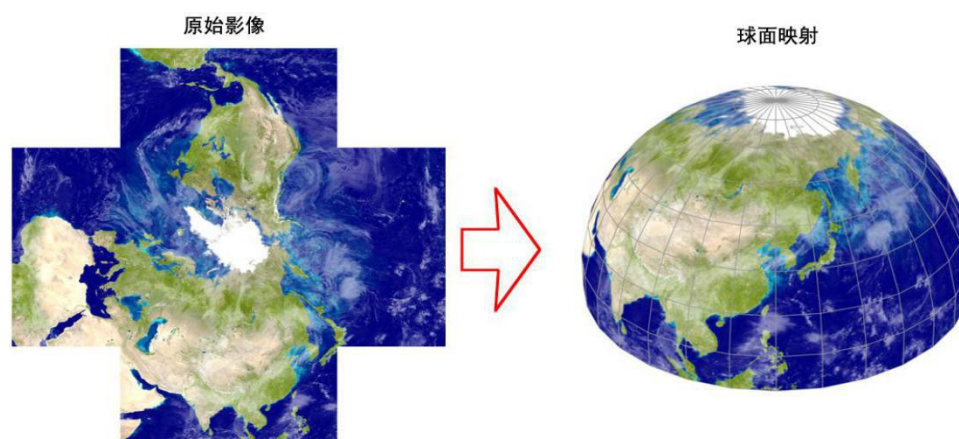


图 3-4 正六面体映射原理示意图

3.3.4 等角方块映射

等角方块映射（Equi-Angular Cubemap，EAC）是正六面体映射的进一步优化格式，其目标是不论取样点在映射面的什么位置，都能获得统一的像素观看密度感。EAC 映射模型在 6 个映射面上对采样点的位置进行优化，使各角度内的像素密度保持恒定不变，如此一来就能让“赤道”周围的画面更加清晰，也能避免“南、北极”部分像素过剩的情况，节省储存空间。相较于单纯使用正六面体映射投影法，EAC 增加了调整区段角度的环节，能够让像素密度更加平均，在

相同的视频分辨率下可以提高细节部分的清晰度，是目前质量最高的一种映射模式。

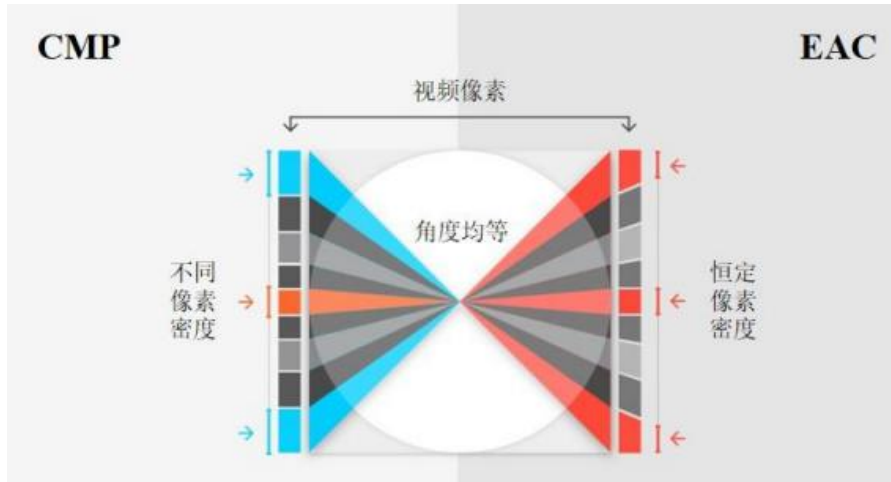


图 3-5 EAC 与 CMP 映射模型对比示意图

其他映射方式还有如八面体、十二面体等各种异形的映射模型，在目前的日常应用中较为罕见，此处不做赘述。

3.4 计算机图像制作技术

计算机图像制作（通常称为 CG 制作）是采用计算机硬件及软件，通过三维建模和动画软件，建立完全虚拟的空间场景，通过对模型场景的逐帧渲染得到动画视频。CG 制作的沉浸式视频灵活性较强，可以自由地设定环境、视角、对象、光照等，不受实景拍摄时的各种限制条件的约束，也不存在多个摄像机拍摄画面缝合带来的画面连续性问题。沉浸式视频与传统 CG 动画的制作方式基本相同，只在影像的渲染、合成等方面略有区别。

沉浸式视频的画面视角覆盖面积远远大于常规 CG 动画，使用常规动画生产手段需要面临较为严峻的渲染压力。在渲染环节，除了 3Ds MAX 和 MAYA 自带的渲染器之外，目前多使用 Unreal 游戏引擎和 V-Ray 两款渲染工具，两者在影像实时渲染、全景合成等方面较为高效，合理使用可以较好地加快生产效率。在合成环节，受到三维动画软件中渲染相机的限制，大多数三维动画软件无法直接输出合适的全景沉浸式视频，一般要把三维的视频映射成二维的序列帧文件后，再进行画面剪辑。

除普通游戏动画、三维动画内容外，科学可视化内容，如星空宇宙的科教片等等，也天然具有全景沉浸式视频的特点。通过格式转换后，可以在 MAYA 软件中通过 Mel 脚本对科学可视化对象进行编辑和输出，最终生成沉浸式视频的内容。

3.5 三维声技术

沉浸式视频要实现沉浸式声音，主要通过三维声技术来实现。三维声技术分为基于声道、基于对象和基于场景的三种技术形式。

3.5.1 基于声道的技术

基于声道的技术，如 5.1.4，就是在传统 5.1 环绕声的基础上，增加了 4 个顶部声道，通过增加声道的方式来补充空间中的声音信息。听众只有在很小的区域内（即通常所说的“皇帝位”）才能听到最佳的声音效果。基于声道的方式在立体声以及环绕声的音频制作中被广泛使用，但是它对扬声器布局绝对依赖，不利于三维声的音频制作。

3.5.2 基于对象的技术

基于对象的技术，是目前主流的三维声技术。基于对象的技术通过赋予声音对象空间坐标的方式使得声音可以定位于空间中的任意一点。基于对象的音频技术的一项核心组成部分就是元数据，它记录声音的位置、扩展度或者运动特性，这些元数据传输到终端，通过渲染后再进行重放。和基于声道的实现方式不同，基于对象的技术不受还音条件（诸如音箱数量）的限制，渲染器可以读取音频对象的元数据信息，使得同一母版能够适配包括耳机在内的各种还音条件。基于对象的音频技术主要有 Dolby Atmos 全景声、DTS-X、Auro MAX、MPEG-H，国内自主研发的技术有雷欧尼斯 HOLOSOUND、AVS2-P3 (WANOS 全景声)。

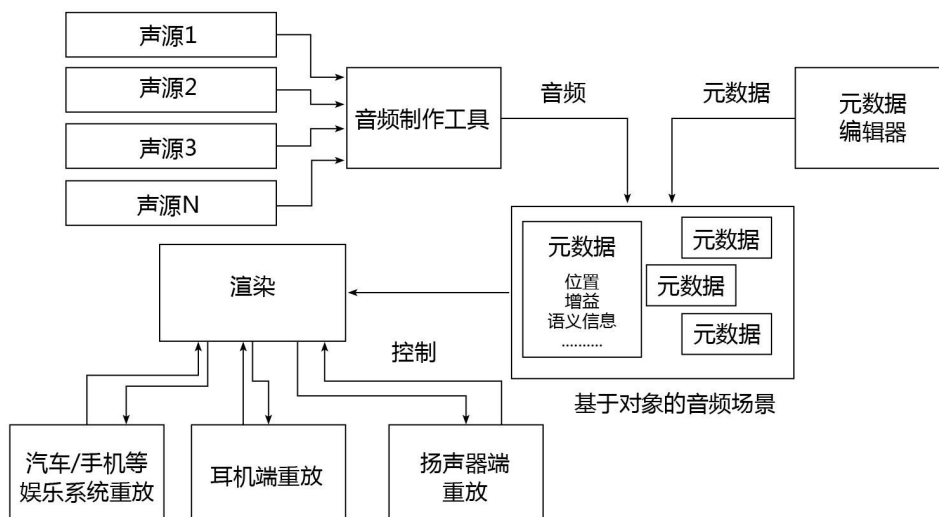


图 3-6 基于对象的三维声制作流程图

基于对象的三维声制作可以借用传统声音录制方式，增加音频对象的制作和渲染重放手段。飞机在头顶从左到右飞过的声音、直升机螺旋桨由远及近的声音、森林里踩踏树叶的嚓嚓声、子弹出膛的声音等等这些，都可作为独立的音频对象。这些音频对象的位置、增益等信息被元数据编辑器记录下来，再通过渲染器运算还原，供扬声器系统或耳机进行重放。

3.5.3 基于场景的技术

基于场景的技术，主要是通过记录三维空间中声压来重现这个空间中的声音。其核心的底层算法是 Higher Order Ambisonic(HOA)，主要以 MPEG-H、AVS2-P3 为代表。麦克风阵列捕获的声音信号，转换为 HOA 系数后，经编解码处理并渲染后，可被映射到任意扬声器布局中，同样也可应用于沉浸式耳机中，因此在 VR 虚拟现实和游戏中的应用需求非常大。

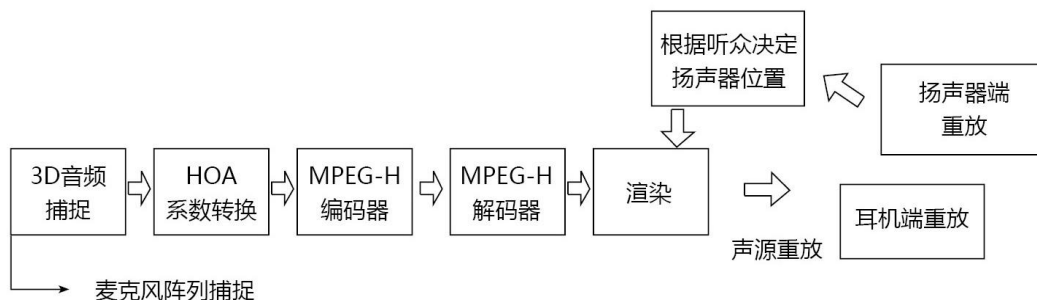


图 3-7 基于场景的三维声制作流程图（以 MPEG-H 为例）

灵活的渲染是基于场景音频制作方式的关键优势之一，不再需要指定扬声器配置，而是可以针对不同数量和位置的扬声器进行渲染，以便在各种环境中对声音场景进行再现。

3.6 编码及传输技术

3.6.1 沉浸式视频的码流

若基于图像分辨率参数，沉浸式视频的码流需求可分为 4K 及 4K 以下、8K 和 8K 以上三个档次。4K 视频的平均码率为 60~100Mbps，传输速率需求为 80~125Mbps；8K 视频的平均码率为 150~200Mbps，传输速率需求为 180~250Mbps，8K 以上视频的平均码率为 300~500Mbps，传输速率需求为 350~600Mbps。对应的视频参数及码流需求见表 3-2。

表 3-2 沉浸式视频的码流需求及视频参数

| 图像分辨率 参数名称 | 4K 及 4K 以下 | 8K | 8K 以上 |
|---------------|---------------|----------------|----------------|
| 分辨率 | 4K 及 4K 以下 | 8K | 8K 以上 |
| 帧率 | 50/60 | 50/60 | 100/120 |
| 色深 | 10bit | 10/12bit | 12bit |
| 压缩编码 | H. 264 AVS | H. 265 AVS2 | H. 266 AVS3 |
| 平均码率 | 60~100Mbps | 150~200Mbps | 300~500Mbps |
| 传输速率需求 | 80~125Mbps | 180~250Mbps | 350~600Mbps |

3.6.2 沉浸式视频的编码

视频编解码技术是数字电视发展的基本条件。应用于沉浸式视频的视频编码技术主要有：国际上 ITU 提出的视频编码标准 H. 264、H. 265 和 H. 266，以及国

内数字音视频编解码技术标准（AVS）工作组制订的具备我国自主知识产权的视频编码标准 AVS/AVS+、AVS2 和 AVS3，这些标准与国际标准对应。

AVS2 和 H. 265 的压缩效率比上一代标准 AVS 和 H. 264 提高了一倍，其中 AVS2 的压缩效率超过国际同类型标准 H. 265。下一代编码技术 H. 266 与 AVS3 的标准正在制定中，目标是实现编码效率比 H. 265 与 AVS2 再提升一倍。

根据码流的需求，目前 4K 沉浸式视频主要采用 H. 264，H. 265 和 AVS2 编码标准。8K 沉浸式视频主要采用 H. 265，H. 266 和 AVS3 编码标准。

3.6.3 沉浸式视频的传输

沉浸式视频的传输根据不同的应用场景和实际条件，可采用有线电视、5G 无线网络、互联网方式进行传输。三种传输方案可适用于沉浸式视频的各种应用需求，本白皮书给出对传输速率及时延指标要求的推荐值，详见表 3-3。

表 3-3 三种传输方案适用场景

| 传输方案 | 指标要求 | 适用场景特点 |
|--------|--|--|
| 有线电视网 | 时延 $\leq 10\text{ms}$ ； 下行速率 $\geq 100\text{Mbps}$ | 可以满足沉浸式视频在带宽及服务质量上的要求，同时拥有良好的入户基础，在超过 2 亿的有线电视用户基础上，可以发挥入户优势，为家庭用户提供更多的应用形式。 |
| 5G 无线网 | 时延 $\leq 10\text{ms}$ ； 下行速率 $\geq 100\text{Mbps}$ | 5G 技术的大带宽、低时延等特性可以满足沉浸式视频的传输要求，同时利用 MEC 以及网络切片技术可以进一步降低成本及保障用户体验，除了可以满足固定场景下的沉浸式视频传输需求外，还可以更好地解决移动场景下沉浸式视频的传输需求，进一步丰富沉浸式视频的应用形式。 |
| 互联网 | 时延 $\leq 100\text{ms}$ ； 下行速率 $\geq 100\text{Mbps}$ | 互联网可以承载包括 IPTV、流媒体等业务的视频。使用光纤传输沉浸式视频可以获得更高的服务质量，无源光网络的模式也降低了建设成本和后期的运维成本，更适合企业用户的应用场景。 |

3.7 媒体播控技术

媒体播控服务器作为沉浸式视频呈现的关键设备，其核心功能为画面分割、拼接融合和视频播放。专业媒体服务器需要支持超大分辨率、高帧率的视频播放，同时通过其内置的拼接融合功能将多台投影机画面或多块 LED 等自显示屏拼接成一副完整的图像。除此之外，媒体播控服务器还涉及到自动校准技术、透视校正技术、环境特效控制等。

3.7.1 视频解码播放

视频播放包含音视频文件的读取、音视频解码、图像渲染等环节。

沉浸式视频涉及高分辨率、高帧率（例如 8K×8K 60fps）、高码率，采用以 H.264、H.265、AVS2 为代表的帧间压缩算法，可以保证较高的压缩比，同时保证优质的图像质量，解码之后的每一帧图像通过显卡的接口输出到投影机或者显示屏。

3.7.2 投影拼接融合

拼接融合技术将多台投影机的输出融合为一幅画面，首先要将一幅完整的画面分割为有重叠区域的区块，然后将区块通过显卡接口分配给投射对应区域的投影机显示。由于沉浸式呈现空间的银幕具有不规则性，对于投影的画面需要对其做相应的几何校正，以实现投影画面和银幕的匹配。同时，由于各投影机的输出画面在安装调试的时候就设计了一定的重叠区域，该重叠区域是为了相邻两幅投影机的画面过渡足够自然、平滑，那么就需要对每一幅画面的重叠区域做羽化处理，使得重叠区域的亮度可以和非重叠区域保持一致。

