

附件 3

5G 高新视频—VR 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，越来越受到社会各界的广泛关注。在 5G、超高清、虚拟现实等新兴技术催生下，广电行业视听内容的生产和传播即将发生新变革。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”的概念。

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。

VR 视频作为高新视频业态的重要组成部分，是指全景视频，如水平 360° × 垂直 360° 全景视频、水平 180° × 垂直 180° 全景视频等，用户可借助 VR 眼镜等虚拟现实设备观看全景视频，并获得身临其境的视觉感受。

为更好地推进 VR 视频发展，指导相关行业与机构开展 VR 视频业务，提高 VR 视频质量，促进 VR 视频业态良性发展，国家广播电视总局科技司组织广播电视科学研究院及行业相关机构、企业等，开展了 VR 视频研究工作，编制了《5G 高新视频—VR 视频技术白皮书（2020 版）》（以下简称白皮书）。

本白皮书分析了 VR 视频的基本概念，深入探讨了 VR 视频关键技术，梳理了 VR 视频国内外技术标准现状，总结了 VR 视频在广播电视及新媒体领域中的应用案例，提出了 VR 视频端到端解决方案，最后分析了目前 VR 视频应用中遇到的问题和未来发展前景。希望本白皮书能为推动 VR 视频的应用提供参考，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

本白皮书主要起草单位：国家广播电视总局广播电视科学研究院、中央广播电视总台、中广电广播电影电视设计研究院、北京未来媒体科技股份有限公司、华为技术有限公司、深圳市天威视讯股份有限公司、大连天途有线电视网络股份有限公司、北京大学深圳研究生院、中兴通讯股份有限公司、四川传媒学院、杭州当虹科技股份有限公司、深圳市腾讯计算机系统有限公司、飞利浦（中国）投资有限公司、北京视博云科技有限公司、北京小鸟看看科技有限公司、深圳市圆周率软件科技有限责任公司、上海海思技术有限公司、深圳创维新世界科技有限公司。

本白皮书编写指导：孙苏川、关丽霞

本白皮书主要起草人：郭晓强、王强、孙可、魏娜、周芸、胡潇、谭阳、范晓轩、周屹、王倩、李大为、王琛、吕福明、王荣刚、罗浚兮、周乐、向东、冉峡、陈勇、许晓中、张峰昌、牛长峰、张瑞生、郭斐、宋海龙、张毅。

目 录

缩略语.....	3
1. 5G 高新视频—VR 视频概述.....	6
1.1 基本概念.....	6
1.2 3DoF VR 和 6DoF VR.....	6
1.3 观看 VR 视频的典型流程.....	7
2. VR 视频关键技术.....	9
2.1 VR 视频端到端系统框图.....	9
2.2 VR 视频节目制作与交换用视频参数.....	9
2.3 VR 视频拍摄制作与拼接.....	10
2.3.1 拍摄制作方法.....	10
2.3.2 视频拼接算法.....	11
2.4 VR 视频映射及压缩编码.....	12
2.4.1 映射.....	13
2.4.2 压缩编码.....	16
2.5 VR 视频传输.....	17
2.5.1 传输所需带宽.....	17
2.5.2 传输方案.....	17
2.6 VR 视频终端渲染显示.....	18
2.6.1 基本流程.....	18
2.6.2 渲染.....	19
2.6.3 显示.....	20
2.7 VR 视频感知交互.....	21
2.8 6DoF.....	21
2.8.1 拍摄制作方法.....	21
2.8.2 编解码.....	24
2.8.3 终端渲染显示.....	25
3. VR 视频国内外技术标准.....	27
3.1 VR 视频国外技术标准.....	27
3.1.1 ITU.....	27
3.1.2 MPEG.....	27
3.1.3 DVB.....	28
3.1.4 SMPTE.....	29
3.1.5 3GPP.....	29
3.1.6 JPEG.....	29
3.1.7 IEEE.....	30
3.2 VR 视频国内技术标准.....	30
4. VR 视频在广播电视及新媒体领域中的应用.....	33
4.1 国外 VR 应用试验.....	33
4.1.1 日本.....	33
4.1.2 韩国.....	34
4.1.3 美国.....	34

4.1.4 欧洲.....	35
4.1.5 国际互联网企业.....	35
4.2 国内 VR 应用试验.....	36
4.2.1 电视台.....	36
4.2.2 有线电视网络/IPTV.....	40
4.2.3 国内互联网企业.....	42
5. VR 视频端到端解决方案.....	44
5.1 VR 视频端到端解决方案系统架构.....	44
5.2 VR 视频拍摄制作域解决方案.....	46
5.2.1 VR 视频拍摄制作工艺流程.....	46
5.2.2 VR 视频拍摄方式.....	47
5.3 VR 视频业务平台方案.....	49
5.3.1 VR 视频直播/点播平台.....	49
5.3.2 VR 视频直播/点播平台技术指标.....	50
5.3.3 VR 视频直播/点播平台部署.....	51
5.4 VR 视频网络传输方案.....	51
5.4.1 基于有线电视网络/IPTV 的 VR 视频传输方案.....	52
5.4.2 基于 5G 网络的 VR 视频传输方案.....	52
5.5 VR 视频终端.....	53
6. VR 视频应用的挑战与前景.....	55
6.1 VR 视频节目制作域.....	55
6.1.1 拍摄指导.....	55
6.1.2 映射.....	55
6.1.3 制作流程和创作工具.....	55
6.2 VR 视频传输分发域.....	56
6.2.1 高传输码率.....	56
6.2.2 VR 分发格式.....	56
6.3 VR 视频终端消费域.....	56
6.4 发展前景展望.....	57
参考文献.....	58

缩略语

缩略语	英文全称	中文全称
3DoF	Three Degrees of Freedom	三自由度
3GPP	3rd Generation Partnership Project	第三代合作伙伴计划
5G	5th Generation Mobile Networks	第五代移动通信网络
6DoF	Six Degrees of Freedom	六自由度
AI	Artificial Intelligence	人工智能
APP	Application	应用程序
AR	Augmented Reality	增强现实
AVS	Audio Video Coding Standard	音视频编码标准
CDN	Content Delivery Network	内容分发网络
CG	Computer Graphics	计算机图形学
CMP	Cube Map Projection	正六面体映射
CPU	Central Processing Unit	中央处理器
DASH	Dynamic Adaptive Streaming over HTTP	基于 HTTP 的动态自适应流
DIBR	Depth Image Based Rendering	基于深度的图像渲染
DNS	Domain Name System	域名系统
DVB	Digital Video Broadcasting	数字视频广播
EAC	Equi-Angular Cubemaps	等角方块
ERP	Equi-Rectangular Projection	等距圆柱映射
FOV	Field of View	视角
GPU	Graphics Processing Unit	图形处理器
HTML	Hyper Text Markup Language	超文本标记语言
HDR	High Dynamic Range	高动态范围
HFR	High Frame Rate	高帧率
HLS	HTTP Live Streaming	基于 HTTP 的流媒体网络传输协议
HMD	Head Mounted Display	头戴式显示器
HTTP	Hyper Text Transfer Protocol	超文本传输协议

缩略语	英文全称	中文全称
HTTPS	Hyper Text Transfer Protocol over Secure Socket Layer	安全套接层超文本传输协议
IEEE	Institute of Electrical and Electronics Engineers	电气和电子工程师协会
ILE	Immersive Live Experience	沉浸式现场体验
IP	Internet Protocol	网际互连协议
ISP	Icosahedron Projection	正二十面体映射
ITU	International Telecommunication Union	国际电信联盟
JPEG	Joint Photographic Experts Group	联合图像专家组
MMT	Multi-Media Terminal	多媒体终端
MPEG	Moving Picture Experts Group	运动图像专家组
MR	Mixed Reality	混合现实
NDI	Network Device Interface	网络设备接口
NPU	Neural Network Processing Unit	嵌入式神经网络处理器
OHP	Octahedron Projection	正八面体映射
OLED	Organic Light Emitting Diode	有机发光二极管
OMAF	Omnidirectional Media Application Format	全景媒体应用格式
PGC	Professional Generated Content	专业生产内容
PPD	Pixel Per Degree	像素/度
QoE	Quality of Experience	体验质量
QoS	Quality of Service	服务质量
RAN	Radio Access Network	无线接入网
ROI	Region of Interest	感兴趣区域
RTMP	Real Time Messaging Protocol	实时消息传输协议
SDK	Software Development Kit	软件开发工具包
SDR	Standard Dynamic Range	标准动态范围
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
SSP	Segmented Sphere Projection	分区域球体映射

缩略语	英文全称	中文全称
TSP	Truncated Square Pyramid	截体金字塔
UDP	User Datagram Protocol	用户数据报协议
UGC	User Generated Content	用户生产内容
VBR	Variable Bit Rate	可变比特率
VR	Virtual Reality	虚拟现实
WCG	Wide Color Gamut	广色域

1. 5G 高新视频—VR 视频概述

1.1 基本概念

VR 通过动态环境建模、实时三维图形生成、立体显示观看、实时交互等技术，生成仿真现实的三维模拟环境，能够构造视觉、听觉等方面高度主观真实的人体感官感受。用户借助交互设备与虚拟场景中的对象相互作用、相互影响、沉浸其中，获得等同亲临客观真实环境的感受和体验。

AR 是在 VR 的基础上，运用多媒体、三维建模、实时跟踪、智能交互、传感等多种技术手段，将计算机生成的文字、图像、三维模型、音乐、视频等虚拟信息模拟仿真后，应用到真实世界中，两种信息互为补充，从而实现了对真实世界的“增强”。

MR 是 VR 技术的进一步发展，该技术通过在现实场景呈现虚拟场景信息，在现实世界、虚拟世界和用户之间搭起一个交互反馈的信息回路，以增强用户体验的真实感。

VR 视频是指全景视频，如水平 $360^\circ \times$ 垂直 360° 全景视频、水平 $180^\circ \times$ 垂直 180° 全景视频等，用户可借助 VR 眼镜等虚拟现实设备观看全景视频，并获得身临其境的视觉感受。考虑到技术成熟度，本白皮书涉及的内容主要讨论 3DoF VR 视频，6DoF VR 视频是未来发展方向，本白皮书在 2.8 节给出一定的技术说明，3DoF VR 和 6DoF VR 的概念见 1.2 节。此外，VR 视频系统还包含配套的音频内容，本白皮书暂不涉及。

1.2 3DoF VR 和 6DoF VR

(1) 3DoF

3DoF 是指用户可以在任何方向（偏转、俯仰和滚动）自由地观看节目素材。典型的应用场景是坐在椅子上的用户通过 HMD 观看 VR 视频内容，如图 1 所示。

(2) 3DoF+

3DoF+是指用户可以在任何方向（偏转、俯仰和滚动）自由地观看节目素材，同时用户头部可以进行一定的平移运动。典型的应用场景是坐在椅子上的用户通过 HMD 观看 VR 视频内容，并且用户头部可以进行小幅度的上下、左右、前后移动，如图 1 所示。

(3) 6DoF

6DoF 是指用户可以在物理空间内任何位置、任何方向自由地观看节目素材。用户移动可以被传感器或输入控制器捕获到，同时支持用户空间位移和头部姿态变化。典型的应用场景是用户自由走动并同时通过 HMD 观看 VR 视频内容，如图 1 所示。

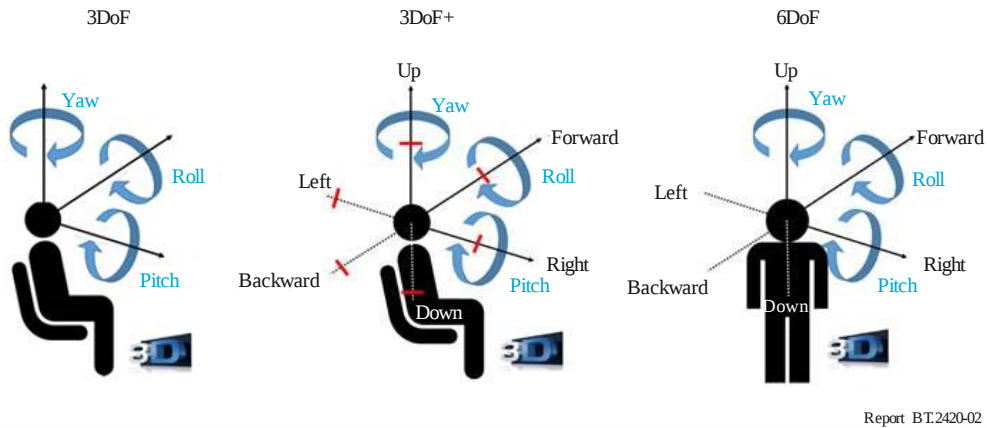


图 1 各类 DoF 示意图

1.3 观看 VR 视频的典型流程

观看 VR 视频时，用户需佩戴 VR 眼镜（配合手机使用）或 HMD，运行在用户手机或机顶盒等设备上的 VR 视频应用软件启动并初始化 VR 眼镜或 HMD 中的陀螺仪等传感器装置，用户通过遥控器、手柄等输入设备与 VR 视频应用软件进行实时交互。当用户头部发生运动时，VR 眼镜或 HMD 中的陀螺仪等传感器装置实时检测头部的运动状态，并将其参数传递给 VR 视频应用软件的处理模块。处理模块获取到用户头部的运动状态参数后，VR 视频应用软件立即计算用户的观看视角并重新渲染新的一帧图像传递至 VR 眼镜或 HMD 上，用户可以进行实时观看。观看 VR 视频的典型系统流程如图 2 所示。

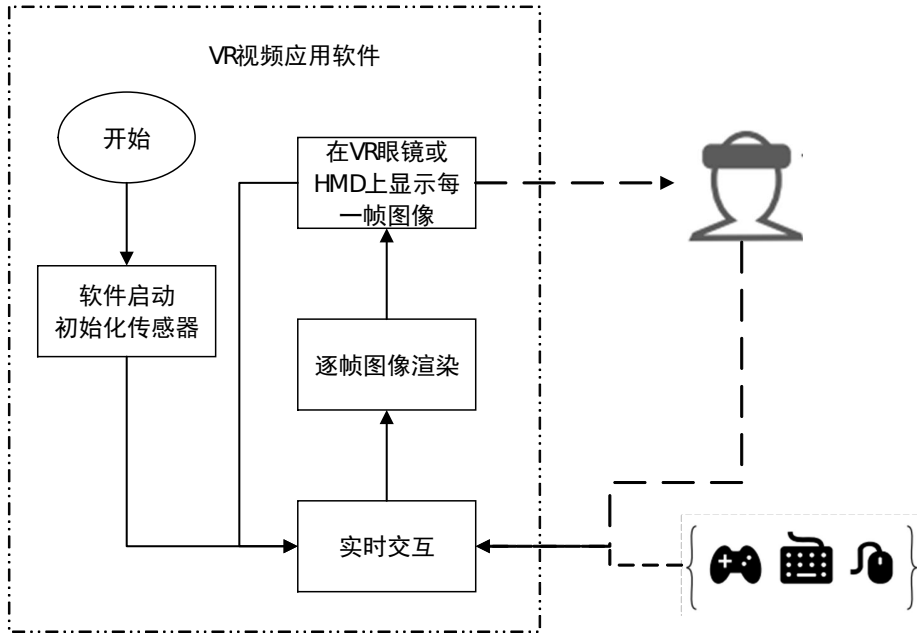


图 2 观看 VR 视频的典型系统流程

2. VR 视频关键技术

2.1 VR 视频端到端系统框图

典型的 VR 视频端到端系统框图如图 3 所示，包括 VR 视频源、VR 视频编码、VR 视频传输和 VR 视频终端接收和显示 4 个部分。图中通路①是实时直播应用场景，通路②为非实时点播应用场景。实时直播应用场景与非实时点播应用场景的区别在于：实时直播采用实时拍摄、拼接的 VR 视频源信号，而非实时点播则采用已经制作完成的 VR 视频源文件。两种信号进入 VR 视频编码环节及其之后环节的技术流程一致。

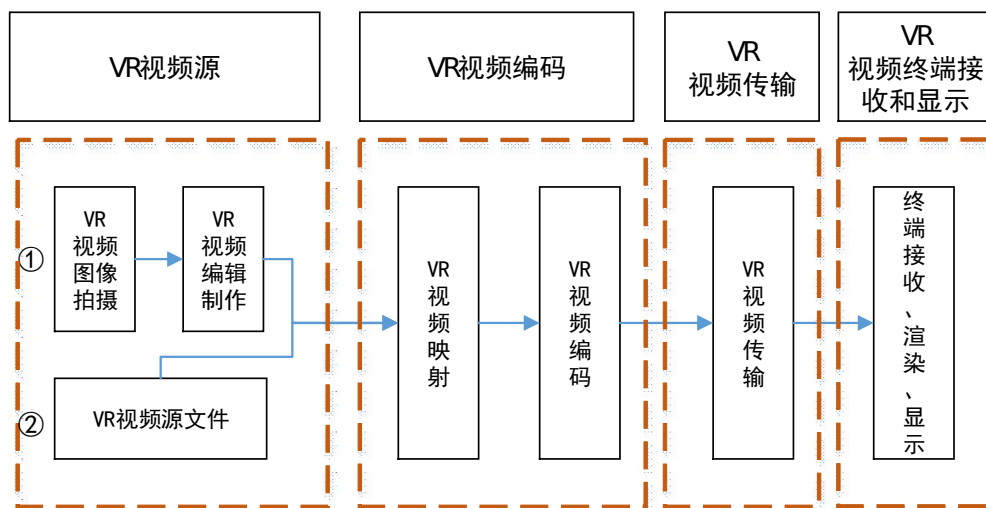


图 3 VR 视频端到端系统框图

VR 视频端到端系统流程首先是使用全景摄像机拍摄，随后将多路视频拼接融合为完整的全景视频进行后期编辑制作，经过高效的视频压缩并传输，最后分发到用户端。用户端通过 HMD 或 VR 眼镜观看 VR 视频内容，通过头部姿态调整，观看不同视角的 VR 视频内容。

VR 视频采集拍摄、拼接编辑制作、编码传输、终端渲染显示等各个环节都会影响用户的观看体验，包括视频质量、流畅性、沉浸感等方面。

2.2 VR 视频节目制作与交换用视频参数

用于 VR 视频节目制作与交换中的视频参数值见表 1。表 1 主要参考了 ITU-R

BT. 2123 标准，映射后图像的有效像素数为 30K×15K，该值基于人眼的空间视觉特性，使得观看者在观看 VR 视频的一部分时感知不到显示屏幕的像素结构。VR 视频要求 30K×15K 的像素数，其他像素数可在实际系统设计时使用，建议目前最低使用 7680×3840 像素数。

表 1 VR 视频节目制作与交换用视频参数值

序号	参数		数值		
1	球面360° 图像到矩形二维图像的映射方法		默认ERP模型，其他模式应该在系统中标识		
2	映射后图像的有效像素数 (水平×垂直)		30720 ×15360 (30K × 15K)、7680 × 3840 (8K × 4K)		
3	像素宽高比		1:1 (方形)		
4	帧率 (Hz)		120, 100, 50		
5	扫描模式		逐行		
			光谱	色坐标 (CIE, 1931)	
				x	y
6	基色	基色红 (R)	单色630 nm	0.708	0.292
7		基色绿 (G)	单色532 nm	0.170	0.797
8		基色蓝 (B)	单色467 nm	0.131	0.046
9	基准白		D65, 符合ISO 11664-2:2007	0.3127	0.3290
10	颜色匹配函数		CIE 1931		
11	信号格式		R' G' B'、Y' C' B' C' R (非恒定亮度)、IC _T C _P		
12	R' G' B'、Y' C' B' C' R 和 IC _T C _P (仅限HDR) 的导出		SDR: 见GY/T 307 HDR: 见GY/T 315		
13	量化电平 (bit)		10、12		
14	色度亚取样		见GY/T 315		
15	数字10比特和12比特整数表示		见 GY/T 315 (SDR: 窄范围, HDR: 窄范围或全范围)		

2.3 VR 视频拍摄制作与拼接

2.3.1 拍摄制作方法

(1) VR 视频拍摄

VR 视频拍摄以采集点为观察点拍摄包含场景所有方向的视觉画面，如 360° ×

360° VR 视频、180° × 180° VR 视频。拍摄 VR 视频的采集装置通常由多个摄像机呈环形或球面排列组成，VR 视频由每个摄像机采集的视频拼接而成。VR 视频提供第一人称视角的视觉体验，用户固定观看位置，支持头部转动，以头部为中心转动选择视角方向，观看相应方向的画面。

目前由多摄像机多镜头组成的 VR 视频采集装置虽没有达到广播级拍摄能力，但已经可以输出高质量的全景视频。配合现有高端手机具备的 8K 解码能力，以及 HMD 和手机具备的灵活交互能力，VR 视频已较好地应用于直播等场景。

(2) 计算机图形学制作

CG 制作是通过计算机实时计算、渲染出虚拟的场景和实体。CG 制作计算量大，对 CPU 和 GPU 的性能要求较高，目前主要用于 VR 动画等。

(3) 实拍抠像结合 CG 制作

实拍抠像是在蓝、绿色抠像影棚拍摄主体影像，包括人物、道具等，后期加入 CG 制作的背景和环境，或者 CG 制作的其它元素，与实拍影像进行高度融合，达到真实的效果。

2.3.2 视频拼接算法

常用的视频拼接算法可分为基于变换的图像拼接算法和基于拼接线的图像拼接算法。

基于变换的图像拼接算法的核心思想是通过对单应性矩阵进行调整，通过网格化的扭曲使重合区域拼接的缝隙尽可能减小，该算法适合小视角变换的情况。代表性算法有 AutoStitch、APAP、ANAP、GSP 等。

基于拼接线的图像拼接算法核心思想是通过对图像拼接线部分的重新调整，确保拼接的自然性，该算法适合大视角变换的情况。代表性算法有：Seam-Driven、Parallax-Tolerance 等。

(1) AutoStitch

AutoStitch(Automatic Panoramic Image Stitching)算法是 2007 年 M. Brown 和 D. Lowe 提出的，该算法主要是通过单应性变换重建视场，进而完成图像拼接。

AutoStitch 算法的特点是：使用概率模型进行特征点筛选、使用光束平差进行位置优化、使用多波段方法进行图像融合。

(2) APAP

APAP (As-Projective-As-Possible Image Stitching with Moving DLT) 算法是 2013 年 Zaragoza. J., Chin. T. J. 和 Brown. M. S. 提出的, 是一种基于变换的图像拼接算法。APAP 算法改进了 AutoStitch 算法中默认光心不动的缺陷, 从而优化了拼接后的重影/鬼影问题。