投影融合的核心功能为画面分割、几何校正、边缘羽化等。对于球幕和异形银幕，需使用 3D Mapping 的方式来完成投影画面的几何校正和融合带校正。

3.7.3 自动校准技术

自动校准技术可以对投影系统进行自动校准。媒体服务器的自动校准功能一般可以通过以下两种方案实现：基于相机的自动校正和基于特征点的自动校正。基于相机的自动校正通过拍照记录不同位置的投影画面进行反算叠加，实现投影系统的自动校正。基于特征点的自动校正主要用于 3D Mapping 以及非正投情况下的投影融合，它通过添加特征点（4~6 点）的方式，将画面中的特征点和物体模型上的特征点匹配，通过矩阵求解，反算出投影机在三维空间中的位置、姿态、投射比等参数，进而实现投影画面的自动选取和融合带的自动生成，只要计算机模型与实际物体空间关系保持一致，就可在无额外的几何校正基础上实现良好的投影融合。

3.7.4 透视校正技术

3.3 中的投影映射是针对标准球幕的计算方法，即便是标准球幕，由于观众观看的位置不一定在球心处，也会带来由于视点引起的画面变形。而对于异形幕，如椭圆形的银幕，将无法采用投影映射方式进行处理；如果遇到更加复杂的胶囊型银幕，其画面变形将更加严重。

透视校正技术是基于全景拍摄已经获取环境中的所有画面信息，采用一套校正的算法，在媒体服务器中将标准球幕规格的内容进行二次渲染，无变形地展现给观众。此校正技术的算法是一种矩阵计算，可以通过 GPU 完成，不需要在制作环节进行预处理。

3.7.5 环境特效控制

为了有更好的沉浸式观影体验，目前在某些沉浸式影院中配有 4D 影院动感座椅、灯光、烟机等设备。其常用的控制协议为 DMX512，串口通信（RS-232、RS-422、RS-485 等）。媒体播控服务器需要具备这些常用的外部控制接口，支持这些控制协议。

以动感座椅为例，媒体播控服务器发布相应的控制指令，经过串口接口将指令发送到座椅中的传感器以及特效控制电路，来触发烟机的喷雾、喷气效果以及座椅的运动。

3.8 投影呈现技术

在球幕上投射画面可通过以下两种方式实现：单台投影机和多台投影机。单台投影机的优点为节约搭建成本，同时减少了画面融合带来的不必要的亮带和画面错位、扭曲等问题。但是单台投影的分辨率有限，屏前亮度很难提升，且亮度均匀性不高，无法在较大的场景中实现，最佳观赏视点往往被投影机占据。多台投影机的优点则在于可以通过多台投影机拼接融合，易于提高画面分辨率及亮度，适合大型的球幕，但需要保证多机安装的物理精度，以避免投影融合拼接带来的问题。除此之外，多台投影呈现还涉及投影空间几何校正技术、异形边沿融合技术、球面投影亮度均一化技术。

3.8.1 投影几何校正技术

针对多台投影机组合方案，需要保证每一台投影机投射到对应屏幕区域的图像内容与物理空间匹配对位。通过在三维空间中构建出投影机和实体屏幕之间的空间映射关系，需要精确转换出屏幕区域对应投影机每一个像素点的图像信息，进而得到科学精密的显示结果。

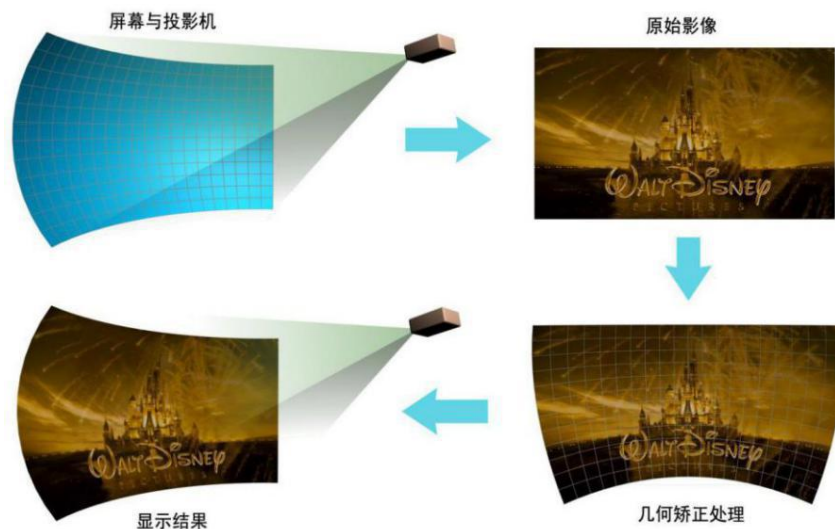


图 3-8 投影几何校正示意图

3.8.2 异形边沿融合技术

针对多台投影机组合方案，需将投影机投射到屏幕上的有效显示画面，通过几何分割的方式切分为不规整的画面形状。由于相邻投影的叠加区域是不规则的几何形状，在不同区域涉及双边叠加，三边、四边甚至更多投影光路的叠加，需要将融合区域按投影空间和屏幕拓扑结构拆分、建模、合理分布，采用多边融合的亮度衰减算法，对不同融合区域进行独立调试。

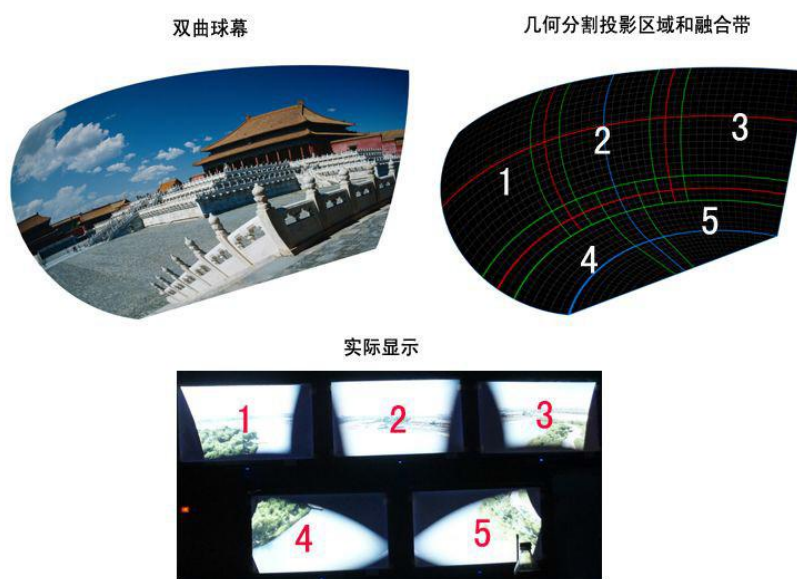


图 3-9 异形边沿融合示意图

3.8.3 球面投影亮度均一化技术

对于进行球幕拼接显示的各个投影设备，需要以非垂直角度将影像投射于曲面屏幕，这就产生随投影光线入射角度变化而带来的投影亮度变化和像素密度变化。

多台投影在不同的角度投射，投射的范围大小也不一致，最终导致整个球面屏幕不能取得均匀一致的亮度显示，自然无法得到高质量的拼接显示效果。需要通过屏幕像素点到投影点的逆向运算，分析出每个像素点的投射角度、有效光照亮度，计算出投射的各个像素点实际的光通量，对投射空间的所有像素点光通量进行均一化计算，最终得到一个图像过滤通道。将其叠加到每个输出通道画面上，即可在最终的实际投射环境中取得均匀统一的全屏画面亮度。但是，统一后的各

投影区域画面亮度较原先有所衰减。因此，在投影方位的设计中，应当尽量控制各投影机的投射范围是一致的，投射角尽量保持垂直，避免局部倾斜投射。

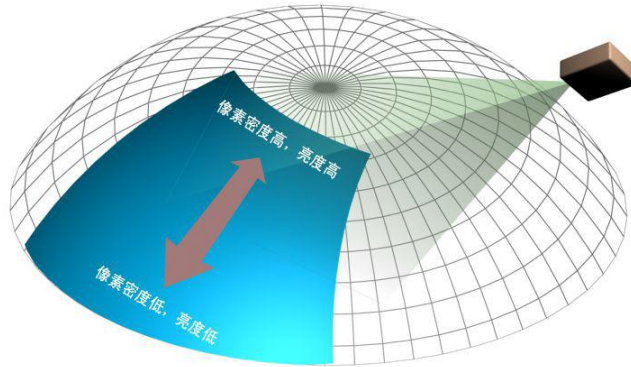


图 3-10 球面投影亮度示意图

3.9 LED 屏呈现技术

近些年来，自显示屏技术进步迅猛，其中最具有代表性的 LED 技术日臻成熟，除了成本因素之外，非常适合制作超大尺寸显示屏幕。LED 自显示屏可以实现超大尺寸的画面显示需求，像素无损且密度均匀，支持像素级单点校正，色域宽，亮度高，噪音低，不存在几何畸变校正。但 LED 屏显示方式需要解决亮度、均匀度以及色彩的问题，根据观影需求，沉浸式视频呈现的最高亮度不宜超过 200cd/m²，球面不应呈现明显的色差。针对沉浸式视频的特殊应用，LED 屏呈现技术需要结合实际要求，重点解决异形无缝拼接、透声及吸声、图像球形矢量变换等问题。

3.9.1 异形无缝拼接技术

球形 LED 影院可通过标准小尺寸的矩阵单元拼接成项目所需的形状。单曲面的显示屏，例如圆柱屏也可以使用标准小尺寸的矩阵单元拼接。而半球状或是近似半球状的 LED 显示屏，属于双曲面。因此，需要确定大型曲面的 LED 屏幕的拼接方法。目前成熟的方法有经纬分割法，以球体的赤道为中心线，球体上下对称，借助软件建模，计算分割出标准的模块尺寸和拼接方法。特别需要综合考虑系统控制和视频拼接等因素，以确定半球状和近似半球状的 LED 显示屏的基本单元的尺寸、带载方式、拼接角度等。

除此之外，需要开发新型 PCB 基板，解决 PCB 板异形、像素密度高、像素排布不规则、元器件多、布局布线复杂等技术难题；需按照球形弧度设计多种不同尺寸的异形弧面模组套件；设计模块化双弧面单元箱体，自动对接定位装置，以提升箱体拼装效率和拼接精度。

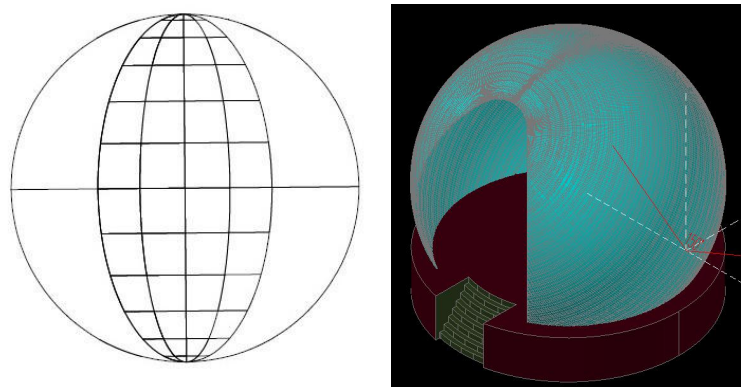


图 3-11 LED 球形拼接示意图

3.9.2 透声及吸声技术

沉浸式视频体验对呈现端系统及环境音质有很高的要求，为了保证良好的音质，需要对包括声压级、频响、环境空间混响时间和声缺陷以及语言清晰度等指标进行控制，使其符合要求。为满足还音系统音箱摆放以及降低球形空间的声聚焦问题，需要小间距的 LED 显示屏具有透声性能和吸声功能。但该技术实现受到 LED 显示面阵板布局的限制，一方面显示屏要求清晰度越来越高，导致像素与像素间可利用的开孔面积越来越小；另一方面有限的开孔面积不利于达到透声和吸声性能，需要更加精细地设计 PCB 板，并尽可能平衡好开孔和灯珠以及其他组件的位置关系。通常灯珠是阵列排布，开孔位于四个灯珠的中间位置，如图 3-12 所示。

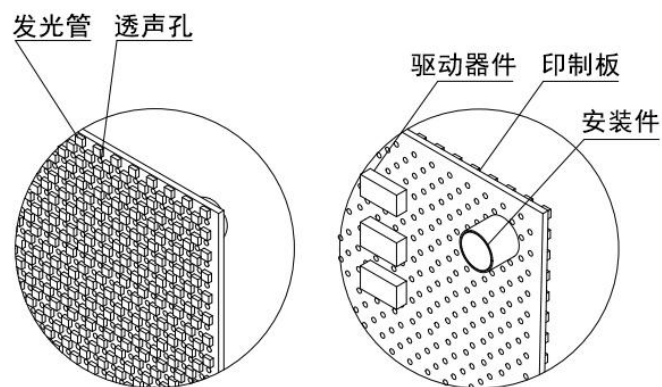


图 3-12 LED 透声板

3.9.3 图像球形矢量变换技术

由于球形 LED 像素的数量从“赤道”开始分别向两极逐渐减少，致使常规的平面矩形规格的视频源在球体表面的呈现会出现像素缺失或严重失真放大等情况。为了保证较好的显示效果，需要通过平面三角坐标、柱面坐标和极坐标的变换，实现将计算机输出的平面图像映射到 LED 球面上，使观众在球心或者最佳观看区域观看到的画面变形尽可能小。同时，需要综合考虑 LED 控制系统、媒体播控系统以及视频制作系统的匹配。

4. 沉浸式视频端到端解决方案

4.1 总体工艺架构

沉浸式视频业务主要可以分为直播和点播两种场景，两种场景分别对应不同的工艺制作流程，其拍摄、制作、编转码、传输分发和呈现系统的总体框架如下所示：

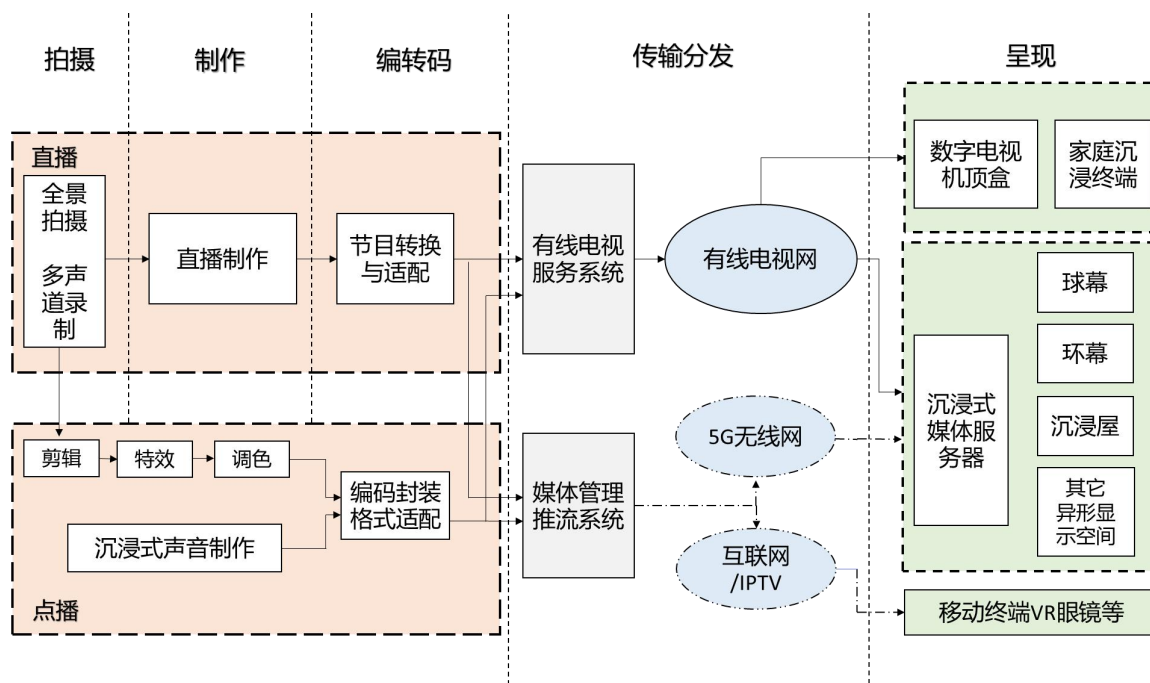


图 4-1 沉浸式视频端到端系统总体框架图

对于直播的沉浸式视频业务，通过全景摄像机采集拍摄，并实时进行拼接缝合合成全景画面，同时完成多声道音频录制，再同步将视音频信号进行适配网络传输的编码压缩、打包封装，通过有线电视网、5G 无线网、互联网将高码率的全景视音频流信号传输至呈现端的媒体服务器，最后经过音视频解码和渲染后呈现在终端显示设备上。

对于点播的沉浸式视频业务，拍摄缝合后的全景画面或者 CG 制作的画面需要进行剪辑、特效合成、调色等后期制作环节，录制的多声道音频文件，需要经

过三维声后期制作，再将制作好的视音频编码、打包输出成片，用于后续录播和点播使用，其传输分发和呈现环节和直播流程相似。

4.1.1 直播工艺流程

沉浸式视频直播工艺流程主要包括：拍摄、实时拼接、实时编转码、直播传输分发和终端播放。

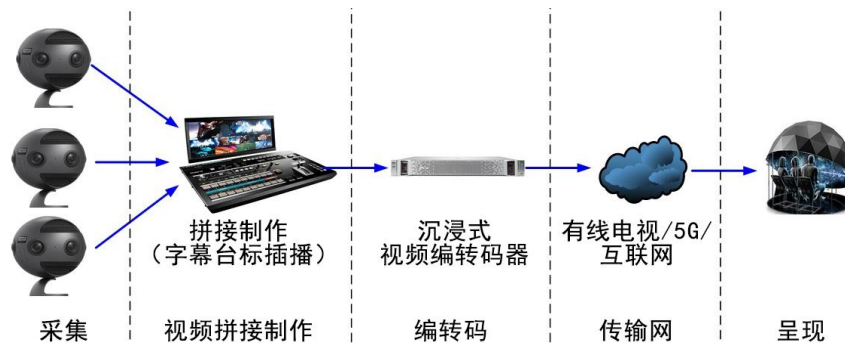


图 4-2 沉浸式视频直播流程图

(1) 沉浸式视频拍摄

直播一般采用一体化全景摄像机进行沉浸式视频的拍摄，由 4、6 或 8 个摄像头组成，从不同角度同时拍摄现场的画面。

(2) 沉浸式视频实时拼接

沉浸式视频实时拼接系统将全景摄像机中所有摄像头拍摄的画面进行实时拼接缝合，拼接成一个 $360^{\circ} \times 360^{\circ}$ 的全景视频。目前主流的专业级 VR/全景相机品牌均已开发了 VR 全景直播软件，可实现较高质量的实时拼接和编码输出。

(3) 沉浸式视频实时编转码

根据沉浸式视频直播的需要对拼接后的视频进行实时编码和转码。编转码的格式和参数要求根据全景摄像机的采集清晰度、网络传输的带宽、终端播放器的要求等进行设定和调整。同时由于用户观看的场景不同，需要编转码器同时提供多路不同分辨率和不同码率的视频流输出。

(4) 沉浸式视频直播传输及分发

沉浸式视频直播对视频清晰度的要求较高，传输和推流系统需要较高的上行和下行速率，同时支持基于 Cable 和基于 IP 两种模式。Cable 模式下，推流服

务器输出的组播信号输出到组播交换机，经过复用器后调制输出。IP 模式下，沉浸式视频流经过交换机可采取单播或者组播的方式推流。

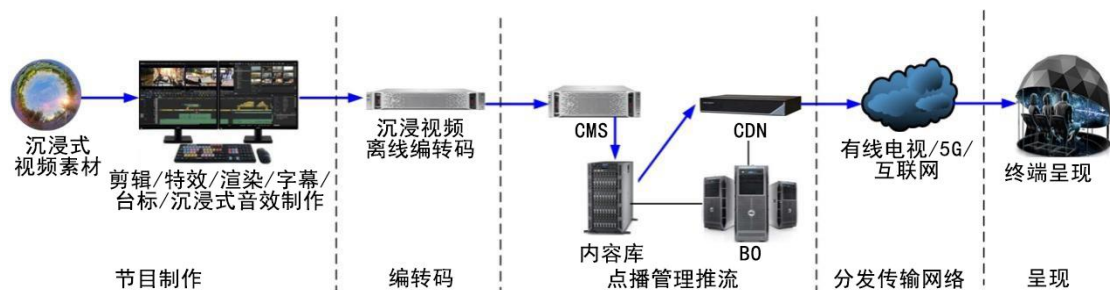
沉浸式视频直播的网络部署较为简单，只需要提供较大的下行带宽或者广播式带宽即可。比如利用广电的 HFC 网络、卫星网络就能实现比较好的覆盖和部署，还可以使用 5G 无线传输的方式，覆盖移动终端或者不方便使用有线网络的场所。商业项目的沉浸式场所可以使用光纤专线的方式，实现更优良的网络服务。

(5) 沉浸式视频终端播放

沉浸式视频的直播终端需要集成相应的直播应用，能够通过该应用获取直播分类、列表等。选定直播视频后，能从直播系统提供的信息下载相应的视频流并实时解码，最后在沉浸式投影设备、LED 等自显示屏或其他播放设备上播放。

4.1.2 点播工艺流程

沉浸式视频点播指的是用户在远端能够根据用户的需要，自由选择播放媒资系统里面的沉浸式视频。沉浸式视频点播工艺流程主要包括：后期制作、编转码点播管理、传输分发和终端播放。



(1) 沉浸式视频后期制作

沉浸式视频的后期制作与传统视频是基本一致的，如剪辑、调色、特效合成等，本质区别是沉浸式视频的编辑通常是在球面投影映射到二维平面后进行的，画面是变形状态，需要后期制作人员有一定的经验。当然，当前主流的后期制作软件（如 Adobe CC 系列等）已对 VR/全景/沉浸式视频有越来越完善的功能支持，如支持局部视野预览、外接设备预览等。

(2) 沉浸式视频编转码

编转码的格式和参数要求根据全景摄像机的采集清晰度、网络传输的带宽、终端播放器的要求等进行设定和调整。同时由于用户观看的场景不同，需要编转码器同时提供多路不同分辨率和不同码率的视频流输出。

(3) 沉浸式视频点播管理

如果沉浸式视频由本地呈现走向远程点播应用，需建立媒资存储、内容分发、节目传输、计费、密钥和用户管理等技术流程。本白皮书建议采用两个已经成熟的技术方向：一是通过网络/有线电视方式实现对远程数字视听节目媒体资源的访问，实时播放；二是通过网络方式访问远程的媒体资源系统，离线下载到本地服务器，本地服务器再播放。

(4) 沉浸式视频传输分发

沉浸式视频的点播环节需要基于 IP 的交互网络的传输，可以在传统 HFC 有线电视网络进行传输和交互，也可以通过 5G 或者无线网络传输。由于沉浸式视频的视频清晰度要求较高，对于视频的下行带宽要求比较大，对于上行带宽要求比较小。根据不同传输网络部署的需要，可以对推流系统进行集中式部署、分布式部署或者集中+分布式联合部署。

(5) 沉浸式视频终端播放

沉浸式视频的点播终端需要集成相应的点播交互应用，能够通过该应用获取视频分类和列表，并且进行视频的检索等操作。选定视频后，能获取相应的视频流并实时解码，最终在沉浸式投影设备、LED 自显示屏或其他播放设备上播放。

4.2 沉浸式视频的拍摄

4.2.1 拍摄设备的选择

(1) 多机组合拍摄

针对全球幕或环幕等特定应用场景，可采用全景视频的拍摄方式，如多机不同角度拍摄后拼接的拍摄方案，详见 3.1.1。也可根据应用场景选择采用少量相机而无需覆盖完整球面，可采用专业摄影机组合以获得高画质，同时亦可视场景采用广角、标准甚至中长焦镜头组合，较为灵活。

多机组合拍摄方案的优点是：没有因画面裁切和球面变换造成的像素浪费和图像质量劣化；画质好，特别是使用多台摄像机拼接后，图像分辨率可以达到8K、12K 或者更高。缺点是后期制作过程中需要对多机位图像素材进行调整和拼接，成片通用性差。

(2) 单机拍摄

对于半球幕、弧幕、沉浸屋等应用场景，沉浸式视频的拍摄也可以采用单机拍摄的视频拍摄方案。例如，针对半球幕等特定应用场景，可以采用单台专业摄影机/单反相机/无反相机，利用接近或大于 180° 视野的鱼镜头拍摄，获得覆盖完整半球屏幕的全景图像。



图 4-4 鱼镜头拍摄

使用单台摄像机+鱼镜头的拍摄方式，设备结构简单，构图灵活方便，拍摄得到的图像无需拼接缝合，只要配合半球幕播放要求进行格式调整即可使用。但此方式存在一定不足，如最终呈现在屏幕上的图像质量受限于单台摄像机分辨率；鱼镜头成像区域无法完全覆盖摄像机感光器件，造成感光器件有效像素的浪费；图像中心质量明显高于四周。

(3) 全景相机拍摄

对于直播这类时效性较高，不能适合采用复杂后期制作流程的应用场景，或者在拍摄过程中受场地、准备时间、器材、预算等限制，无法采用多机组合拍摄方式的，可以使用一体化全景相机进行拍摄，详见 3.1.2。

4.2.2 与传统视频拍摄的差异

全景视频拍摄记录了周围场景全部的光影信息，而传统视频只是记录了单个镜头视野范围内的信息，这造成了全景视频与传统视频在拍摄方法上的差异。具体主要有以下几个方面：

(1) 安全距离

由于沉浸式视频内容的宽视角的要求，需要对多摄像机拍摄的画面进行缝合，而相机/镜头不是一个理想的无体积的点。从光学系统的光路上看，全景视频拍摄时相邻镜头的视野重叠是在大于一定的距离才会发生，在这个距离之内或附近的对象在进行图像拼接时会出现变形、错位、丢失等异常现象。所以在全景视频拍摄时，需要保证拍摄对象位于摄像机的安全距离范围之外，且主要拍摄对象应尽量避免出现在拼接线附近，或来回穿越拼接线。不同的全景摄像机或者组合系统其安全距离是不同的，可通过测试拍摄确定。通常可以近似认为体积越大的拍摄系统其所需安全距离越大，目前常用的一体化全景相机的安全拍摄距离宜保持在 2 米以上，同时全景视频应避免在过于狭小的空间中拍摄。

(2) 消除影响沉浸感的图像内容

因全景拍摄是水平 360°（环幕）或水平 360° × 垂直 360°（全球幕）的全景区域内的图像，很多时候不可避免地会在拍摄的图像中出现摄像师、摄像机支撑设备、灯光、布景等影响沉浸式视频体验的内容。为消除这种情况，通常可利用场景内的遮蔽物（如墙体或建筑物）进行遮挡，也可以通过后期制作对特定物体与人物进行擦除、替换等，还可以使用具有远程遥控功能的摄像机进行拍摄。

(3) 照明

关于全景视频的照明主要有两个问题，一是除专业摄影机组合外，大部分全景相机在低照度条件下的成像质量很差；二是由于需要避免穿帮，无法像传统拍摄一样架设各种补光设备。

针对第一个问题，全景视频拍摄宜尽量在亮度充足的环境下拍摄；在照度条件不理想的环境中拍摄，尽量采用低照度成像素质较高的专业级一体化全景相机或专业摄像机系统。针对第二个问题，确需补光的，需要根据拍摄环境和条件，确定既不会明显穿帮又可以实现补光效果的照明架设方案，通常以伪装为主，且需针对具体场景具体分析。

(4) 拍摄视点

沉浸式视频视点的选择自由度与视频的视场角呈反比关系。视场角越小，视点选择越自由，类同于传统视频拍摄。通过适当地变焦，有限改变视场角，可以

达到在视线方向移动视点的效果。通过适当地“摇镜”，可以达到在垂直视线方向移动视点的效果。

对视场角接近 180°（半球）或 360°（全球）的沉浸式视频来说，无法实现“推镜”和“拉镜”，只能通过“移镜”的方式，依靠移动拍摄设备来改变视点。在沉浸式视频拍摄过程中，必须将拍摄机位真实地靠近被摄目标，才能获得传统视频中一些长焦镜头才能实现的效果。

在小空间拍摄场景时，由于受到镜头视差的限制，需要采用镜头间隔距离更小的拍摄设备，降低拍摄系统的安全距离，从而达到传统视频中微距镜头的效果。

(5) 观看舒适度

沉浸式视频通过扩大观众的视场角来帮助构建沉浸式的空间定位、自我运动感知以及姿态平衡。以观众为中心，沉浸式视频覆盖的角度越大，对观众带来的影响越大。

视场角较小视频的制作中，使用的运镜方式所产生的感官刺激，会随着沉浸式视频视场角的增大而成倍加强，超过人的耐受极限后会导致观众眩晕、失衡等不适。在制作视场角较大的沉浸式视频时，需要谨慎处理摇镜和移镜。镜头移动的角速度、线速度和加速度对不同年龄观者会造成不同的影响。同样的运动视觉刺激，神经发育尚未成熟的儿童及神经调节能力较弱的老年人更容易产生不适的感知。除了镜头移动所施加影响以外，画面亮度、闪光、音强、影片节奏也需要针对特定的观者进行单独设计。

4.2.3 常用拍摄方式

沉浸式视频的视角一般以人的视角为主，在传统视频拍摄的“推、拉、摇、移”四种镜头语言中，“推”、“拉”镜头不适用，能够使用的只有“摇”和“移”。一般来说，沉浸式视频的拍摄方式可分为不移动和移动两大类。

(1) 固定机位

固定机位主要分为地面固定和其他位置固定。地面固定主要利用三脚架、独脚架等支架立于地面上；其他位置固定则是指依托于墙体、天花板、舞台灯光架等现场条件，通过各种类型的支架、夹具、吊具等将全景相机侧悬或倒悬于地面之上。

(2) 移动机位

移动机位拍摄是固定机位拍摄的扩展。地面移动机位可利用移动拍摄车、轨道、平衡车或固定在移动拍摄对象上；吊装移动拍摄可利用摇臂、飞猫、蜘蛛等器材进行辅助操作，可以实现特殊的拍摄效果；此外也可利用无人机、直升机吊装进行航空拍摄。