APAP 算法的特点是: 引入了网格化, 针对图像的不同区域进行加权计算, 优化了重合处的效果。

(3) ANAP

ANAP (Adaptive As-Natural-As-Possible Image Stitching) 算法是 2015 年 Lin. C. C., Pankanti. S. U. 和 Ramamurthy. K. N. 提出的, 该算法主要是通过对单应性矩阵进行线性过渡, 考虑旋转角度, 尽可能地恢复图像的自然性。

ANAP 算法的特点是: 对重叠区域和非重叠区域进行不同处理、采用全局相似变换提高图像自然性、对变换矩阵采用线性过渡。

(4) GSP

GSP (Natural Image Stitching with the Global Similarity Prior) 算法是 2016 年 Chen. Y. S. 和 Chuang. Y. Y. 提出的, 该算法主要是提高全景图自然性的同时, 提高拼接的准确性。

GSP 算法的特点是: 使用 APAP 增加优化的匹配点数, 使用不同优化项进行优化, 使用直线检测, 设置旋转阈值, 进行二维和三维优化。

(5) Seam-Driven

Seam-Driven (Seam-Driven Image Stitching) 算法是 2013 年 Gao. J., Li. Y. 和 Chin. T. J. 提出的, 该算法主要是找到两幅图片的拼接线, 通过拼接线对图像进行融合。

Seam-Driven 算法的特点是: 使用拼接线对两幅图片进行拼接, 求解多个单应性模型并进行评估, 选择最优结果。

(6) Parallax-Tolerance

Parallax-Tolerance (Parallax-Tolerant Image Stitching) 算法是 2014 年 Zhang. F. 和 Liu. F. 提出的, 该算法主要是提高大视差场景下的效果。

Parallax-Tolerance 算法的特点是使用了视频去抖动方法的优化项。

2.4 VR 视频映射及压缩编码

VR 视频映射是 VR 视频编码的预处理环节，目前专门针对 VR 视频的编码标准仍在研究中，通常将 360° 的二维球面视频图像映射成为二维平面矩形视频图像，再送入编码器进行编码传输。VR 视频有多种映射表示方式，从压缩编码的角度看，不同的映射方式其压缩效率也不同。

2.4.1 映射

VR 视频映射是指将球面全景视频表示为适于压缩编码的平面视频，即将球面全景图像转化映射成为二维平面图像。

球面全景图像映射成为二维平面图像的映射模型有多种，为了进一步提高编码效率以及达到尽可能好的映射效果，映射模型还在不断地研究和丰富中。MPEG 等标准中提出了多种映射模型，最常用的是等距圆柱映射和正六面体映射模型。典型映射模型介绍如下：

(1) 等距圆柱映射 (ERP)

ERP 映射模型是在 VR 视频中使用最广泛的映射模型，如图 4 所示即是 ERP 映射模式的二维图像平面。经纬图模型只有一个投影面。YouTube、Samsung Gear、优酷、爱奇艺等均采用此种投影格式生产 VR 媒体文件。

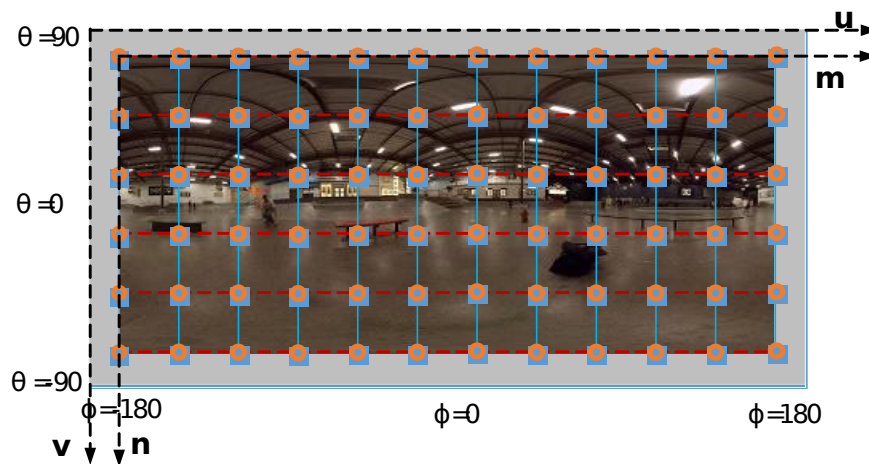


图 4 ERP 映射模式

(2) 正六面体映射 (CMP)

CMP 映射模型有 6 个映射面，分别将 360° 球体视频图像映射到 6 个面上，如图 5 所示。

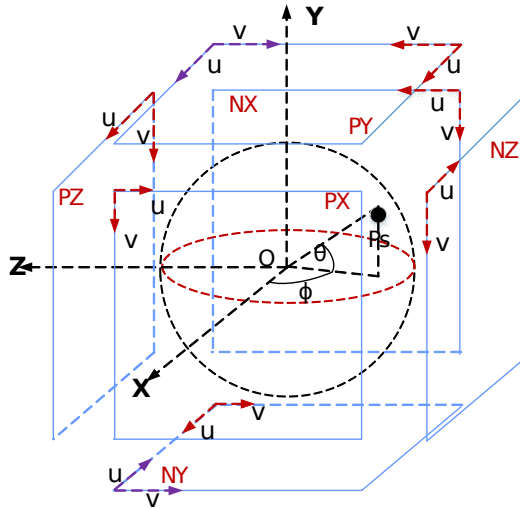


图 5 CMP 映射模式

(3) 等角方块映射 (EAC)

EAC 等角方块映射投影法的目标是不论取样样点在映射面的什么位置，都能保证观看者在各处获得统一的像素观看密度感。EAC 映射模型在 6 个映射面上，各角度内的像素密度是保持恒定不变的。EAC 与 CMP 映射模型的对比示意图如图 6 所示。

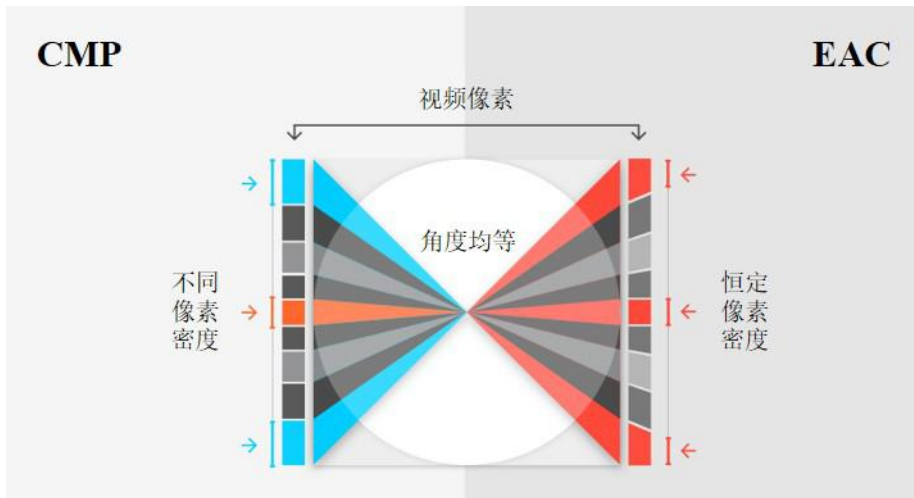


图 6 EAC 与 CMP 映射模型对比示意图

(4) 正八面体映射 (OHP)

OHP 映射模型有 8 个三角形的映射面和 6 个顶点，OHP 模型不同于 CMP 模型的特殊之处在于它包含三角形的映射面。三角形映射面需要仔细处理以保证面与面交界处的视觉连续性以及编码效率提升，如图 7 所示。

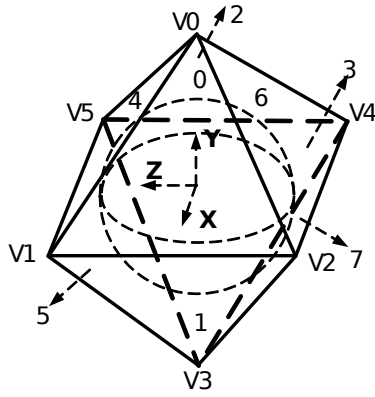


图 7 OHP 映射模式

(5) 正二十面体映射 (ISP)

ISP 模型有 20 个三角形的映射面和 12 个顶点。非紧凑型 and 紧凑型 ISP 的示意图如图 8 所示。

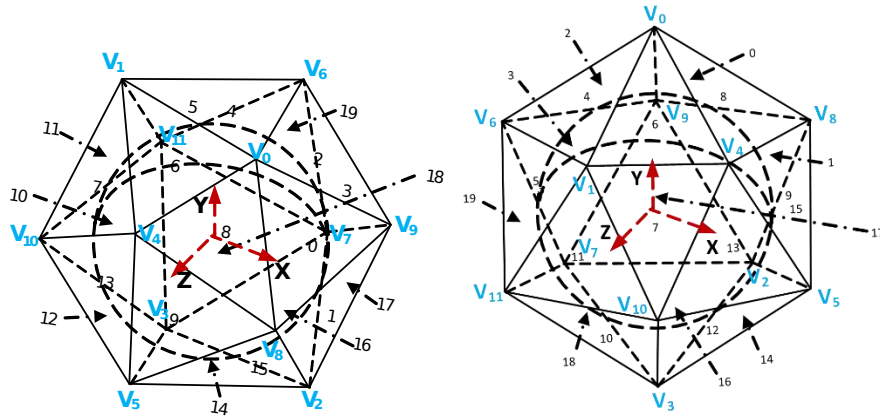


图 8 非紧凑型 (左) 和紧凑型 (右) ISP 映射模式

(6) 分区域球体映射 (SSP)

SSP 映射法将球体分为北极区圆、南极区圆和中间赤道部分等三个区域进行映射，如图 9 所示。

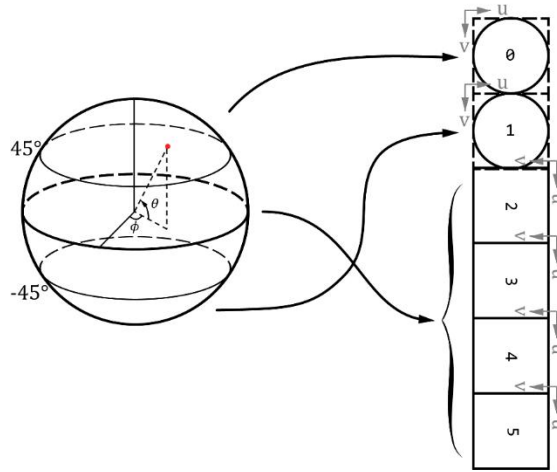


图 9 SSP 映射模式

(7) 截体金字塔 (TSP)

TSP 映射模型使用顶端截断的金字塔形的立方体几何模型，并将该立方体的六个面整合形成一个紧凑的矩形帧。观看正面是一个整形的正方形，观看侧面和背面则逐渐缩小，如图 10 所示。

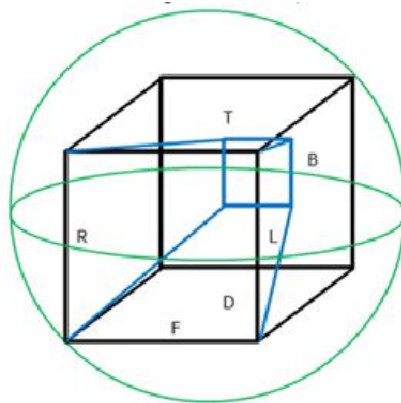


图 10 TSP 映射模式

(8) 非均匀映射方法

除上述多种映射模型外，还可以采用非均匀映射的方法。在主视点区域使用较高的采样密度进行采样，在非主视点区域使用逐渐降低的采样密度进行采样。使用非均匀映射方法的好处是可以大幅降低服务器端的接收码率和解码复杂度，可用于流切换等场景。常用的非均匀映射方式包括非均匀经纬图、非均匀球体和非均匀正六面体等。

2.4.2 压缩编码

映射后的 VR 视频可采用普通视频的编码技术进行压缩。目前应用较多的视频

编码技术是 H. 264、H. 265、AVS2 等，在保证同等画质的前提下，H. 265 和 AVS2 的压缩效率大约比 H. 264 提升 50%左右。下一代编码技术 H. 266 与 AVS3 的目标压缩效率比 H. 265 与 AVS2 提升一倍。

码率是影响 VR 视频在终端接收观看清晰度的重要指标。8K/50P 的视频，采用 H. 265 或 AVS2 编码，码率需 80~100Mbps。

2.5 VR 视频传输

2.5.1 传输码率

VR 视频业务的发展是画质、交互感不断提升，沉浸感不断增强的过程。传输码率很大程度上决定了 VR 视频业务的质量。现阶段 8K VR 视频需要 80~100Mbps 码率，强交互模式下运动感知时延应小于 10ms，弱交互模式下运动感知时延应小于 20ms；未来 30K VR 视频预计需要 800~1000Mbps 码率，强交互模式下运动感知时延应小于 5ms，弱交互模式下运动感知时延应小于 10ms。本白皮书 2.7 节详细描述了 VR 视频的感知交互模式，VR 视频对传输码率的具体需求见表 2。

表 2 VR 视频对传输码率的需求

参数	8K	30K
VR 视频分辨率	7680 × 3840	30720 × 15360
典型单眼分辨率	1920 × 1920	7680 × 7680
量化电平 (bit)	10	12
编码标准	H. 265、AVS2 等	H. 266、AVS3
帧率 (Hz)	50	100
典型码率 (Mbps)	80~100	800~1000

2.5.2 传输方案

VR 视频传输主要有两种技术路线：全视角传输方案和基于 FOV 的传输方案。

(1) 全视角传输方案

全视角传输方案就是将 360° 环绕画面都传输给终端，当用户头部转动需要切换画面时，所有的处理都在终端本地完成。VR 视频在相同单眼可视分辨率情况下，由于帧率、量化电平、360° 环绕等原因，码率要比普通平面视频大很多，前者一般是后者的 5~10 倍，这对于传输来说是个极大的挑战，成本也大大增加。

(2) FOV 传输方案

虽然整个 VR 视频是 360° 的，但是观看者在观看时，实际只能看到当前视野部分，看到的内容只是占了部分带宽，采用全视角传输方案对带宽资源造成了比较大的浪费。针对这种情况，业界提出了基于 FOV 的传输方案。

FOV 传输方案主要传输当前视角中的可见画面。一般都是将 360° 全景视野划分为若干个视角，每个视角生成一个视频文件，只包含视角内高分辨率和周围部分低分辨率视觉信息，终端根据用户当前视角姿态位置，向服务器请求对应的视角文件。当头部转动视角发生变化时，终端向服务器请求新视角对应的视角文件。

以基于金字塔投影的 FOV 传输方案为例，如图 11 所示，将用户在虚拟环境中的视觉信息对应的全部球面数据放入金字塔投影。用户视点正前方的平面为 FOV 平面，使用高分辨率编码，其余四个平面为非 FOV 平面，分辨率从与 FOV 平面相交的边到视角反方向的顶点逐渐降低。传输网络根据终端返回的用户视角信息，向终端提供 FOV 范围内的高质量画面和非 FOV 范围内的低质量画面。

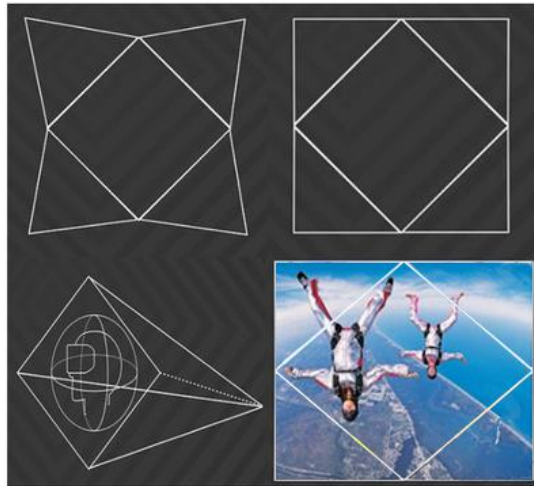


图 11 FOV 视角传输示意图

2.6 VR 视频终端渲染显示

2.6.1 基本流程

VR 视频以流媒体形式经网络传输到达用户终端侧，或者 VR 视频以文件形式存储在终端侧，终端对 VR 视频进行解码、渲染和显示。目前，终端有两种形态，一种是 HMD 形式，HMD 主要是显示设备，解码渲染等处理能力在 PC、机顶盒等终端上，HMD 要与终端连接；另一种是 VR 眼镜，可以是具备处理能力的一体式 VR 眼镜，也

可以是连接手机使用的 VR 眼镜。

运行在终端上的 VR 视频应用软件，实时追踪用户头部的 3DoF 姿态，即头部的偏转、俯仰和滚动参数，并根据这些参数实时渲染出用户当前姿态下应看到的每一帧 VR 视频画面，追踪、渲染、显示实时完成。

2.6.2 渲染

在 VR 视频应用中，渲染主要是指根据用户头部姿态从全景视频中截取出用户当前应看到的画面，并根据 HMD 和 VR 眼镜的透镜特性，对画面进行的变形处理。

HMD 和 VR 眼镜的显示屏幕配合距离人眼只有 4~5cm 距离的凸透镜，达到增大视角范围的效果。VR 视频应用软件预先对图像进行“桶形失真”变形处理，图像经透镜折射后，“桶形失真”与透镜折射引入的“枕形畸变”相互抵消，最终用户在 HMD 和 VR 眼镜里可看到正常的画面。“桶形失真”和“枕形畸变”示意图分别如图 12、图 13 所示。

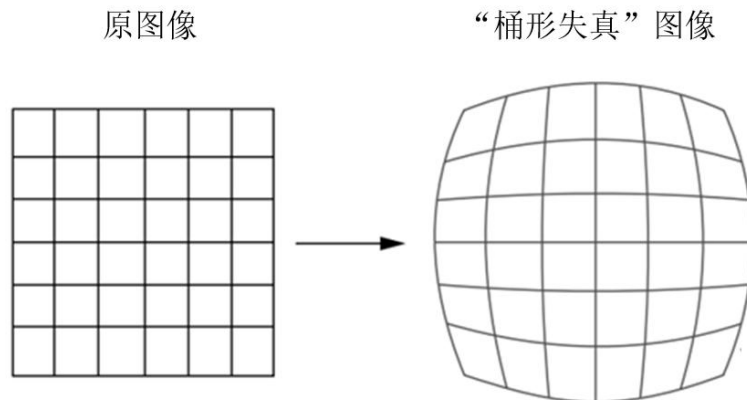


图 12 VR 视频图像“桶形失真”示意图

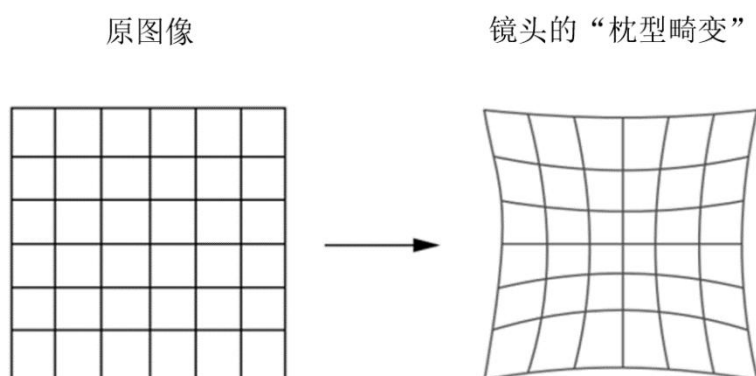


图 13 VR 镜头“枕形畸变”示意图

2.6.3 显示

(1) VR 视频像素密度

人肉眼视网膜中心凹（视网膜中心凹是视觉敏锐度最高的区域）的分辨率极限是 60 PPD（因人而异，平均为 60 PPD），即如果一幅 60×60 像素的图像落入视网膜中心凹 $1^\circ \times 1^\circ$ 的区域，则人眼无法将其与一幅 80×80 像素且落入 $1^\circ \times 1^\circ$ 视网膜中心凹区域的图像区分开。根据人眼 60PPD 的分辨能力，可以计算得到，为达到良好视觉效果，VR 视频水平方向 360° 应具有至少 21600 像素。

目前，VR 视频分辨率通常是 4K，与理想分辨率还有较大差距。未来拍摄 $30K \times 15K$ 的超高分辨率 VR 视频，达到人眼观看时 60PPD 的视网膜感受，才能带来画面质量质的飞跃。

(2) 显示屏分辨率

HMD 和 VR 眼镜的显示屏在显示 VR 视频时会分为左右相等的两个显示区域，分别显示左右眼观看的画面。目前，终端显示屏的最高分辨率可达到 4K，按照人眼视域 124° 和 60PPD 计算，终端显示屏要达到 16K 才能实现理想的显示效果，因此终端屏幕分辨率还需要进一步提升。

HMD 和 VR 眼镜内置了凸透镜，用于放大显示屏幕的图像，放大倍数一般为 5~7 倍。因此，用户通过 HMD 和 VR 眼镜观看 VR 视频时，相当于近距离用一个放大 5~7 倍的放大镜观看显示屏，显示屏的物理结构被放大显现出来，这就是业内常说的纱窗效应，类似于隔着纱窗看窗外的世界，如图 14 所示。因此，HMD 和 VR 眼镜的显示屏分辨率需要进一步提升，降低纱窗效应。

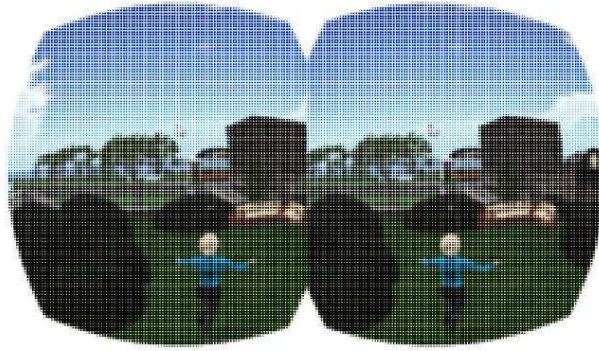


图 14 显示屏“纱窗效应”示意图

2.7 VR 视频感知交互

用户佩戴 HDM 或 VR 眼镜观看 VR 视频时，用户头部姿态的变化，如偏转、俯仰和滚动，可以被 HDM 和 VR 眼镜中的陀螺仪等装置捕获到，VR 视频系统根据这些捕获的姿态参数，渲染显示用户当前应看到的画面。

在 VR 视频应用中，用户与 VR 视频系统的交互属于弱交互。弱交互是指用户与虚拟环境中的实体不发生实际的互动，用户可以在一定程度上选择视点和位置。在弱交互环境中，用户体验是相对被动的，体验内容也是预先规划好的。VR 视频、VR 旅游等业务属于典型的弱交互范畴。