4.3 沉浸式视频的制作

沉浸式视频分为实拍类视频和三维制作类视频，这两种视频的后期制作方法稍有区别。

4.3.1 实拍类视频的制作

实拍类的视频制作流程包括缝合素材、视频剪辑、特效制作、声音制作、调色、合成输出。

通过实拍的沉浸式视频素材一般有两种，一种是经过一体化全景相机拍摄设备直接导出相机内已有的素材（一般是直播后存储下来的视频）；一种是通过组合相机镜头拍摄的多镜头素材。第一种素材可以直接按照传统的视频后期制作方式进行剪辑、特效制作、调色、音频处理、渲染等步骤输出视频，第二种素材需要先进行缝合才能进入传统视频的制作步骤。

目前市场上用于视频缝合的主流软件有 NUK、Video stitch，图片缝合软件有 Ptgui。除了以上主流的易用软件外，VR 一体摄像机厂家为了帮助后期制作人员降低缝合时间成本，提高工作效率，推出了专门针对本厂家相机的缝合软件，提供高效的一键缝合。

目前阶段，实拍类视频主要通过投影算法将拼接后的视频，按照比例还原成 2D 视频后，在传统制作软件中完成视频剪辑及制作。此方法虽然能够较快的完成视频的剪辑，但是无法实时观看到视频剪辑后的效果，尤其是无法在视频中体验画面的层次感。部分厂家为解决此问题，采用将视频导出到 HTC vive 等 VR 头盔或 VR 一体机上进行画面监看，或通过 3D 算法，将实拍画面转成 3D 视频，模拟展示环境的方式观看效果。沉浸式视频的特效制作、调色、合成与传统视频一致，此处不再赘述。

需要注意的是，全景沉浸式视频是一种现场还原性较强的体验形式，在添加文字内容的时候，需要将文字层的投影形式与背景画面的投影形式保持一致。另外，全景沉浸式视频的画面视觉覆盖角度巨大，常规的底部字幕形式，可能无法获得观者的足够重视。

画面的稳定对优质的全景影像也至关重要。在前期拍摄过程中尽量使用三脚架、背负式稳定器、小型手持稳定器等稳定设备。如果全景拍摄设备本身具有稳定功能，也必须优先打开。在后期处理过程中，将稳定处理放在工作流的前端，根据需要选择相应全景沉浸式视频稳定插件。

此外，合成输出时，应根据沉浸式视频呈现终端的不同，来选择不同的映射方式与格式。图 4-5 分别示意了 360° 球幕（裸眼 VR 影院）、垂直半球（飞行球幕）、穹顶半球（天象厅）、倾斜 30° 半球（IMAX 球幕影院）、120° × 70° 弧面终端的视频格式在等距圆柱映射中的对应范围。

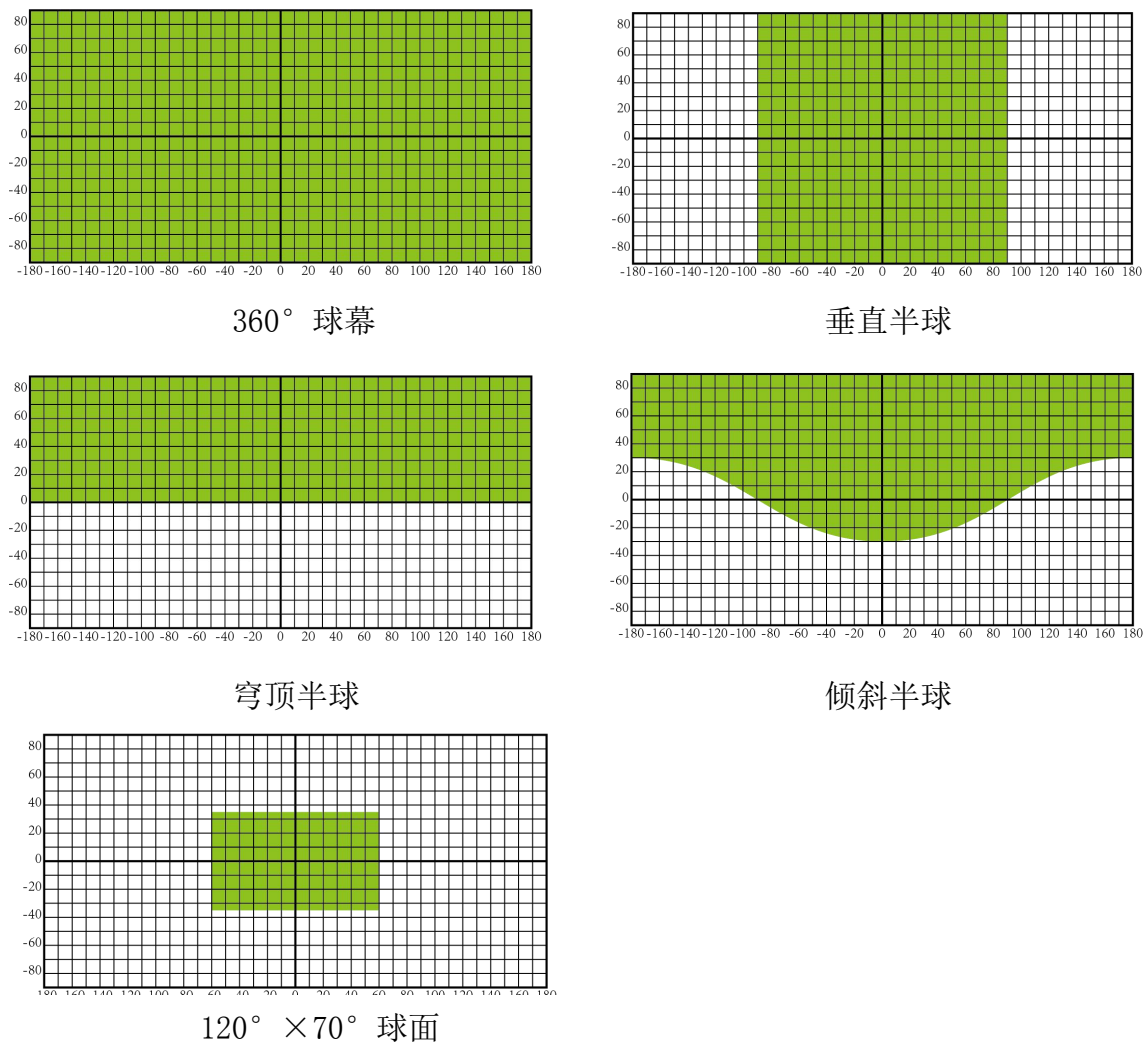


图 4-5 不同终端的视频格式对应范围示意图

4.3.2 计算机图像制作

目前沉浸式视频实景拍摄受设备、技术、场地限制，并不能适用于任意的场景，这就使得一些分辨率要求高、拍摄难度大或无法拍摄的镜头，不得不使用抠像合成或计算机图像（CG）制作的方式生成三维类的沉浸式视频。

计算机图像制作的主要流程包括模型搭建、材质设定、动画制作、灯光渲染、特效制作、配音配乐、剪辑输出等环节，其中采用的技术与传统数字电影的特效镜头制作是基本一致的。CG制作常用的软件有3Ds MAX和MAYA。3Ds MAX更适合于建筑模型与一些棱角分明的模型，MAYA适合于生物体的建模，如怪物、人体等。在渲染环节，除了3Ds MAX和MAYA自带的渲染器之外，目前多使用VRay这款渲染插件，它渲染速度快，调节方便，效果逼真，还有非常多的渲染选项，如制作全沉浸影片所必须的360度全景渲染。在后期环节，一般要把三维的视频映射成二维的序列帧文件后，再进行画面剪辑。常用的后期制作软件有After Effects、Premiere、Final Cut Pro、Combustion、Fusion等等。

需要注意的是，通过计算机合成或生成的沉浸式视频，仍需重点考虑沉浸式视频的观众体验感所带来的限制（如4.2.2节与传统视频拍摄的差异中所述），面向不同的观影环境和设备，尤其要注意虚拟镜头的移动，例如，对于固定座椅的倾斜小于30°的全景球幕影院，就不适合镜头快速移动或摇晃的画面场景等等。

4.4 沉浸式视频的传输

沉浸式视频的传输根据不同的应用场景和实际条件，可采用有线电视、5G无线网络、互联网三种方式进行传输。

4.4.1 有线电视网络

有线双向高速数据传输网络包括基于DOCSIS技术的光纤同轴混合接入网络、基于HINOC技术的接入网络以及基于PON技术的FTTx接入网络。

(1)DOCSIS网络是以有线电视光纤同轴电缆混合网络（HFC）为传输介质的高速数据业务承载技术系统，能够接入、汇聚、转发用户终端设备产生的数据流

量，实现用户终端设备与业务服务器之间的通讯功能。DOCSIS 技术标准由 CableLabs 国际标准化组织制定与发布。基于 DOCSIS 技术标准的 CMTS 设备产品与 Cable Modem 设备产品在广电行业已经大规模商用多年，承载业务稳定可靠。目前，DOCSIS 系统下行信道最高速率为 10Gbps，上行信道最高速率为 1Gbps，同时具备完善的多业务 QoS 保障机制，能够对沉浸式视频业务提供支持。

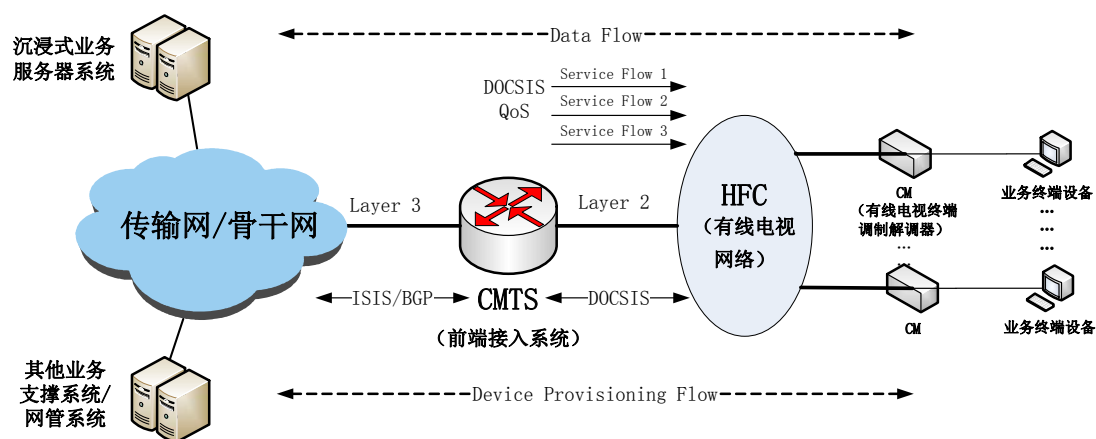


图 4-6-1 沉浸式视频有线电视（DOCSIS）组网示意图

(2) HINOC 网络是以有线电视同轴电缆网络（COAX）为传输介质的高速数据业务承载技术系统，能够接入、汇聚、转发用户终端设备产生的数据流量，实现用户终端设备与业务服务器之间的通讯功能。HINOC 技术标准由国内 HINOC 芯片厂商、设备制造商、运营商共同制定与发布，最新技术标准为 HINOC2.0。目前，HINOC2.0 系统商用产品下行信道最高速率为 1.4Gbps，上行信道最高速率为 1.4Gbps，同时具备多业务 QoS 保障机制，能够对沉浸式视频业务提供支持。

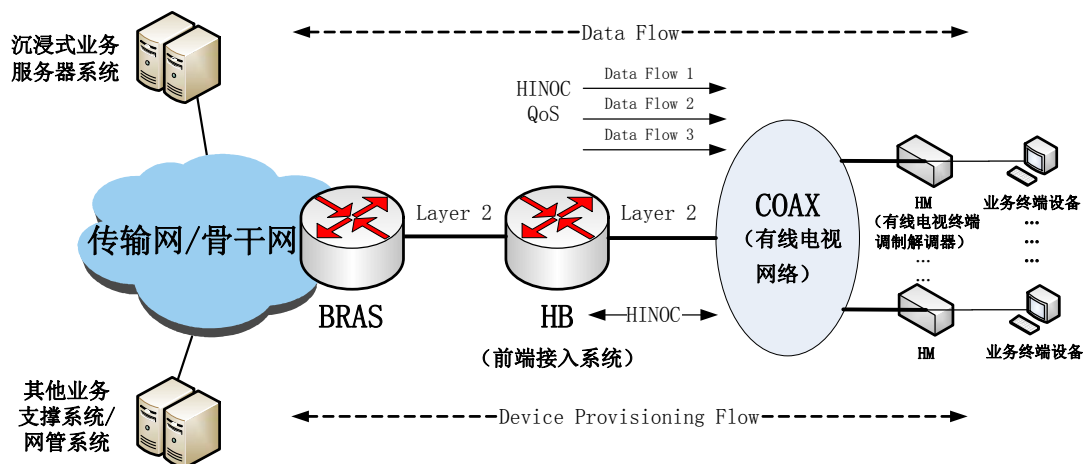


图 4-6-2 沉浸式视频有线电视（HINOC）组网示意图

(3)PON 网络是以无源光网络（ODN）为传输介质的高速数据业务承载技术系统。有线电视 PON 网络与传统电信运营商 FTTx 网络在技术及功能层面类似，可用于承载 Internet 业务、OTT 业务、IPTV 业务等用户业务。EPON/10G EPON 系统为对称传输系统。EPON 系统上/下行信道最高速率为 1Gbps，10G EPON 系统上/下行信道最高速率为 10Gbps。GPON 系统为非对称传输系统，下行信道最高速率为 2.5Gbps，上行信道最高速率为 1.25Gbps，同时具备完善的多业务 QoS 保障机制。无论 EPON 还是 GPON 系统均能支持沉浸式视频业务。

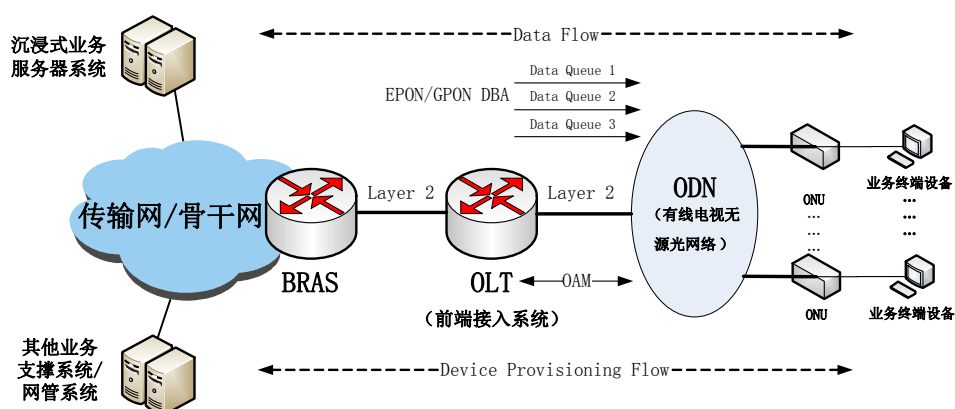


图 4-6-3 沉浸式视频有线电视（PON）组网示意图

4.4.2 5G 无线网

5G 网络作为新一代信息化基础设施，具有高速率、大带宽、低时延等特点，商用后能够满足沉浸式视频的无线传输需求。5G 网络以 eMBB、eMTC、URLLC 三大应用场景为核心，针对沉浸式视频为代表的 eMBB 类应用，5G 网络以全新的网络架构，商用网络可提供 1Gbps 左右的下行速率，同时结合移动边缘计算（MEC）、网络切片等特性，可进一步满足沉浸式视频的数据传输要求。



图 4-7 沉浸式视频 5G 传输典型网络示意图

典型 5G 传输方案如下：沉浸式视频内容通过核心网传输到承载网、接入网，再到 5G 基站。5G 基站设备使用高容量大并发的 AAU（64T64R），典型室分设备的速率可达 800Mbps，针对速率为 80Mbps 的 4K 沉浸式视频可容纳 10 位用户；典型宏站设备的速率可达 1.6Gbps，针对速率为 80Mbps 的 4K 沉浸式视频可容纳 20 位用户。通过建设 MEC，可进一步增强传输带宽，降低传输时延，带来更好的观看体验。

对于直播场景，5G 网络可支持现场媒体数据采集过程的移动性及数据传输需求。例如，针对目前的超高清电视直播业务，4K 60fps 直播业务的上行速率约为 60~100Mbps，下行通常采用更高压缩比的策略，速率约为 20~50Mbps；8K 60fps 直播业务的上行速率约为 120~250Mbps，下行速率约为 80~150Mbps。针对沉浸式视频业务，由于视频的画幅比更大、画质要求更高，速率需进一步提高。时延方面，目前广电媒体行业普遍使用的传输协议为基于 UDP+ARQ 的 RIST 和 SRT 协议，在 5G 网络典型丢包率下，RTT 时延达到 60ms 即可满足 99.9999% 的业务可靠性（帧完整传输成功率）和端到端直播时延 1s 的要求。

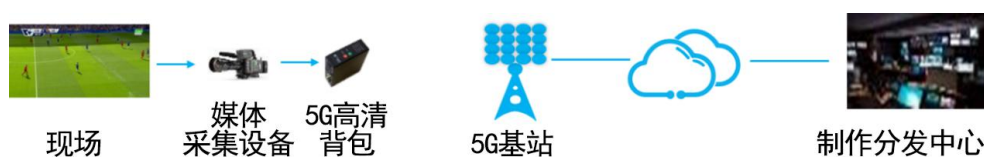


图 4-8 直播场景 5G 网络传输示意图

4.4.3 互联网

互联网可以承载包括 IPTV、流媒体等业务的视频。核心网、承载网和接入网都会对网络性能造成较大的影响，特别是复杂的接入网部分，下面只阐述同轴接入和无线接入之外最重要的无源光接入技术。为实现大带宽的视频传输，可以采用全光视频业务承载方案，基于 OTN+PON 技术的全光品质专线，可以实现快速覆盖和高性价比传输，这个方案全光专线在光传送网（Optical Transport Network, OTN）设备上增加 OTN lite 新型板卡，在光交箱/楼宇部署无源 WDM 分光器，利用 ODN 光纤，实现无源光纤覆盖；P2MP WDM 分光器实现主干光缆共享，节约光纤资源；OTN lite E2E 硬管道提供了 OTN 的高品质 SLA 体验。传送速率支持从 10Mbps~1Gbps 任意颗粒度等数据业务。

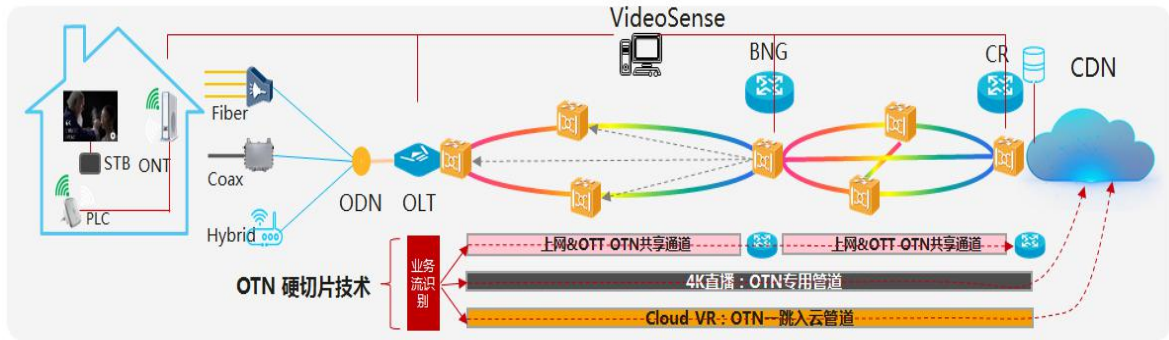


图 4-9 沉浸式视频全光网络示意图

4.5 沉浸式视频的呈现

沉浸式视频的呈现主要包括视频播放服务器、投影或 LED 自显示屏系统、还音系统、动感系统，除此之外还有信号传输及控制系统。下面以投影呈现的球幕影院为例，介绍沉浸式视频的呈现系统的技术实现重点内容。

4.5.1 视频播放系统

视频播放系统是呈现系统的信号源端，需根据具体项目的视频分辨率、帧率，以及需要的输出通道数量等确定视频服务器的选型和数量。若使用多台服务器，需要保证服务器之间完全同步，高端的媒体服务器往往在同步性上有着非常好的表现。同时，使用多台服务器可能需要事先把画面进行分割，让每台服务器仅播放画面的某个部分，这无疑会增加呈现环节的工作流程和时间。目前有些服务器支持自动画面分割，可以通过软件计算需要分割的画面，再自动输出相应的画面给投影机或 LED 等自显示屏。

对于高分辨率、高帧率视频（例如 $8K \times 8K$ 60fps）的解码，现有的 CPU 主频和并行度都很难完成。随着 GPU 技术的快速发展，主流的 GPU 都集成了硬件解码单元，这种技术可以实现将硬盘中高速码流直接读取到 GPU 中，并由 GPU 的专用硬件解码单元完成视频的解码，由于采用专用芯片技术，可以实现极高的效率和并行性。但 GPU 的硬件解码单元是相对有限的，存在计算瓶颈。为了进一步提高计算效率，高端的媒体服务器会自主研发基于 CUDA 的专用的解码器，充分挖掘 GPU 的并行计算能力，真正实现 $8K \times 8K$ 60fps 的实时解码。

沉浸式视频一般使用专业的媒体服务器，主要的外国品牌有 7th Sense、Disguise (D3)、Coolux、Green Hippo、Skyscan 等，国内的品牌有 Hirender、Light Magic (光魔)、Hecoos 等。

目前，对于品质要求高且节目内容未实现集中传送的沉浸式视频系统，其节目内容均采用拍摄制作后，内容拷贝到本地媒体服务器播放的方式，服务器的存储空间不宜小于 1T 或存放不短于 200 分钟的沉浸式视频节目。若内容为序列帧方式，需服务器具备更大的存储空间，一般不小于 8T。

4.5.2 投影系统

在小间距 LED 技术大规模使用前，特效影院的画面一般选择投影拼接融合的方式。其关键是通过光路分析，选择合适的投影机型号及亮度、不同投射比的镜头，既不影响观众观看，又保证显示区域内像素间距一致、亮度一致，同时需考虑投影机的安装位置，以及确保设备投资的性价比。

投影机的选择对画面质量的影响，将直接反映到最终的显示效果，最好是选用高亮度的工程投影机，这样能够确保显示效果的稳定性。投影机从最初的灯泡投影机逐步向激光投影机升级，目前采用的机器根据预算不同分为灯泡投影机和激光投影机，亮度一般在 5000~30000 流明之间。同时，工程机具有多种可变镜头，并且部分机型具有较为完善的自动校正功能，可以在一定程度上减少调试中由于操作不当导致的误差。

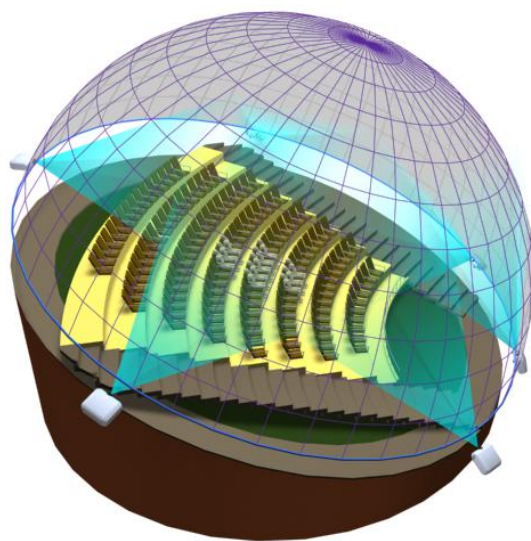


图 4-10 球幕影院的光路分析示意图

投影系统的设计还可以通过三维场景预演软件，将实际场景的模型导入到预演软件之中，在其所建构的虚拟仿真场景中添加虚拟投影机，通过投影仿真、照度测试等功能，调整投影机的参数（亮度、分辨率、镜头等），来确定最终使用的投影机及其摆放的位置。在实际的工程搭建中，可以按照预演软件中相对应的位置来架设投影机，简化搭建过程，减少反复修改投影机位置导致的搭建成本和时间成本，提升了投影画面的效果和准确性。

4.5.3 动感系统

对于飞行影院及需要座椅或平台来配合影片内容表现的，应选择动感系统。对于场地较大或观众区坡度较大的影院，尽量采取动感座椅方案；小空间、坡度小、观众少可采取平台方式。对于 3D 方式的球幕影院或没有特殊运动场景的影片内容，不宜采用动感系统方案，否则，在快速运动的影片场景下，大多数观众出现眩晕、恶心等不适，严重影响观影感受。

动感座椅或平台的驱动方式一般包括液压、气动、电动三种方式，液压方式模拟真实运动效果逼真，但对环境要求较高；气动方式成本相对低，但结构复杂；电动是相对新兴的技术方式，更加环保清洁。

运动维度是座椅或平台在空间运动的自由度。机械系统在空间运动自由度共 6 个：即上下（升降）、前后（移动）、左右（移动）、俯仰、摇摆、旋转。根据影片内容和观众感受的预期需求，在以上 6 个运动自由度上确定动感座椅和平台的运动维度。多数动感座椅选择 3~4 个自由度，动感平台选择 4~6 个自由度。若动感座椅运动维度选择太多，势必加大排间距和左右间距，影响观众容量；动感平台属整体运动，不存在座位之间相互影响问题。

4.5.4 周边系统

周边设备包括屏幕、信号处理设备、集中控制设备、线缆、接插件、机柜、电源等。

屏幕：对于球幕影院而言，银幕的设计至关重要。根据需要选择球幕的大小、结构和观看方式，球幕大小从 3 米到 30 米不等；球幕的结构可分为半包围或全包围球幕；观看方式可以分为穹顶式、倾斜式或垂直立式。另外，需确保整张投

影幕无缝，以确保画面完整性。球幕影院的银幕亮度系数一般在 0.35~0.65 之间。

信号处理设备：信号处理设备主要包括信号源类型的转换、信号的切换、信号的放大与传输等，对于画面显示非常重要。

集中控制设备：可以更好地管理各个设备，方便快捷地操作控制设备，提高工作效率和准确度。

4.6 沉浸式视频的声音方案

沉浸式视频的声音录制可以采用声场合成技术的球面声场 (Ambisonics) 传声器拾音。一阶球面声场系统 (FOA: First Order Ambisonics) 包含四个心形指向的振膜，所拾取的原始信号叫做 A 格式 (A-Format)，经过处理后得到的输出信号称为 B 格式 (B-Format)。B 格式包含 4 个通道的信息，即全方向的 W 信号、前后深度的 X 信号、左右宽度的 Y 信号和上下高度的 Z 信号，形成一种扩展后的三维化 M/S 拾音制式，从而获得水平面和垂直面的三维信息。一阶球面声场系统的缺点是空间解析度不够好，高阶系统 (HOA: High Order Ambisonics) 通过增加通道数来解决这一问题，二阶球面声场系统包含 9 个通道，三阶球面声场系统包含 16 个通道，四阶球面声场系统包含 25 个通道，如此递增。不过这也使得系统输出的数据量大大增加，运算的复杂程度也大为提高。

除此之外，沉浸式视频的声音还可以采用能记录水平和高度信息的环绕传声器组来拾音。例如，利用上下两层传声器组合来拾音，组合的模式往往借鉴已有的三维扬声器布局，如 7.1.4 的布局，下层设置 7 只传声器，上层设置 4 只。这类拾音方式获得的声音声道分离度较好，但在声场整体感和连贯度上有所欠缺。

除了现场直接拾取三维声之外，还要补充拾取更多的声音细节，采用的方法和传统电影的拾音方法类似，此处不再展开。

三维声音的编辑首先需要支持多路母线输出的音频工作站，常见的音频工作站有 ProTools、Nuendo、Reaper 等。其次，需要解决 3D 视频还放及与音频工作站的同步问题。用于声音编辑的 3D 视频还放，最好是将球形视频展开之后在电脑屏幕上还放，同时能利用鼠标调整观看角度，最终精做时再配合球形视频对声音进行精细调整。沉浸式视频的声音制作可以借用 VR 的声音制作工具，例如杜

比 VR 全景声工具里的 Video Player 可以和 Pro Tools 工作站同步。最后，沉浸式视频声音制作的重点是在三维空间内的声像定位（3D panning）和三维空间感处理（3D reverb），这两项工作需要专门的制作插件来完成。

编辑与混录完成之后，需要导出声音母版，母板格式如 Dolby Atmos Master File、ADM BWF、IMF IAB、AVS2-P3 lossless、Ambisonics 等，之后进行相应的压缩编码，然后与视频进行封装，得到 MP4、MPEG-TS、VPx 等格式的文件送往播放平台。

沉浸式视频的呈现空间的形状、尺寸，画面的形状和尺寸等等都不尽相同，对于扬声器布局，建议根据每个项目的具体情况来设计。

4.6.1 基于对象的三维声方案

杜比全景声（Dolby Atmos）是基于对象的三维声技术，在电影领域有广泛的应用，在广电、新媒体等领域也有了一些成功的应用案例。利用杜比公司的技术可以通过有线电视、地面电视、IPTV/新媒体平台、OTT 运营商、VOD 点播和直播这几种方式进行全景声节目的分发。在传输上，杜比全景声可以通过 Dolby Digital Plus JOC（Joint Object Coding）编码方式传输，JOC 传输元数据，用于音频对象的解码。在直播的应用场景中，Dolby Digital Plus JOC 可以与视频一起封装至 TS、HLS 等进行传输；另外，还可以使用文件编码器将杜比全景声的音频信号与视频封装为 MP4、TS 等文件格式，用于点播节目传输。家庭终端用户可以通过布置 5.1.4/7.1.4 等还音系统、条形音箱 Soundbar 或者支持杜比全景声的电视和手机来获取三维声的体验。杜比推出了成熟的制作工具帮助声音后期制作人员完成声音的编辑、混音、母版制作及内容质检等工作流程。Protools 和 Nuendo 均支持杜比全景声声像定位 Pan，混音师可以灵活使用声床和对象，最多支持 128 个声音元素的渲染。

DTS-X 也是基于对象的三维声技术，主要应用于数字电影、主题公园的飞行影院等项目中。DTS-X 不受固定扬声器摆位或具体声道信号的限制，可以根据呈现空间的模型来设计扬声器的布局，因此能较好地适应球形影院的需求。目前，DTS-X 支持 32 通道输出，配合相应的音频处理器，可以扩展到 64 通道。

HOLOSOUND 是国内自主研发的基于声道+对象+声场的三维声技术，主要应用于数字电影领域。HOLOSOUND 的内容制作采用插件与内容制作系统结合的形式，可以实现 3D 音频制作的监听、录制与回放。其中插件可以与主流的音频工作站兼容。在应用端，HOLOSOUND 可根据不同的应用场景进行编码输出，最多可支持 128 通道，支持 5.1.4, 7.1.4, 22.2 等。