相对于弱交互，强交互是指用户可通过交互设备与虚拟环境进行互动，通过虚拟环境中的物体对交互行为做出实时响应，使用户能够感受到虚拟环境的变化。在强交互中，虚拟环境中实体的变化与用户输入有关，这与 VR 视频预先规划的内容体验是不一样的。VR 游戏等业务属于强交互范畴。

2.8 6DoF

2.8.1 拍摄制作方法

(1) 摄像机拍摄

6DoF 中，用户可以在场景中移动。在拍摄 6DoF 的 VR 视频时，必须通过“真实”摄像机在整个视区中记录足够的视图，以允许最终在用户的渲染设备中进行高质量的视图合成。

在实际拍摄中，相机的数量、位置、角度等取决于所需内容的质量，而质量又取决于许多因素，例如：视区的大小、与相关物体的距离、物体的类型、用户的预期运动等。如果有一些物体离用户很近，则相对较小的用户动作将显著改变物体的

视线（即大视差），并且会迅速遮挡物体的某些部分，相反，如果物体离得较远则看不见。为了捕获此信息，与拍摄远处的物体相比，将需要更多的摄像机。因此，6DoF 摄像机有多种设置，如图 15、图 16 所示。

相机往往基于固定的角度间隔摆放，角度间隔越小，视角切换的平滑度会越高，但同时相机的数量和系统成本也会随之增加，所以如何利用尽可能少的相机拍摄出平滑度高的 6DoF VR 视频是前端采集的关键任务。除此之外，采集系统的同步性和标定精度也是影响拍摄质量的两个主要指标，为了拍摄出接近静止的多角度精彩瞬间，要求相机支持毫秒级别的同步触发拍摄。直播图像要围绕一个焦点旋转，支持焦点的数量和可选范围的大小也是衡量一个拍摄系统的关键所在。

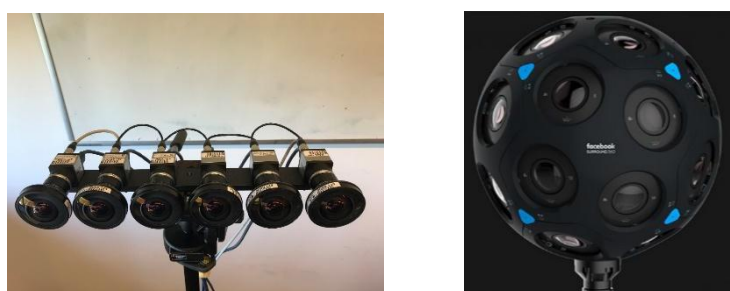


图 15 较小的摄影机装备，可在较小的观看区域内捕捉 6DoF VR

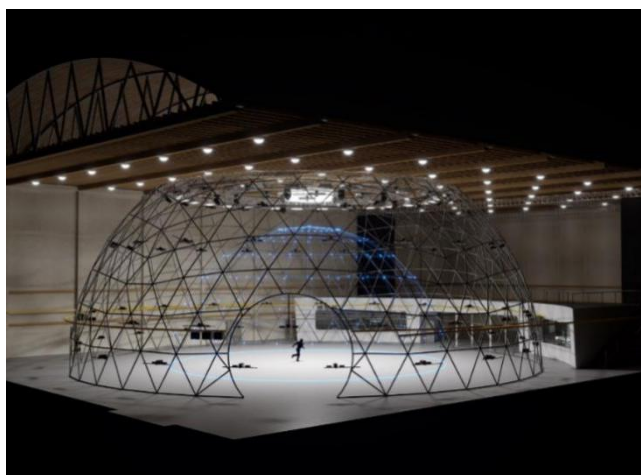


图 16 具有许多摄像头的 6DoF VR 设置的 Intel Studio

(2) CG 仿真制作

6DoF 内容也可以由 CG 仿真制作，可以从 CG 模型中渲染出所需的“真实”视图，而无需物理摄像机。例如，对于图 17 中的场景，图 18 显示了 CG 所生成的虚拟摄像机视图集。从这 15 个视图集合中，实际上可以在视图合成中生成任何其他视图。



图 17 VR 视频全景视图和相应的深度图

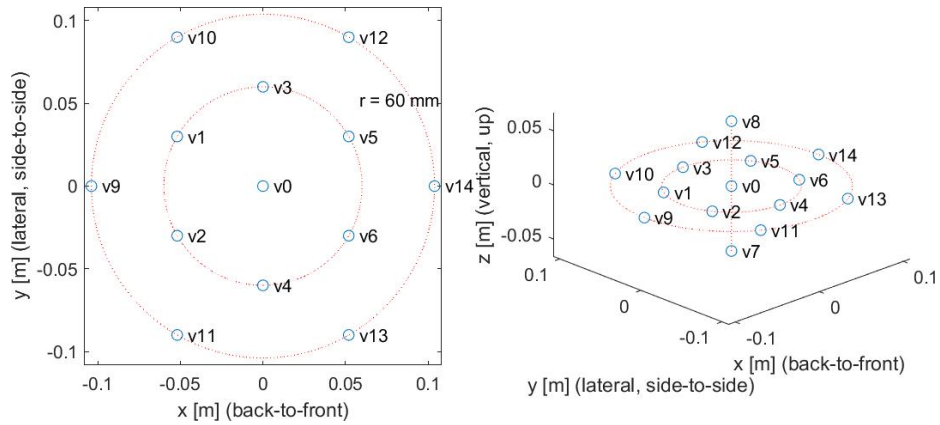


图 18 15 个虚拟 VR 全景摄像机的位置

通常，在拍摄 6DoF VR 场景时，有两种方法：由外而内和由内而外。图 19 展示了由外而内拍摄时的摄像机布置，可捕捉球场上的动作。这种方式也适用于用户与正在进行的事件相对较远的其他场景。

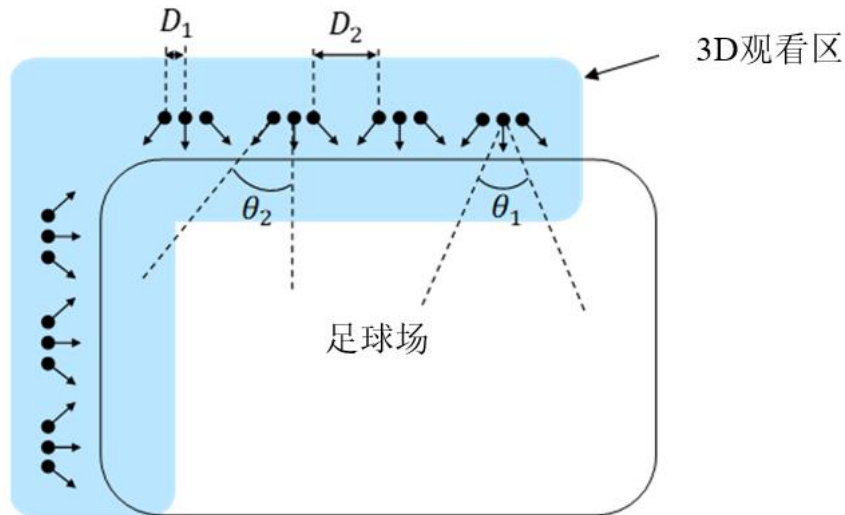


图 19 由外而内 VR 拍摄体育比赛(点代表摄像机，箭头代表镜头对称轴的方向)

对于用户想沉浸在动作中间的用例，例如，身临其境的旅行或一级方程式赛车的驾驶舱，由内而外的拍摄更为合适。图 20 所示为这种情况下典型的摄像机布置。

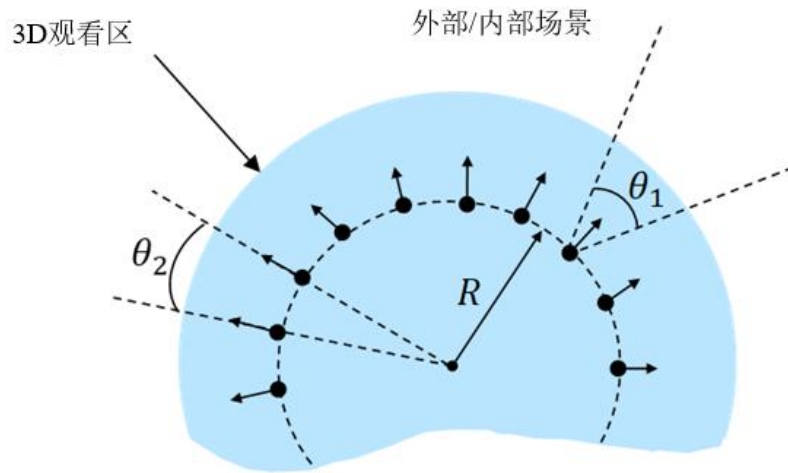


图 20 由内而外的 VR 拍摄的典型摄像机布置

2.8.2 编解码

典型的编码问题是发送方采用预定义类型的输入，将其编码为码流并发送。接收器了解编码方案，可使用该方案对接收到的码流进行解码，从而重建输入视频。广播应用的视频压缩是有损编码，这意味着无法完全重建输入视频。在某种程度上，6DoF VR 视频也会发生同样的情况，但情况要复杂一些。6DoF 编解码器的系统框图如图 21 所示。

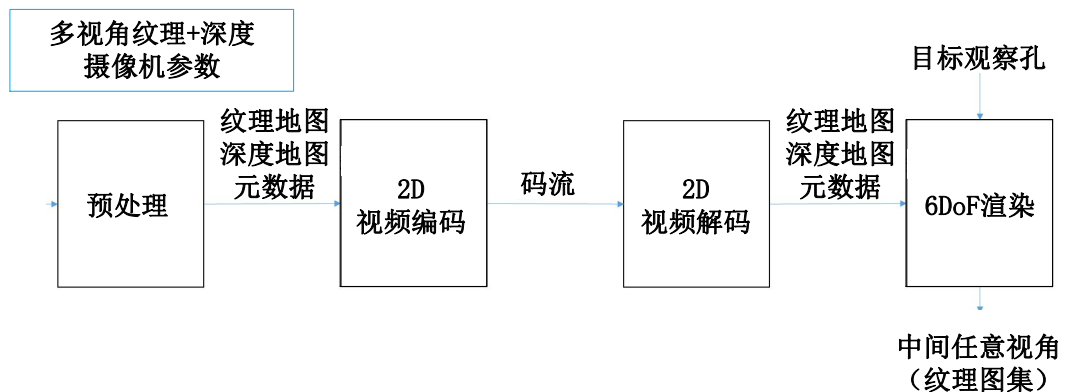


图 21 6DoF 编解码器的系统框图

(1) 6DoF 编码

6DoF 编码器由可转换输入数据的预处理器组成。预处理器的第一步是搜索冗余像素并裁剪它们。例如，平坦的白墙的外观在很大程度上不依赖于视点，因此，如果从一个视图中看到它可能就足够了。对于树木或镜面物体等复杂对象，应保留多个视点。第二步是将保留的区域打包到一帧图像中。将图像小块打包成帧图像在计算机图形学中称为纹理图集。此外，对于 6DoF 还存在相应的深度图集。图集元