AVS2-P3 是中国自主知识产权的三维声技术，能够同时支持多达 128 个声道和 128 个声音对象的高效编码，包括通用音频编码（CBA）、无损音频编码（lossless）和对象元数据编码（OBA）等三个层次。目前，AVS2-P3 及其应用解决方案 WANOS 全景声已经应用于电影、有线电视、IPTV、OTT、VOD、直播、有声读物和音像节目数字出版物等领域。

4.6.2 基于场景的三维声方案

Ambisonics 麦克风阵列录制的声音场景信息，可以使用 MPEG-H 技术进行制作、编码和传输。MPEG-H Audio 音频标准（ISO/IEC 23008-3，第三部分音频）是国际动态图像专家组 MPEG 最新开发的针对下一代广播电视流媒体应用的音频系统，目前已经实现了商用。MPEG-H 能传输基于声道的信号或基于场景的声音元素，还支持基于对象的静态或动态元数据的编码和传输，最多可支持 128 个对象的渲染，在广电以及流媒体应用中，为了降低复杂度，实现家庭端的实时还放与渲染，MPEG-H 的基础子集可传输 32 个对象，同时解码 16 个声道（在特别模式下可支持 22.2 声道解码），能够支持多种三维声格式。

在后期制作环节，Fraunhofer 研究院推出了免费插件 MHAPi，能够实现在 Pro Tools 上生成 MPEG-H 元数据，达芬奇调色软件也能够支持 MPEG-H 音频的混音，元数据生成以及编码，生成 MP4 文件，实现节目点播应用。在还音系统上，MPEG-H 采用 ITU-R BS. 2051 标准对还放扬声器位置进行定义，按照球形空间的模型对声音对象进行定位以及渲染，可以用于球幕影院这一类球形空间的沉浸式视频项目。在家庭端可以通过固定的扬声器布局，例如 5.1.4/7.1.4，或者条形音箱 3D Soundbar 的方式实时解码渲染还放三维声内容。同时可根据还放条件，自动将多种三维声格式内容与之相匹配。

4.7 呈现空间的音视频指标要求

沉浸式视频呈现的空间主要为球幕影厅、环幕影厅及沉浸屋，需要对呈现空间的亮度、对比度、声道数量、扬声器布置、声压级、频率响应等指标做出约束。球幕影厅的音视频指标可参照规范《科技馆球幕影院功能配置与放映技术要求》和国际巨幕协会制定的沉浸式数字影院规范；环幕影厅及沉浸屋的音视频指标可参照电影院相关规范。

4.7.1 亮度

在沉浸式视频的球幕呈现空间里，画面或显示介质“包裹”着观众，画面成像介质之间存在光线反射干扰，如果画面亮度过高，导致厅内的反射光互相干扰，反而影响画面对比度实现，观影效果差。因此，建议普通模式画面亮度 $17.1 \sim 20.5 \text{cd/m}^2$ ，3D 时约 6.8cd/m^2 较稳妥。采用如 LED 等自显示屏的影院，画面亮度可不受此约束，但此类放映方式正处于市场试验推广期，各类标准并未严格约定，但是考虑到人眼疲劳等生理因素，画面亮度不宜过高。

4.7.2 对比度

画面对比度是很重要的影像技术质量指标，顺序对比度与显示设备性能关联度较大，相对容易实现；帧内对比度不仅取决于设备性能，且受内饰材料光学特性、杂散干扰光、画面关系以及银幕反射影响较大。建议沉浸式球幕呈现端的画面对比度满足：顺序对比度应不低于 2000:1，帧内对比度应不低于 10:1。

4.7.3 声道数量

沉浸式视频系统的声音需要实现不低于 5.1.4 声道的声音系统，因此，声道数量应该不少于 10 个。根据实际经验，16 个声道方案是能满足多数项目需求和声音沉浸感，尺度较大的特效影院需要部署更多的声道。

4.7.4 扬声器布置

主声道扬声器的布置需保证其声中心处在主画面高度的 2/3 附近，确保声画位置对应；环绕效果声扬声器布置位置要能覆盖到大多数观众，除非内容对声音的位置表现有特殊需求，一般不应布置过高或过低；尽量将每个声道对应的扬声器集中布置，除非因环境尺度大需要一个环绕声道设置多只扬声器来覆盖多排观众。

4.7.5 声压级

根据国内数字影院技术标准和国际巨幕协会对沉浸式数字影院做出的规定，沉浸式影厅各声道的基准声压级确定为 85dBC，主扬声器的峰值声压级为 105dBC（参考电平为-20dBFS），次低频扬声器的峰值声压级为 110dBC。

4.7.6 频率响应

根据国内数字影院技术标准和国际巨幕协会对巨幕影厅和球幕影厅的指标要求，规定主扬声器的频率响应需满足 30Hz~16kHz（+3/-6dB），环绕扬声器的频率响应需满足 40Hz~16kHz（+3/-6dB）次低频扬声器的频率响应需满足 25Hz~120Hz。

4.8 呈现空间的环境指标要求

球幕、环幕及沉浸屋是沉浸式视频呈现的主要形式，球幕影厅、环幕影厅及沉浸屋是声音和画面重放的主要环境。为了获得优良的声画重放效果，沉浸式视频需要对呈现空间的声学和环境光进行必要的约束。

4.8.1 声学环境要求

优良的声学环境是声音沉浸感的基础。声学环境由两部分组成，即室内音质和噪声控制。其中，室内音质涵盖了混响时间设计、声缺陷的避免等内容，噪声

控制则需综合考虑房间围护结构的隔声设计（空气声、撞击声等）以及机电设备的噪声振动控制（如暖通消声、机组减振等）。

(1) 音质

室内音质是反映沉浸式视频最终声音质量的关键。为了获得良好的室内音质，需要对混响时间、混响时间频率特性和声场均匀性进行控制，以避免出现声学缺陷。

参照规范《电影院建筑设计规范》（JGJ58），沉浸式影院应做到：

a. 混响时间及混响时间特性

■ 沉浸式影厅的混响时间，应根据观众厅的实际容积按下列公式计算或从图 4-11 中确定：

500Hz 时的上限公式为：

$$T_{60} \leq 0.07653V^{0.287353}$$

500Hz 时的下限公式为：

$$T_{60} \geq 0.032808V^{0.333333}$$

式中 T_{60} ——观众厅混响时间(s)；

V ——观众厅的实际容积(m^3)。

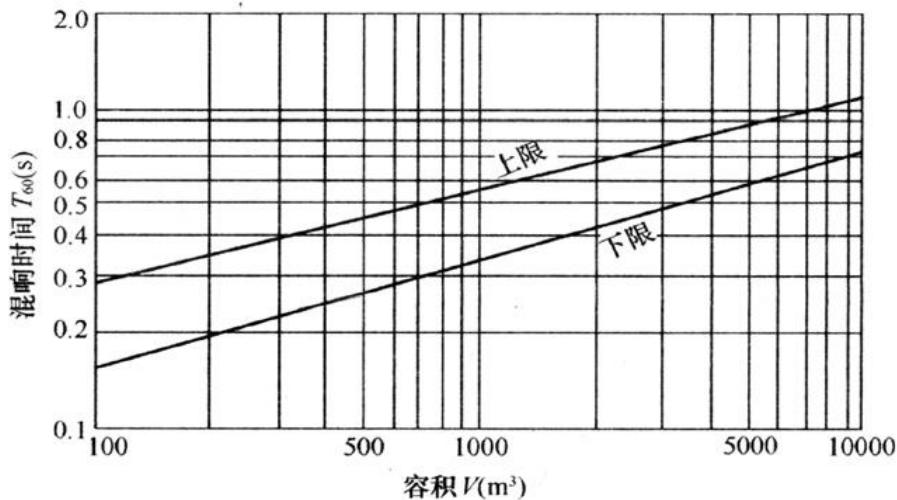


图 4-11 混响时间与房间容积对应关系图

■ 沉浸式影厅混响时间的频率特性（相对于 500Hz 的比值）宜符合图 4-12 的规定。

| f (Hz) | 63 | 125 | 250 | 500 | 1000 | 2000 | 4000 | 8000 |
|-------------------------|---------------|---------------|---------------|------|---------------|---------------|---------------|---------------|
| T_{60}^f/T_{60}^{500} | 1.00~ 1.75 | 1.00~ 1.50 | 1.00~ 1.25 | 1.00 | 0.85~ 1.00 | 0.70~ 1.00 | 0.55~ 1.00 | 0.40~ 0.90 |

图 4-12 混响时间频率特性表

b. 声缺陷

观众厅内听音区不得出现声聚焦、共振和回声，包括多重回声、颤动回声等缺陷。这需要针对不同重放空间的声学特性，如球幕和环幕影院易产生的声聚焦，矩形空间易产生颤动回声等，分别进行声学分析和处理。

(2) 本底噪声

影院的围护结构隔声及相关设备的噪声振动，是控制影院本底噪声的关键。参照规范《剧场、电影院和多用途厅堂建筑声学设计规范》（GB/T 50356-2005），沉浸式影院的本底噪声应控制在不低于 NR30 的要求。

a. 隔声

隔声的重点是通过制定合理的建筑墙体、楼板等围护结构，确保影厅不会受到外界（如交通噪声）或相邻噪声源（如机房）的干扰，确保与其他相邻影厅或声学敏感用房互不干扰，确保影厅特殊设备（如投影机等）的噪声不对观众观影造成不良影响。具体隔声做法的选择，主要依据影厅周围噪声源情况和影厅内背景噪声的限值目标。

b. 设备噪声

设备噪声主要是空调管道的消声，应该避免因风噪引入厅内影响观众观影。

c. 设备振动

设备振动包含建筑设备和厅内设备的振动，应该避免建筑设备振动（如各类泵、机组、变压器、水管等），通过建筑墙体、楼板等刚性构件形成的固体传声对观影环境形成振动干扰及二次噪声干扰。

4.8.2 照明技术要求

对于上述环境空间，照明技术需满足以下要求：

- 场所内应优先选用节能型光源和高效灯具；

- 场所内环境光的照度不宜过高，宜低于 150lx，同时可以通过交互灯光互动装置艺术打造“沉浸式”的体验感觉。不同的场所内应充分利用光的强度、颜色、氛围三个主要特质，创造出与众不同的艺术特性；

- 对于设置观众厅的观演空间，其踏步灯或座位排号灯的供电电压应不大于 36V 的安全电压；

- 设置疏散应急照明的场所，其通道上地面的最低水平照度不应低于 1.0lx；观众厅内地面的最低水平照度不应低于 3.0lx；疏散应急照明的设置应符合现行国家标准《建筑设计防火规范》（GB 50016）的规定。

4.8.3 暖通技术要求

对于上述这些环境空间，暖通技术需满足以下要求：

- 应设有通风换气和空气温湿度调节设施，保证影厅的温度，夏季不低于 26℃，冬季不低于 18℃；相对湿度，夏季不高于 70%，冬季不低于 30%；风速，夏季不低于 0.30m/s，冬季不低于 0.20m/s；新风量不小于 25m³/人·h。

- 应有空气质量保证设施，空气质量应符合《室内空气质量标准》（GB/T 18883）的相关要求，且颗粒物 PM10 不应大于 50 μg/m³、PM2.5 不应大于 35 μg/m³。

4.9 沉浸式视频系统小结

近年来，随着沉浸式视频不断地应用探索及各种系统、设备和软件性能地提升，已形成基于不同应用场景的端到端解决方案，多数方案已趋于成熟化、产品化，并在局部领域形成了专业市场，这为沉浸式视频未来在各领域地实际应用奠定了较好的技术基础，各系统的技术状况可归纳如下。

4.9.1 拍摄制作

目前已摸索出符合沉浸式视频特点的拍摄制作技术路线，市场中已有的系统设备、软件基本能胜任各种拍摄制作要求。

在拍摄方面，多机组合、单机、全景相机等拍摄方式，在拍摄安全距离、照明、拍摄视点、镜头应用方式等诸多影响沉浸式视频效果的方面，总结出了经验性的技术规范。

在制作方面，已形成基于实拍和计算机图像两类制作的完备制作流程。实拍类的制作流程包括素材缝合、三维图像映射、视频剪辑、特效制作、声音制作、调色、合成输出等环节；计算机图像类的制作流程包括模型搭建、材质设定、动画制作、灯光渲染、特效制作、配音配乐、剪辑输出等环节。

在指标方面，单镜头的画面质量已达到 8K 10bit 30fps，全景一体机的拍摄质量已达到 11K 8bit 30fps；可拼接出完整的 360° 全景画面。三维声技术已完全支持至少 32 通道输出，实现 5.1.4、7.1.4、22.2 等声道制式。

未来的技术发展，需要集中在镜头、传感器、摄影机集成度、算力等性能方面进行技术提升。

4.9.2 编码

目前，已经成熟的视频编码标准 H.264、H.265、AVS+、AVS2 基本已胜任大多数沉浸式视频应用的内容制作和播放，正在制定中的下一代编码技术标准 H.266、AVS3 在编码效率、适应 8K 分辨率、多路视频不同码率及新型网络传输的适应性等方面均有解决方案，可完全覆盖未来沉浸式视频各种应用下的内容制作和播放要求。

但由于沉浸式视频存在视场角大、画面存在主辅区域、内容可能附加动感信号等特点，未来需要深入研究沉浸式视频的画面和声音质量要求、核心要素和特点，借鉴多视角编码 MVC 和点云技术，对现有标准进行改进，形成针对沉浸式视频的编码方案，以满足沉浸式视频的功能和性能质量要求。

4.9.3 传输

沉浸式视频的传输是其在应用广度层面形成的新需求，特别是在现有应用端的内容丰富性、大型赛事活动的沉浸式体验应用等方面，视频传输是必要的手段。现有成熟的视频传输方式均可对常规沉浸式视频业务实现良好的支持。对临时性

或时效性要求高的应用，未来的低时延 HFC 和 5G 无线网络技术已经可以达到 1ms 时延和上行 1Gbps 的传输速率，也能实现对业务的良好支撑。

沉浸式视频内容节目传输，现有两个成熟的技术方向：一是通过互联网或有线电视网络，访问远程数字视听节目媒体资源系统，实时播放；二是通过网络方式访问远程的媒体资源系统，离线下载到本地服务器，本地服务器再播放。

未来，沉浸式视频内容传输需要解决如下问题：媒资管理、版权交易、内容点播分发平台、密钥分发、用户权限管理、计费等等。同时，完善的节目加密和授权方式是方便节目交流传播的前提，对保证版权方利益至关重要。

4.9.4 终端呈现

沉浸式视频在终端呈现层面的应用起步最早，形式最丰富，技术最成熟，市场最活跃。

(1) 显示技术方面

投影技术在几十年的发展中已从 LCD、CMOS 发展到 DLP，光源从卤素灯、氙灯发展到激光和 LED 光源，单台投影机物理分辨率均已达到 2K 以至 4K 指标，无论在亮度、色彩还原、清晰度等方面已充分适应沉浸式视频的各种应用场景。同时，针对沉浸式视频的各种呈现形式，投影技术在投影画面拼接、边沿融合、图像映射、亮度均一化、自动校准、透视校正等方面不断进步，形成观众感受体验不断提升的现象，也成为推动沉浸式视频市场的最强因素。

(2) 播控技术方面

播控技术和产品已基本成熟，能支撑各种内容格式、各种声画质量的影片播放，能实现精确同步，并协同控制环境特效及动感平台的运动，支持自动画面分割，支持实时解码以及无压缩序列帧播放。

(3) 环境技术方面

沉浸式视频呈现空间与其他影视及演出呈现空间有诸多相似之处，各种国际国内标准对呈现环境在室内声学、照明、振动、空调、配电及环境质量等方面的技术要求和指标控制手段完全成熟，只需在应用中根据沉浸式视频特点进行有针对性地归纳和细化。

(4) 核心指标方面

沉浸式视频在终端呈现形式上类似电影放映，又类似现场演出，又类似 VR。其核心指标是根据呈现方式的特殊性，借鉴其他呈现形式的要求总结形成的，包括视场角、动态范围、色域、帧率、色深、分辨率、亮度、对比度、声道数量要求、声压级、频率响应等诸多内容。围绕沉浸式体验的要求，现有产品技术已能很好保证这些核心指标的实现，并在分辨率、亮度、三维声等方面不断提升。

(5) 新技术方面

近年出现的采用激光光源、120fps 帧率、单机 30000 流明输出的投影机，微间距 LED 自显示屏，基于对象的三维声，6 自由度的动感平台等诸多新技术，为沉浸式视频的推广应用提供了市场卖点。

(6) 终端呈现在未来技术发展中需要解决的问题

基于点播方式的应用，需解决基于特定的呈现方式下，播控系统在硬件和软件层面对不同内容和不同终端的自适应技术，这是沉浸式视频市场广泛应用的关键技术。

另外，需研究如何利用多显卡 GPU 并行计算，合理分配计算资源，增加单机渲染输出能力，从而实现高分辨率、高帧率视频（例如 8K×8K 60fps）地顺畅播放。

投影技术需要在融合带处理、画面边缘融合、投影机镜头光源衰减和镜头畸变、画面变形拉伸等方面进一步优化。

动感系统设计需要更加符合人体工程学，减轻人体恶心眩晕等不适反应，使观众沉浸式体验更轻松、更舒适。

5. 沉浸式视频的应用场景

由于沉浸式视频带来的特殊影像呈现方式，目前已广泛应用于体育赛事、科技场馆、展陈场馆、游艺场馆、商业场馆、家庭终端、教育培训、虚拟拍摄等领域。

| 场景类型 | 应用情况 |
|-------------------|--|
| 体育赛事直播/ 重大活动直播 | 通过采用全景摄像技术进行广播电视内容的采集录制、视频内容创作、视频内容拼接、编解码、分发传输等，并使内容在球幕终端上呈现，让观众获得人与内容场景互动的体验。 |
| 科普场馆 | 应用形式为球幕影院，显示系统基于投影来实现。沉浸式视频通常会呈现出天空、海洋、森林等大自然景象，配合三维声与其他感官特效，创造出沉浸的氛围。 |
| 展陈场馆 | 应用形式主要有 L 型折幕、三折幕、沉浸屋 CAVE 空间、球幕等，显示系统基于投影来实现，在党建展陈中已广泛应用。 |
| 游艺场馆 | 应用方式通过飞行影院、球幕影院等设备布置实现娱乐游玩和科普教育的功能，显示系统基于投影或 LED 等自显示屏来实现。 |
| 商业场馆 | 应用形式通常是利用独特的穹顶结构，结合数字 3D 投影技术打造出超震撼的沉浸式体验空间或是投影艺术互动装置，已广泛应用到新品发布会、商业会议、宣传活动、场秀、派对、音乐会等商业活动中。 |
| 教育培训 | 应用方式通常为球幕虚拟仿真系统，形式为弧幕或三折幕，系统基于多通道投影技术和三维同步成像设计，适用于虚拟仿真教学实训、视景仿真演练等。 |
| 家庭终端 | 应用形式可为直径 1.5-3 米的小型球幕，用单台投影或多台投影机拼接实现画面的呈现，用环绕声系统实现声音的包围感，还可以结合动感飞行乘骑设备，通过吹风、水雾和不同 |

| | |
|------|---|
| | 的气味以及震动，呈现给观众沉浸式的视听及体感体验。 |
| 虚拟拍摄 | 将拍摄背景实时呈现投影碗幕或 LED 三折屏幕上，营造出沉浸式空间，演员在大屏幕前进行表演，具有非常强的视觉冲击力和沉浸感。可以应用在虚拟演播室、影视节目拍摄中，不仅为拍摄带来便利，同时也为观众带来裸眼 3D 的效果。 |

5.1 重大活动应用

5.1.1 项目概述

新华社、中央广播电视总台等多家媒体选择在传统报道的基础上，加入 VR 直播这一新形式，在线下搭建球幕影院，向市民开放，对国庆阅兵等重要会议和重大事件进行直播。

5.1.2 整体架构

新中国成立 70 周年阅兵活动中，低时延、高速率的 5G 网络全程覆盖天安门、长安街两侧，有了 5G 网络，多家媒体选择 VR 全景相机设备进行 VR 直播。

新华社客户端构建了 VR 直播矩阵，依托 VR+5G+8K，现场采用 12 机位将这场规模空前的盛宴即时 VR 传输，提供导播切换视角与自由选择不同机位，为上千万在线观看的观众营造身临其境的现场感。



图 5-1 国庆 70 周年阅兵 VR 直播矩阵

5.1.3 部署方案

阅兵式 5G+VR 直播，现场根据不同题材，信号源部分部署 12 台 Insta360 Titan 和 Pro2 全景相机，提供导播切换视角与自由选择不同机位，采集 6 或 8 路视频流，经过千兆网线传输到拼接服务器，推流到本地服务器后，通过千兆路由，利用 5G 网络将 RTMP 的视频流实时传输到观看终端。



图 5-2 国庆 70 周年球幕直播系统示意图

5.1.4 应用效果

新华社客户端直播超 100 万人在线观看，球幕开放时间内接待近 400 人次，整体传播曝光量超 3000 万。



图 5-3 国庆 70 周年球幕直播采集及呈现现场

5.2 游艺场馆应用

5.2.1 项目概述

飞行影院是沉浸式视频在游艺场馆的创新应用。飞行影院和球幕影院具有天然的沉浸感和立体感，不需要带立体眼镜，适合多人同时观影，这种形式在娱乐、科普、主题等活动均有广泛的应用，其中，天文类教育节目多采用现场互动的实时操作，其他多采用制作好的影片回放形式。飞行影院一般为垂直的球幕影院，可以通过多台投影拼接或 LED 球形屏来实现。观众坐在悬挂的座椅里，观看到超大视角的画面，结合动感飞行乘骑设备，六自由度的运动平台，配合影片节奏通过吹风、水雾和不同的气味特效以及量身定制的三维声，呈现给观众全新的视觉、听觉、感觉三位一体的凌空飞翔的互动体验。

5.2.2 整体架构

某飞行影厅由直径 20 米的 LED 球幕、8K×8K 分辨率的节目播放系统、20.2 声道沉浸式音响及 72 座 6 自由度动感平台组成，还包括等候区及预演厅的音视频播放及相应灯光控制系统，统一由演出控制系统通过网络协议进行时序控制。

5.2.3 部署方案



该飞行影院由演出控制系统通过网络协议统一控制音视频播放系统、乘骑座椅运动系统、灯光、特效等系统的时序动作，采用洛普 LED 球屏替代了传统的投影拼接方案。LED 球屏实现了面板的均匀拼接，在低亮度下依然有良好的色彩表现，透声电路板的设计实现了按照声场分布的需要，在屏幕后方合理摆放音箱，系统兼容等距鱼眼以及柱面展开两种主流球幕节目源，实现了 8K×8K 60fps 非压缩图像及 20.2 声道的 DTS-X 沉浸式环绕声音的播放。

5.2.4 应用效果

采用 LED 屏幕的飞行影院，在影片播放亮度、对比度和色彩、均匀度等指标优于传统的投影拼接方案的飞行影院，增强了游客的体验感，提高了观赏性，同时，降低了维护费用，提升了平均无故障运行时间。



图 5-5 飞行影院现场

5.3 展陈场馆应用

5.3.1 项目概述

沉浸式视频在展陈场馆应用主要以现在的规划馆、展览馆、博物馆以及企业展厅居多，主要的应用方式有 L 型折幕、三折幕、CAVE 空间，有时也会结合实物沙盘共同投射的方式。

以某感官式体验馆为例，该馆具备数个主体空间，提供电影式叙事、循序渐进的体感实验，通过体感、音乐、色彩、影像的变化，给观众从听觉，到嗅觉、听觉、视觉、体感的融合体验。观众在包裹感的画面下，可随意依靠在地面移动沙发，体验置身宇宙中央或浮游海上。穹顶画面采用音乐可视化的技术手段使得音乐、影像同步，突破感官边界。

5.3.2 整体架构

由一台媒体服务器主机采集外接的互动程序主机画面，将处理好的数据输出到两台负责显示的媒体服务器上，再输出给投影机 and LED 大屏，形成沉浸式投影场景布置。

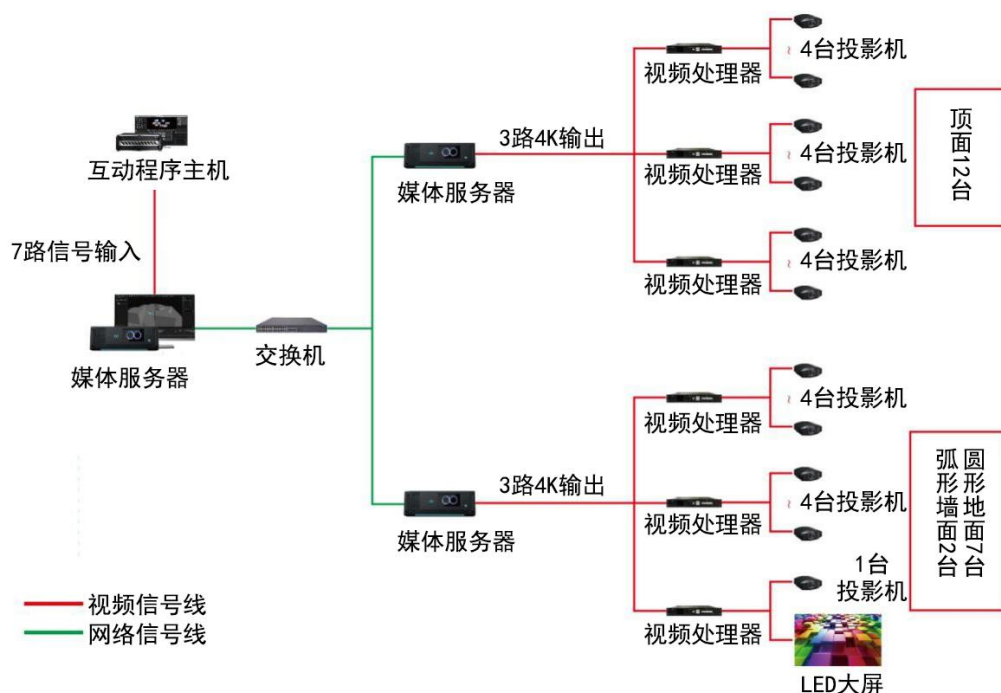


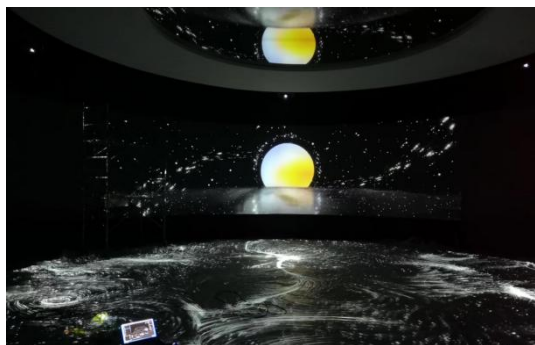
图 5-6 感官式体验馆系统示意图

5.3.3 部署方案

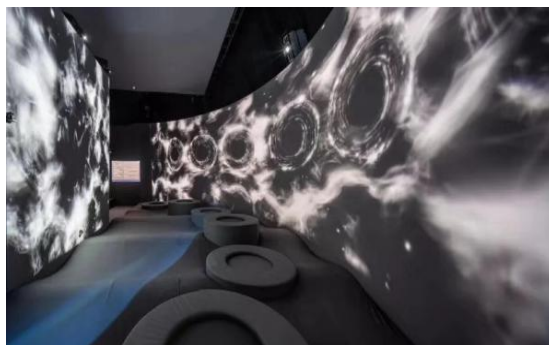
媒体服务器主机通过网线将数据输出到交换机，然后分配给两台显示端媒体服务器。每个媒体服务器输出 3 路 4K 视频信号，通过拼接处理器，将每一路 4K 视频信号分割为 4 路 2K 视频信号。将每个 2K 信号分配给具体的投影机来进行画

面的输出。顶面使用了 12 台投影机进行拼接融合，并采取背投的方式；圆形地面使用 7 台投影机，弧形墙面使用 2 台投影机，还有一路视频输出到 LED 大屏。

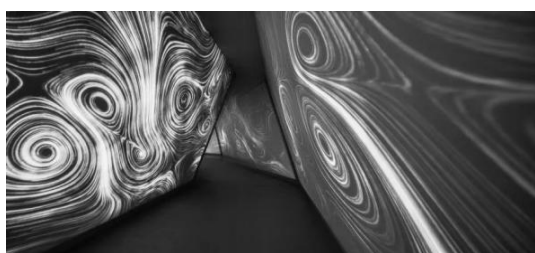
5.3.4 应用效果



场景 1



场景 2



场景 3



场景 4

图 5-7 感官式体验馆场景 1-4

5.4 教育培训应用

5.4.1 项目概况

沉浸式视频在教育培训的应用非常广泛，以空中客车 A320 模拟驾驶项目为例，通过 3 台 DLP 激光投影机投射到弧形幕，让一个室内场景虚拟还原出天空飞行的场景，体验者可以在仿造的飞机驾驶舱里体验驾驶的感觉，高分辨率的投影机可以营造出逼真的天空景色。

5.4.2 整体架构

播放服务器将画面分配给 3 台松下激光投影机，通过拼接融合系统组成一个完整画面输出。投影机通过可网络进行控制。



图 5-8 空中客车项目系统示意图

5.4.3 部署方案

通过 3 台投影机将融合的画面投射在一个完整球带幕上，球带幕的画面再通过球面镜反射到操作人员眼睛里。投影机部署在球面镜上方，球面镜和球带幕按照一定角度布局，保证操作人员看到画面为正常标准的画面。这种安装方式最大的好处可以在较短距离内投射出比较大画面，让模拟效果更佳。

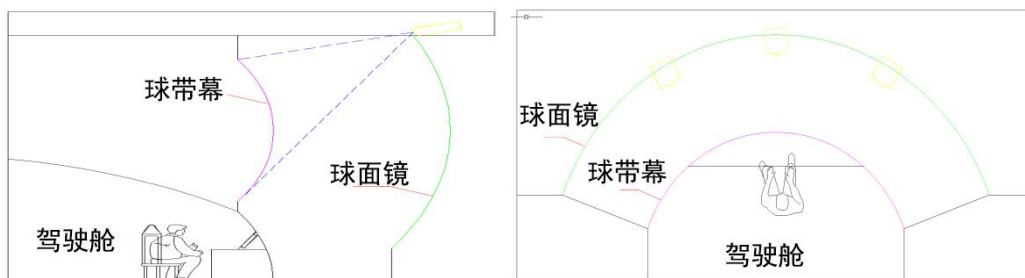


图 5-9 空中客车项目光路分析图

5.4.4 应用效果



图 5-10 空中客车项目现场

5.5 家庭终端应用

5.5.1 项目概述

家庭终端的应用具备场地小，沉浸感、互动性强等特点。球幕播放具有强烈视觉冲击力的画面，可以实现多人互动，外围观众能够与体验者同步看到相同的画面，同样具有视觉冲击力。体验者结合动感飞行乘骑设备可以佩戴或不佩戴VR眼镜，乘着由六自由度的运动平台特效，通过吹风、水雾和不同的气味特效以及量身定制的沉浸式震动强烈的真实感，呈现给观众视觉、听觉、感官及震动的互动体验。