数据对图集中的图像小块与裁剪后的源视图之间的关系进行编码，该元数据对于重建至关重要。预处理器流程图如图 22 所示。

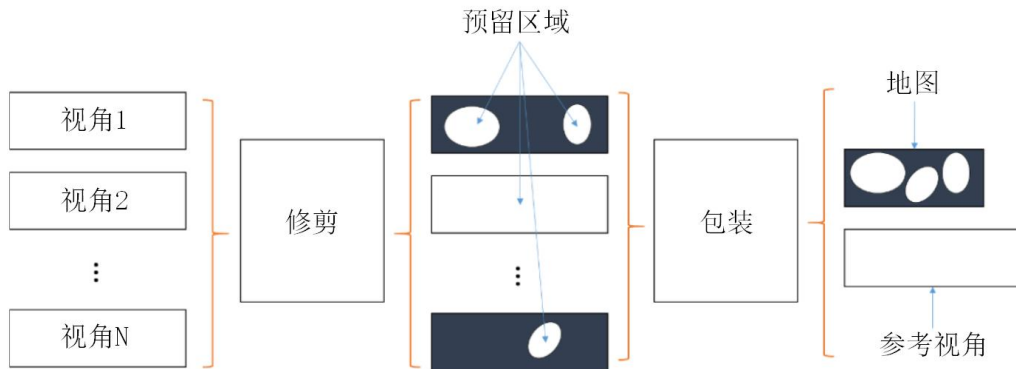


图 22 6DoF 预处理流程

经过预处理器之后的数据通过一个或多个二维视频编码器（例如 H. 265、AVS2）进行传输。

6DoF 编码器可接收场景的多视图表示，包括：每个视图（摄像机）的纹理（颜色）框、每个视图的深度图、描述摄像机的摄像机参数（内部，例如焦距）和摄像机的空间布局（外部，即相对位置和方向）。

6DoF 编码器将上述多视图表示作为视图空间的采样，并传输一个支持视图插值的视图空间表示。

(2) 6DoF 解码

6DoF 解码器能够在保真度损失情况下重建上述所有信息。另外，6DoF 解码器还可以重建中间视点。因此，6DoF 解码器包括二维视频解码器和渲染器。

2.8.3 终端渲染显示

6DoF 视频的典型渲染模型具有以下步骤：

(1) 实时解码一个或多个基本视频流，例如，一个纹理图集和一个深度图集。通常，解码将直接进入 GPU 可访问的内存。

(2) 实时解码元数据，包括源视图参数以及如何将其打包到视频流中（即图集集中的图像小块）。

(3) 呈现帧图像时，接收对视口位置和方向（姿势）的预测。

(4) 将元数据和目标视口姿势转换为 GPU 渲染指令，以并行合成块或面片，例如：重新投影像素以创建带纹理的网格、栅格网格。

(5) 混合在目标视口中重叠的面片。

(6) 修补所有丢失的像素。

(7) 呈现视口。

3. VR 视频国内外技术标准

国内外多个视音频技术标准组织都在 VR 领域进行技术研究和相关标准制定工作。

3.1 VR 视频国外技术标准

3.1.1 ITU

(1) ITU-R SG6

ITU-R SG6 研究组在 2019 年 1 月发布了 Recommendation ITU-R BT.2123-0 “Video parameter values for advanced immersive audio-visual systems for production and international programme exchange in broadcasting” 标准，规范了 VR 视频的基本参数。在此之前的 2018 年 4 月，发布了报告 Report ITU-R BT.2420-0 “Collection of usage scenarios and current statuses of advanced immersive audio-visual systems”，对近年 VR 视频的发展进行了梳理。

(2) ITU-T SG16

ITU-T SG16 组研究制定 ILE 相关标准，已发布 ILE 业务需求、架构、业务场景、媒体传输等标准。

(3) ITU-T SG12

ITU-T SG12 组研究制定服务质量 (QoS) 和体验质量 (QoE) 的国际标准，目前在 VR 领域正在研究制定 “QoE-VR”、“QoE-AR” 和 “使用 HMD 观看 VR 视频的主观评价方法” 等标准。

3.1.2 MPEG

针对 VR 沉浸式音视频数据格式、编码、传输等关键技术，MPEG 开启了 “Coded Representation of Immersive Media ” (简称 MPEG-I) 系列标准 (标准号: ISO/IEC 23090) 的制定工作。按照工作组的计划，标准制定分为两个阶段：第一阶段面向 3DoF，采用现有的 H.265 视频编码和 MPEG-H 三维声音频编码，预计 2020 年完成；

第二阶段面向 6DoF，音视频编解码采用新的编码标准，预计 2021 或 2022 年完成。目前 MPEG-I 系列标准主要包含 9 部分标准：

第一部分：沉浸式媒体架构（23090-1）。该部分规定了后续标准所使用的术语和定义、系统概述、用例描述及与媒体质量相关参数（如音视频信号质量、分辨率、延迟等），目前处于建议的技术报告草案阶段；

第二部分：全景媒体应用格式（23090-2）。该部分规定了全景媒体的编码、存储、传输和渲染，为全景媒体系统提供视频、图像、音频、文字在内的媒体封装接口。支持 3DoF 的版本 1 标准已于 2019 年 1 月正式发布，支持 3DoF+ 的版本 2 目前处于国际标准最终草案阶段；

第三部分：VVC 视频编码（23090-3）。该部分标准由 MPEG 和 ITU 联合制定，重点面向超高清、360° 全景和 HDR 等新型视频提出新的编码技术，目标是相同质量下编码效率比 H.265 提升一倍，目前处于国际标准最终草案阶段；

第四部分：沉浸式音频（23090-4）。该部分提供面向 6DoF 场景下的音频编码技术，目前还处于技术探索阶段，目前处于提案征集阶段；

第五部分：基于视频的点云压缩（23090-5）。该部分面向由点云序列呈现的三维视觉媒体编码，目前处于国际标准最终草案阶段；

第六部分：沉浸式媒体评估指标（23090-6）。该部分针对沉浸式媒体质量和体验评估提出了沉浸式媒体评估指标和测量框架，目前处于国际标准草案阶段；

第七部分：沉浸式媒体元数据（23090-7）。该部分规定了在不同应用和系统环境下统一使用的元数据，包括坐标系定义、映射格式、纹理到球体的映射、旋转参数等，目前处于委员会草案阶段；

第八部分：基于网络的媒体处理（23090-8）。该部分规定了媒体上传至网络的流程、处理节点、媒体数据接入等，目前处于国际标准最终草案阶段；

第九部分：基于几何的点云压缩（23090-9）。该部分研究使用三维几何定向编码技术，主要用于车辆激光雷达、三维绘图、文化遗产和工业等应用，目前处于国际标准草案阶段。

3.1.3 DVB

DVB 成立了 CM-VR 研究组，该研究组的总体目标是把 VR 的商业需求传递给相关的 DVB 技术模块（TM）小组，根据 DVB CM 的规定开发针对 DVB 网络上的 VR 内容

交付的技术规范。

(1) 2017 年, DVB CM 批准了 CM-VR 官方小组, CM-VR 的目标是将商业需求传递给相关的 DVB 技术模块 (TM) 小组, 旨在通过 DVB 网络传送 VR 内容。

(2) 2018 年, 该小组重点调研了 VR/360° /3DoF 内容方面的商业兴趣和需求, 并计划未来关注 AR、MR 和 6DoF VR 方面。

目前, 该工作组已暂停相关研究工作。

3.1.4 SMPTE

SMPTE VR/AR 研究组成立于 2018 年 2 月。该小组对图像/声音捕获和发布的标准化方法的需求进行研究, 以创建 VR/AR 的分发和显示系统, 同时研究可能的标准应用场景; 对 VR/AR 内容制作和后期制作的现有标准化和所需标准化进行差距分析。为其所需的未来标准和工作提出建议, 并形成实践/工程指南。目前, SMPTE VR/AR 研究组已暂停工作。

3.1.5 3GPP

3GPP 是移动通信国际标准制定者之一。在 VR 方面, 3GPP 组织目前已经研究发布了 3GPP TS 26.118 《Virtual Reality profiles for streaming applications》和 3GPP TR 26.999 《Virtual Reality (VR) Streaming Interoperability and Characterization》。

3GPP TS 26.118 标准分别规定了基于文件下载、基于 DASH 流服务的 VR 端到端系统架构, 还规定了 VR 音视频操作点、VR 音视频媒体类等内容。

3GPP TR 26.999 标准规定了 VR 音视频的媒体配置文件, 以及不同的配置选项, 该规范还对 VR 360° 视频应用程序的互操作性基本要求进行了描述。

此外, 3GPP 针对 5G 传输的 FS_XR5G 项目研究涉及数据格式、点云压缩等内容。

3.1.6 JPEG

JPEG 组织正在研发: JPEG XT, 一种描述 360° 全景视频图像编码的技术标准; JPEG XS, 一种具有极低延迟和极低复杂度的压缩算法, JPEG XS 的应用涵盖虚拟现实、增强现实、空间图像、自动驾驶汽车和专业电影编辑。未来, JPEG 将研究应用点云、光场、全息图像等技术。

3.1.7 IEEE

IEEE 1857 工作组针对沉浸视频内容专门立项了 IEEE 1857.9 标准 (Standard for Immersive Visual Content Coding), 该标准提供了高效的编码工具集, 用于压缩、解压缩和重建沉浸式视频内容。目标应用和服务包括但不限于虚拟现实场景, 例如基于无人机的 VR、增强现实、全景视频、自由视点视频、全景立体声视频、以及其他支持视频/音频的服务和应用, 例如沉浸式视频流、广播、存储和通信等。

3.2 VR 视频国内技术标准

2016 年初, 中国数字音视频编解码技术标准工作组 (简称“AVS 工作组”) 启动了 VR 国家标准的制定工作 (以下简称 AVS VR)。该标准的主要目标是定义 VR 视频紧凑表示和编码方法, 促进 VR 设备互联互通, 提升 VR 视频压缩效率。标准的制订工作分为两个阶段, 第一阶段完成 3DoF VR 视频编码, 第二阶段完成 6DoF VR 视频编码。第一阶段已于 2017 年底完成, 目前正在进行第二阶段的研究。

2019 年 7 月, AVS VR 视频编码标准正式被国标委立项, 项目名称:《信息技术虚拟现实内容表达第 2 部分: 视频》, 项目编号: 20192086-T-469。

针对 3DoF VR 视频, AVS VR 定义了如图 23 所示的框架, 该标准规定了 VR 视频映射模型和基于 VR 视频映射模型的 VR 视频重建过程, 支持常用 VR 视频表示模型, 兼容 AVS3/AVS2/AVS+/H.265/H.264 等传统平面视频编码标准。

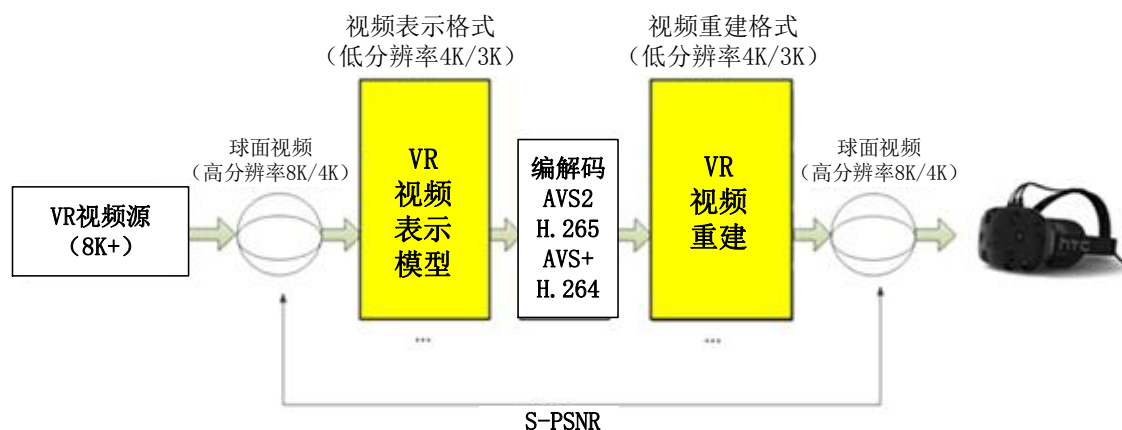


图 23 AVS 3DoF VR 视频编解码框架

针对全视角传输场景, AVS VR 定义了两种映射模型。第一类模型基于 ERP 映射, 支持二维 ERP、三维 ERP 和 ROI ERP 三种映射模型, 其中, ROI ERP 模型允许用户定义感兴趣区域, 对于感兴趣区域内的场景按照全分辨率映射, 对于感兴趣区

域周边的场景逐渐降采样映射，该映射模型可以节约 20%的码率，特别适合舞台类演出场景，如图 24 所示；第二类模型主要针对 360° 视频的展开冗余进行消除，其中等积双极方形对展开冗余的消除效果最好，与 ERP 映射模型相比最多可以节约 30%的码率，如图 25 所示。

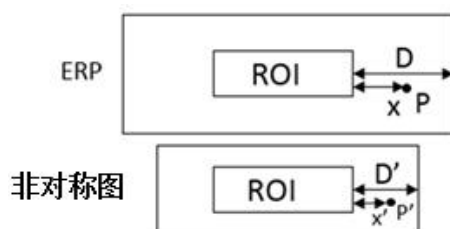


图 24 ROI ERP 映射模型



图 25 等积双极方形映射模型

针对 FOV 传输场景，AVS VR 定义了两种映射模型，其中等角四棱锥映射模型的码率节约效率最高。该模型对 FOV 区域采用全分辨率等角映射，对于非 FOV 区域进行逐步降采样映射，并结合逐级低通滤波进一步降低传输带宽需求。比金字塔或截断金字塔等非均匀映射模型码率可以进一步节约 10%~20%的码率，与 ERP 相比，可以节约超过 70%的码率，如图 26 所示。

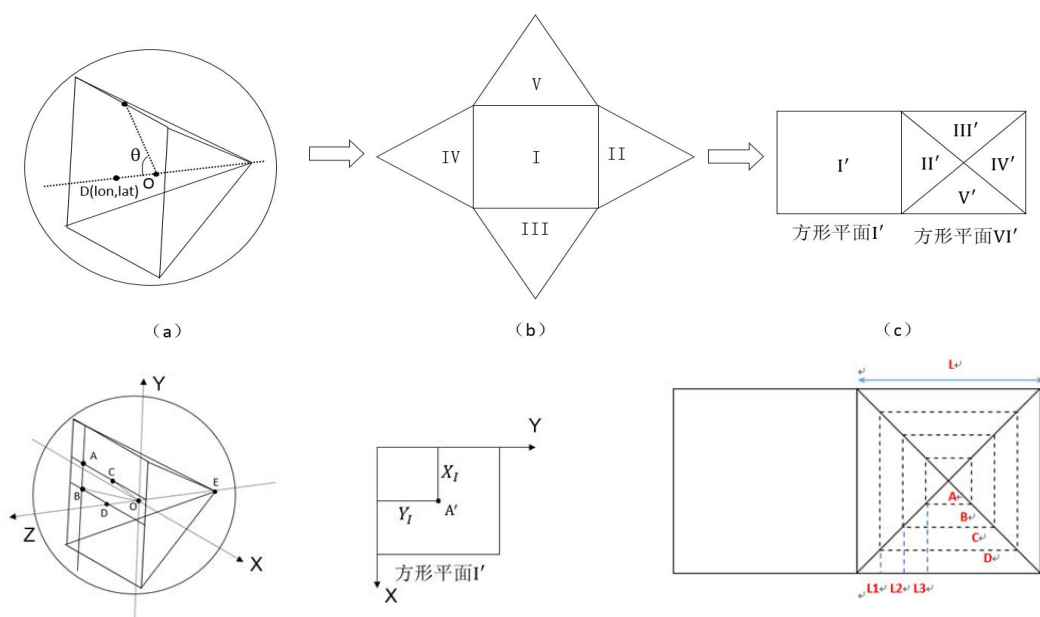


图 26 等角四棱锥映射模型

基于 AVS 3DoF VR 标准的端到端 VR 服务视频系统已在阿里优酷上线应用，VR 视频的传输质量得到了显著提升，可支持 8K VR 视频服务。

最近，AVS 工作组开始进行 6DoF VR 标准的制定工作。AVS 6DoF VR 标准的特点是完全复用了平面视频的编解码过程，和平面视频编解码基础设施完全兼容，可以快速部署实施，基本框架如图 27 所示。AVS 6DoF VR 基于多视角多深度图方案，在内容端采集多视点信息，并基于多视点信息估计出场景的深度信息，编码端将多视点和多深度信息拼接成超高清平面视频，采用新一代 AVS3 等标准高效编码，客户端依据用户想要观看的视点位置，运用 DIBR 算法实时绘制出相应的虚拟视点。目前这套 6DoF VR 视频方案已经率先被阿里优酷应用在 CBA 转播中，并计划在 2022 年北京冬奥会上进行应用示范。

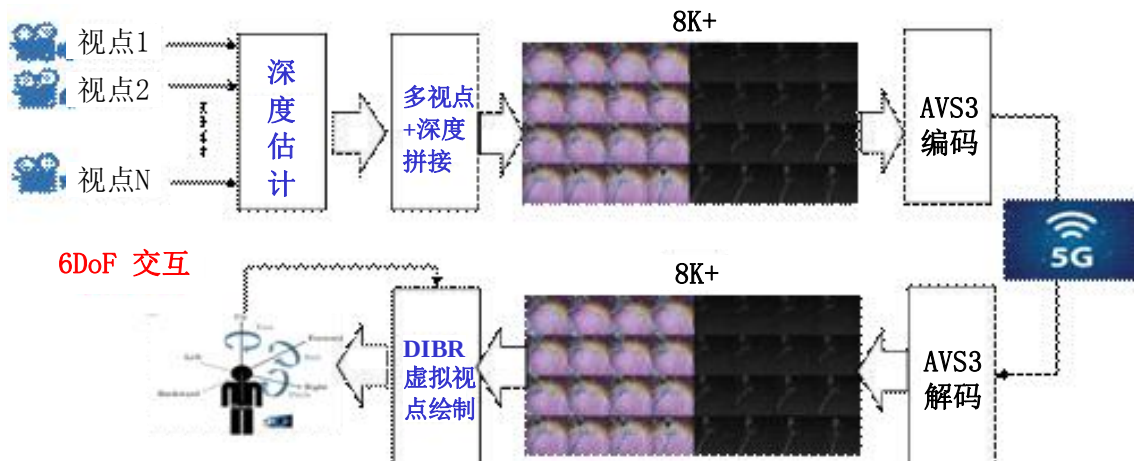


图 27 AVS 6DoF VR 视频端到端框架

4. VR 视频在广播电视及新媒体领域中的应用

国内外广播电视节目制作者、运营商、媒体用户都对 VR 制作和节目非常关注。许多广播电视从业者都将 VR 应用于广播电视节目制作流程中，如：BBC、SKY、Canal+、NBC、NHK、CCTV 等。在传输方面，目前也已经有一些运营商通过 VR 点播频道向用户提供 VR 媒体流，用户可以通过 HMD 观看 VR 节目。

4.1 国外 VR 应用试验

4.1.1 日本

NHK 试行的 VR 项目包括：“NHK VR NEWS”、“Panorama Tour”、“8K VR Theatre”等，VR 节目内容采用 VR 全景摄像机拍摄并播放，可以在 NHK 网站上进行下载。

VR 视频有两个最为关键的问题：一是 VR 全景图像的视频分辨率，另一个是 HMD 的硬件分辨率。这两个问题将直接影响用户观看 VR 视频的效果。为解决上述问题，NHK 使用 $30K \times 15K$ 分辨率的 360° 拍摄及 $8K \times 4K$ 的 OLED HMD 显示。

(1) VR 视频拍摄

360° 高分辨率球面图像使用多个 5472×3648 分辨率的子摄像机拍摄，共计 144 个子画面 (12×12)。随后 144 个子画面被拼接、使用 ERP 投影格式形成一个矩形的 55184×21524 ($55K \times 22K$) 分辨率图像。由于 $30K \times 15K$ 的图像已足以清晰地展示 VR 视频，因而 $55K \times 22K$ 分辨率的图像被下变换至 30720×15360 ($30K \times 15K$)，并在图像的底端插入黑条。

(2) 8K 分辨率的 VR HMD

此 HMD 由 OLED 面板、光学元件、运动传感器单元和图像处理电路组成。运动传感器单元由 3 角度加速度传感器、3 角度角速度传感器和 3 角度地磁传感器组成，能够实时检测用户的观看方向。OLED 面板的尺寸是 103.68×184.32 毫米，左右眼图像的大小为 103.68×92.16 毫米。光学元件的焦距为 38.67 毫米，使 FOV 约为 100° ，与 8K 分辨率显示相匹配。具体参数如表 3 所示。

表 3 8K OLED 屏幕参数

屏幕尺寸	8.33 inch
分辨率	7680 × 4320
像素间距	1058 ppi (24 μm)
帧率	60 Hz

(3) 8K×4K 显示

每隔 10ms，HMD 上的运动传感器就会检测到用户面对的方向。按照方向，重新裁剪球面图像成单眼 3840×4320（4K×4K）的大小，然后对裁剪后的图像进行校正、补偿镜头失真等，最终在 HMD 上双眼显示。

此外，日本电视台、朝日电视台、富士电视台、东京广播电视台等也开展了 VR 的制播试验，通过互联网等传输渠道，向公众提供了 VR 视频节目。早期的 VR 视频节目，视频分辨率普遍很低，清晰度不好。

4.1.2 韩国

韩国进行了云 VR 5G 网络环境试商用验证及 5G 网络 VR 业务拓展验证，搭建了云 VR 平台，支撑云 VR 业务云渲染与云流化能力，打造内容管理、系统服务、终端优化为一体的云 VR 平台。依托 5G 网络，面向用户推广 5G 场景下的云 VR 新业务场景，丰富 5G 的业务领域，增强用户对于 5G 的体验感知。当前项目已经覆盖首尔共计 98 个线下营业厅，为市民提供云 VR 线下体验，并将 5G 云 VR 加入运营套餐中进行客户培育。

韩国的主要 5G 网络运营商，均在积极推动 VR 相关的应用，并将其作为吸引 5G 用户注册的主要手段之一。其中，LG Uplus Corp. 针对不同的用户人群，创造了 5 款不同的应用软件来进行 VR 内容的消费。该公司还于 2019 年 7 月，展示了世界上第一款基于 5G 的 VR 云游戏。

4.1.3 美国

美国 NBC 电视集团在 2018 年冬奥会期间，直播了超过 50 个小时的 360° 视频，同时提供了对应的 VR 重放和集锦。这些节目由奥林匹克广播服务（Olympic Broadcast Service）所制作，并可在各种不同的设备上通过 NBC 体育 VR 应用程序观看。