5.5.2 整体架构



图 5-11 家庭球幕系统示意图

Fulldome Pro 的沉浸式家庭影院终端由一个 3 米球幕和一组动感骑乘装置组成，3 米球幕呈现 3K 分辨率的节目和 5.1 声道沉浸式音响以及 4 座 6 自由度动感平台组成，可配合或不使用 VR 眼镜，统一由球幕控制系统通过网络协议进行时序控制。

5.5.3 部署方案

Fulldome Pro 的媒体播控服务器通过网络连接影片数据库，下载并播放沉浸式影片，并统一控制投影系统、乘骑座椅运动系统、灯光、特效等系统的时序动作。单台主机带动至少 4 台投影机，其拼接软件实现屏幕的均匀拼接（黑白度、饱和度、图像边缘切割等），在黑场部分（如星空、海底画面）呈现了专业影院的色彩表现。音箱悬挂在圆形银幕支架上，形成了 5.1 声道环绕声。

5.5.4 应用效果



图 5-12 家庭球幕现场

5.6 虚拟拍摄应用

5.6.1 项目概述

沉浸式视频在虚拟节目制作的应用领域，目前比较前沿的技术是扩展现实（XR）技术。通过 XR 技术将拍摄背景同步呈现在投影碗幕或 LED 三折屏幕上，演员在大屏幕前进行表演，再用摄像机进行拍摄，真正的所见即所得。

5.6.2 整体架构

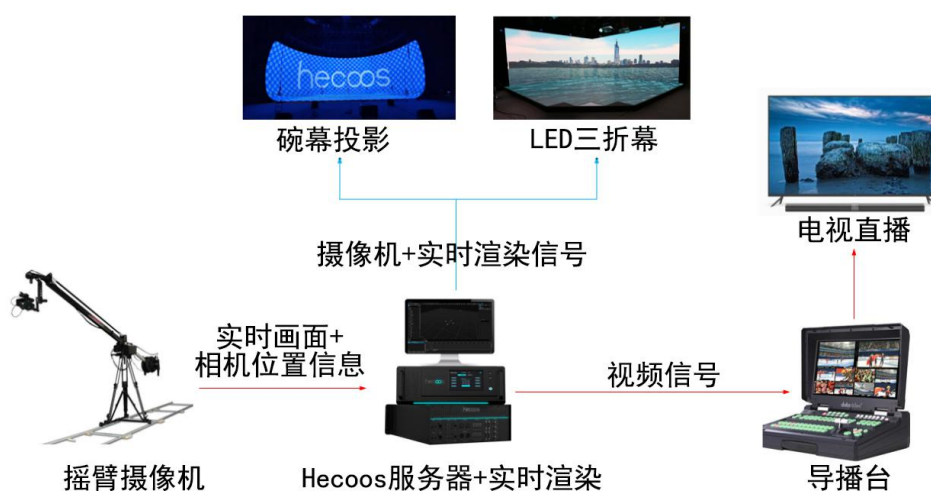


图 5-13 虚拟拍摄系统示意图

虚拟拍摄系统由集成了实现渲染功能的媒体服务器，负责拍摄的摇臂摄像机、负责呈现的碗幕投影或 LED 三折屏幕，以及导播台等设备组成。

5.6.3 部署方案

(1) 摄像机追踪系统

利用摄像机追踪系统为摄像机所在摇臂的每个关节安装编码器，经计算后将编码信息转换为描述真实摄像机的 8 组参数，用于确定摄像机的位置、姿态及镜头参数。

(2) 实时渲染与前景叠加

摄像机追踪系统通过以太网将真实摄像机的运动信息实时传输至 Heccos 媒体服务器，媒体服务器对该信息进行处理，然后与 Notch 实时渲染引擎完成无缝对接，驱动引擎进行画面实时渲染，这些画面分为前景和背景两部分，通过媒体服务器，前景可叠加在导播台的输出画面上，背景可输出至投影机进行透视投射。

(3) 透视投射与播控

媒体服务器将实时渲染引擎渲染出的画面通过透视投射映射至屏幕，为画面建立摄像机观测视角下的正确透视关系。直播者可以与真实变幻的环境进行实时互动，从而有效提升节目表演的趣味性。

(4) 技术关键

XR 技术应用直播场景的关键，在于媒体服务器将真实摄像机数据、透视投射数据、实时渲染引擎工程中的摄像机，三者坐标系保持一致并正确匹配。这个前提贯穿工作流程始终，为 XR 的完整实现奠定基石。

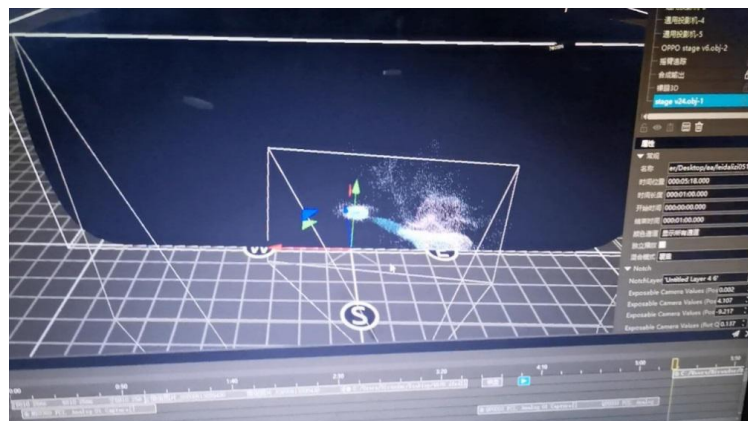


图 5-14 实时渲染引擎工作画面

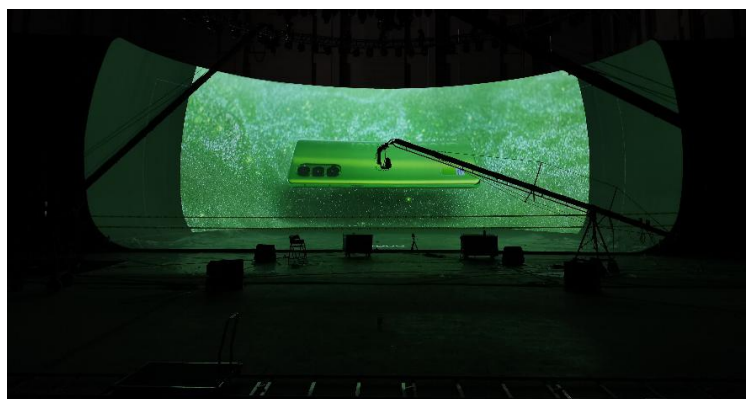


图 5-15 沉浸式碗幕的虚拟拍摄场景

5.6.4 应用效果

使用 XR 技术在虚拟节目制作中可营造出沉浸式空间，为拍摄带来便利，同时也为观众带来裸眼 3D 的效果，具有非常强的视觉冲击力及更好的沉浸式体验。



图 5-16 虚拟制作现场效果

6. 沉浸式视频应用的挑战与前景

沉浸式视频的发展是从巨幕、球幕、多屏、多声道等小众特殊行业逐渐衍生出来的，长期以来是一个封闭的、独立的行业，其技术标准、产品和内容也偏向于定制化。

随着新兴技术的广泛应用，沉浸式视频也逐渐进入大众传播领域。作为新兴的一种视音频业态，沉浸式视频还处于产业酝酿期，整个产业链存在不少问题，必须建立公共的技术标准、丰富的内容、完整的产业生态，其发展才能有序、有生命力。

6.1 存在的问题

6.1.1 内容制作域

目前，沉浸式视频的内容储备少，高质量的沉浸式视频更为欠缺。由于节目交换不多，缺乏相应的版权保护机制，造成了内容短缺，成本居高不下。在内容制作方面，沉浸式视频暂无统一标准。虽然已有不同场景下的应用探索，但受技术瓶颈制约，仍有部分环节的问题亟待解决。例如，在前期拍摄上，8K摄影机和镜头非常沉重，8K监看设备缺乏，因此在对焦、构图、拍摄速度上有一定程度的限制。在后期制作上，受到计算机运算处理能力和超高清海量素材量的制约，在存储速度和生成速度上还存在问题，制作效率也受到一定影响；另外，相关的周边配套软硬件还没有形成完备的体系，如多机位帧同步、导播切换系统、字幕特效系统等，无法直接使用现有的传统直播/转播设备。在三维声方面，国内内容制作产业支持较为滞后，节目源较少。

6.1.2 传输分发域

在传输方面，基于 5G 网络的沉浸式视频直播已开始探索，点播录播方式的传输链条还未建立，内容分发模式将成为影响市场的关键因素。沉浸式视频为达到较好的收看体验，需要超高清分辨率和百兆以上的码率，基于 5G 网络点对点的 4K 传输趋于成熟，但是 8K 传输仍然受制于编码效率和网络状况，8K 传输应用较少。另外，直播时延还处于秒级，由于全景画面拼接处理、格式映射转换、音视频编码、流式传输、播放服务器实时解码等造成的信号处理时间和网络缓存时间等原因，还不能达到真正意义的“直播”。

另外，节目的加密问题也亟待解决，完善的节目加密和授权方式是方便节目交流传播的前提，对于保证版权方的利益也至关重要。若结合压缩编码标准，进一步开发节目加密和授权方式，将对沉浸式视频的管理和推广起到积极的作用。

6.1.3 终端显示域

由于球幕、环幕等特效影院以往多以定制化节目或私有格式为主，很少进行内容交换，因此节目多以本地播放为主，终端媒体播放服务器还不支持流式信号内容的接收。球幕影院中投影融合及内弧形银幕会带来对比度不高的问题，HDR 很难实现；另外，LED 内弧屏带来的声聚焦问题和音箱透声问题均还没有彻底解决。三维声方面，球幕影院已开始使用多声道技术，未来需要统一制作技术和听音环境，研发传输编解码技术。家庭场景的解决方案还在探讨中，需要研发家庭显示终端、传输编解码、接口等。

另外，由于沉浸式视频行业以往不存在传输环节，目前沉浸式视频传输分发平台也未建立，还以本地服务器播放为主。对此，建议未来产业落地采用分步推进的方式，第一步先实现下载后本地播放，第二步实现边下载边播放（高时延），第三步实现实时播放（低时延）。

6.2 发展前景展望

5G 环境下，基于广播电视和网络视听发展的新形势和新需求，“科创”和“文创”相结合，将促进先进技术与应用场景的深度融合，打造文化消费主流，引领更多高新视频业态，为观众带来全新的视听体验，从而更好地满足新时代人民群众日益增长的精神文化新需求和新期待。

在沉浸式视频产业发展初期，可以通过广电网络手段，利用成熟的编解码标准，将各制作机构现有的内容推送给公益性展览展示场所、商业娱乐场所，帮助内容制作企业形成产品增值，丰富各场所的内容量，同时也为沉浸式视频走向大众视野奠定社会基础。

在沉浸式视频产业发展中期，利用 AVS 等具有民族知识产权的标准，建立内容拍摄、制作、传输、分发、呈现各个环节的标准技术体系，特别是拍摄、传输、播放、显示设备乃至处理芯片全产业链的完整生态链。



图 6-1 VisionStation 视效工作站和华为 VR 眼镜

在沉浸式视频产业发展的爆发期，是全面走向 C 端，进入家庭与柔性屏的结合，进入移动终端与手机和 VR 穿戴式设备结合，沉浸式视频将会成为视频观看体验的新模式，使消费者得到沉浸式的视觉和听觉体验。沉浸式视频的应用不仅仅是影视内容的观看消费，更可以将沉浸式视频引入到如点餐、购物、娱乐等日常生活消费中，将沉浸式视频的应用推广至文旅、文博、文娱、文创等新兴数字场景中。

参考文献

- [1] Henson, D.B. (1993). Visual Fields. Oxford: Oxford University Press.
- [2] Martin Howe (2004) A Proposal for Dome Standard Spatial Resolution, Luminance and System Contrast.
- [3] Ka Chun Yu, Filmmaking for Fulldome: Best Practices and Guidelines for Immersive Cinema (Part I), Planetarian (2016.12).
- [4] Ka Chun Yu, Filmmaking for Fulldome: Best Practices and Guidelines for Immersive Cinema (Part II), Planetarian (2017.3).
- [5] 张俊, 《大型数字球幕电影数字制作中的关键问题及进展》, 中国科学院计算技术研究网, 2012年.
- [6] 陈江, 《从人眼的特性分析和研究数字影院升级到4k分辨率图像放映系统的必要性》, 现代电影技术, 2013年7月.
- [7] DIGITAL IMMERSIVE GIANT SCREEN SPECIFICATIONS, (2018.01) 《DIGSS V 2.0 Image, Audio, and Theater Specifications》.
- [8] 4K超高清电视技术应用实施指南(2018版), 国家广播电视总局科技司.
- [9] GY/T 316-2018《用于节目制作的先进声音系统》.
- [10] GY/T《科技馆球幕影院功能配置与放映技术要求》(草案).
- [11] Report ITU-R BT.2420-0 (2018.04) 《Collection of usage scenarios and current statuses of advanced immersive audio-visual systems》.
- [12] Recommendation ITU-R BT.2123-0 (2019) 《Video parameter values for advanced immersive audio-visual systems for production and international programme exchange in broadcasting》.
- [13] DVB. (2016.11) 《Executive summary DVB study mission on virtual reality》.
- [14] Jaunt, Inc. (2017.01) 《The cinematic VR field guide - a guide to best practices for shooting 360°》.
- [15] 孙苏川、潘国林等, 《“观潮”-面向5G高新视频应用的思考》, 影视制作, 2019年第八期.

- [16] 上海交通大学媒体技术实验室, (2018.8), 《沉浸式全景视频技术指南》.
- [17] 汪芮, 覃毅力等, 《三维声 (3D Audio) 音频技术方案与标准测试》, 广播与电视技术, 2017 年第 12 期.
- [19] 王珏, 《VR (虚拟现实) 电影声音制作流程探析》, 现代电影技术, 2017 年 1 月.
- [20] 周耀平, 《沉浸式特效影院工程的关键工艺技术》, 现代电影技术, 2019 年第 10 期.
- [21] 周耀平, 《沉浸式特效影院工程的关键建筑技术》, 现代电影技术, 2020 年第 2 期.
- [22] Dolby. (2014. 10), 《Dolby Atmos Home Theater Installation Guidelines》.
- [23] Dolby. (2019.08), 《Dolby Atmos Home Entertainment Studio Certification Guide》.
- [24] <https://www.audioblog.iis.fraunhofer.com/cn>
- [25] <https://www.dolby.com/cn/zh/technologies/dolby-digital-plus.html#3>
- [26] <http://www.pci-china.com/content/7-solution-duoweidul>
- [27] <http://www.lightmagictech.com/case/read.aspx?cid=1&ID=222>
- [29] <http://www.fulldome.pro>
- [30] 《显示无处不在——视觉体验升级带动数据爆发增长》, 华为、京东方、信通院联合发布.
- [31] GB/T 33475.3-2018 《信息技术 高效多媒体编码 第 3 部分: 音频》