美国 ABC 电视集团将其 VR 方面的尝试聚焦在增强已有节目的故事体验上，例如新闻类、综艺类节目的幕后花絮及全视角的信息体验。

4.1.4 欧洲

BBC Sport 360 为 2016 年里约奥运会提供虚拟现实报道和比赛视频，为用户开启新的奥运会观看方式。用户可在手机通过安卓、IOS 和三星 GEAR VR 下载安装 BBC Sport 360 APP，匹配使用 Gear VR、Cardboard 等 HMD 即可享受 VR 观看体验。

BBC 还在 2018 年俄罗斯世界杯期间运用 VR 技术进行世界杯直播，提升球迷们的观赛体验。在 BBC Sport VR- FIFA World Cup Russia 2018 APP 中为观众带来 33 场赛事的实况报道，此 APP 兼容 IOS 和安卓系统，用户可自主选择不同分辨率画面，帧率为 50P。

泛欧体育广播机构欧洲体育台在 2018 年平昌冬奥会上首次进行冬奥会 VR 直播。VR 直播内容包括高山滑雪、冰壶、花式滑冰、冰球、短道速滑、高台滑雪、俯式冰橇、单板滑雪以及开闭幕式，此外还以 VR 分享全点播回放和每天的集锦。

4.1.5 国际互联网企业

(1) Facebook

Facebook 是美国的一个社交网络服务网站，Facebook 注重虚拟现实，曾花费 20 亿美元收购 Oculus，其网站支持使用 360 Gear 等设备的用户进行 VR 直播推流。Facebook 与 RED 于 2018 年联合推出三维 VR360 摄像一体机。

(2) 谷歌

谷歌旗下拥有包括 YouTube VR、Earth VR、Cardboard、VR180 等多个 VR 相关产品。2017 年 12 月 15 日谷歌正式为旗下的视频网站 YouTube 推出了 Steam VR 版本，用户可以在 Steam VR 上观看 YouTube VR 虚拟现实视频。而 Google Earth VR 已经收录了来自地球 5 亿平方公里的真实地貌和街景，用户可以戴上 VR 设备足不出户游览亚马逊河、曼哈顿中心、罗马古迹或者是世界上的任何一个角落。同时，用户也可以从太空视角来观望我们所在的这颗蓝色星球，体验漂浮在宇宙中的感觉。

(3) NYT VR

NYT VR 是纽约时报和 Within 合作推出的高质量 VR 新闻视频网站。在 NYT VR

网站客户端中，《纽约时报》将会提供沉浸式的报道，其中包括纪录片、新闻现场视频等。通过虚拟现实技术，新闻企业能够提供 360° 场景的报道，读者可以体验到身临其境的感觉。这种报道方式特别适合某种内容，利用新技术可以引发读者的同情和关注。

(4) Hubblo VR

Hubblo VR 是一个聚焦于可商用化的 VR 直播生态系统，同时推出 VR 直播所需要的各类硬件设备及配套软件，致力于为 VR 用户贡献更多的消费内容。Hubblo VR 提供在线 4K 高质量清晰 VR 视频；支持手机本地二维/三维/360° 全景视频的播放观看；适配目前市场上主流 VR 设备，保证提供良好的 VR 观影体验。

(5) NextVR

NextVR 创办于 2009 年，产品包括以 VR 方式广播体育赛事、游戏等直播类的节目。目前正在直播的节目包括 2019 年 NBA 赛事，2019 国际足球冠军赛等。

4.2 国内 VR 应用试验

4.2.1 电视台

中央广播电视总台通过成立 VR 体验实验室，开展 5G+VR 技术的研究与应用，开通 VR 频道等 VR 技术应用试验，成功完成了多档 VR 节目的拍摄和制作，并成功实现了 VR 内容的 5G 直播，为用户带来多元、全新的视听体验。

(1) 基于 VR 技术的台内节目制作应用

中央广播电视总台 VR 体验实验室于 2016 年建设成立，是国内电视台首次成立专门研究 VR 技术应用与节目制作的实验室。实验室成立至今完成了针对不同节目形态的 VR 拍摄和制作，包括：体育频道《CCTV 体坛风云人物颁奖盛典》、《CBA 全明星赛》、《排球全明星赛》、《亚洲乒乓球锦标赛》、《扬州鉴真国际半程马拉松赛》、《天津全运会》；大型节目中心《2017 年春晚》、《2018 年春晚》、《五一国际劳动节心连心特别节目》—“中国梦·劳动美”；综艺频道《艺术人生》、《五一乐三天》；中文国际频道《中国舆论场》、《国家记忆》；科教频道《中国记忆》、《百年地理大发现》；财经频道《美丽中国行》；音乐频道《全球中文音乐榜上榜》、《一起音乐吧》、《音乐公开课》、《四季音乐会》；戏曲频道《盛世梨园》、《空中剧院》；少儿频道《最野假期》等多个频道的多个节目。节目类型包括体育赛事类、大型综艺节目晚会类

以及专题节目类，累计播放条数超过 350 条，素材总时长累计超过 120 小时。

(2) 5G+VR 技术在台内的研究与应用情况

2018 年，“5G 新媒体平台”在中央广播电视总台开建，技术系统搭建了基于 5G 网络的移动端呈现模拟实验环境和 VR 制作整体试验系统，完成了基于 5G 的超高清 VR 信号从制作到分发的全流程调试，在实验室环境下首次实现了 5G+VR 直播。2018 年 12 月 28 日的“新闻直播间”对此进行了报道。

(3) 新闻频道《我要看春晚》5G+VR 直播试验

新闻频道《我要看春晚》节目是一档以春晚前期预热宣传为主的强时效性新闻节目，每年在春晚前期彩排阶段进行直播，通过多内容、多样态和多种制作手段组合的方式对春晚主会场及各地分会场重点报道。

此次直播选取 5G 信号覆盖较好的长春和深圳分会场，构建 8K VR 信号采集和传输链路。通过 5G 接入终端实现无线采集现场多角度的景观信号并实时拼接缝合，然后将超高清 VR 视频信号通过 5G 基站及网络传输至北京的中央广播电视总台新闻演播室，实现电视端直播过程中的 VR 视频交互点评，带给观众通过电视画面首次全景预览春晚分会场的沉浸体验。另外，VR 内容也将通过中央广播电视总台新媒体内容集成发布平台分发至央视新闻移动端客户端及 HTML5 页面等进行同步直播，有效扩展春晚相关内容的融合传播途径，如图 28 所示。

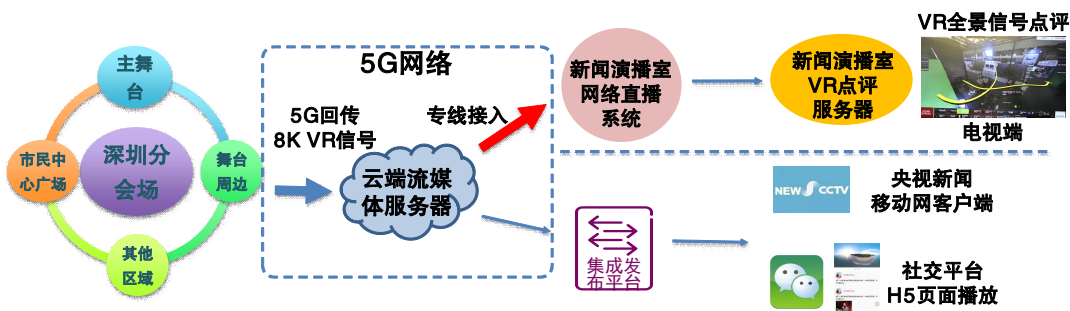


图 28 新闻频道《我要看春晚》5G+VR 直播技术方案

(4) 体育赛事类转播《2019 国际体操联合会成都跑酷世界杯》5G+VR 直播试验

中央广播电视总台在 2019 年 4 月 6 日至 7 日《2019 国际体操联合会成都跑酷世界杯》节目中实现基于 5G 网络传输制作的 VR 体育赛事直播，通过体育频道 CCTV5 APP 客户端完成多机位信号播出，这是 5G 技术和超高清 VR 制作技术首次结合应用于体育赛事转播中。

根据跑酷运动的特点，技术团队利用 5G 低延迟、大带宽、大容量的优势，设计了整体 VR 视频直播方案，从技术上缩短 VR 视频实时信号从采集制作到分发端的时延并提升其清晰度；另外赛事相关 VR 视频内容的精编制作涵盖了主办城市风光、赛场周边花絮、VR 视频航拍以及独特的参赛选手第一视角等，如图 29 所示。

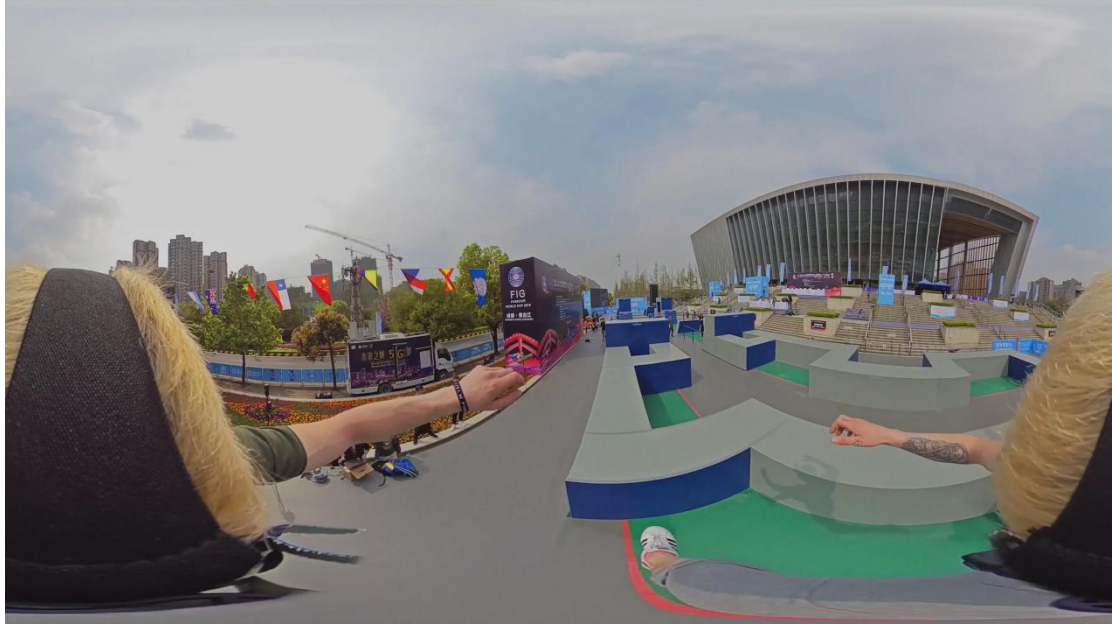


图 29 参赛选手 VR 视频第一视角

VR 视频定点、移动拍摄和航拍技术，将城市风貌和比赛相关内容全方位多角度的展现，使用户能够近距离感受比赛氛围。

在具体直播技术方案的设计实施过程中，构建了适配超高清 VR 视频信号在 5G 网络的采集、传输、制作链路，在比赛区域拍摄现场多角度信号并实时拼接缝合，再将多路超高清 VR 视频信号通过前方 5G 基站及网络传输至北京的总台光华路办公区 5G 媒体应用实验室，在后方完成 VR 视频的切换、字幕包装、特效等全部环节，制作完的信号通过总台新媒体内容集成发布平台分发至 CCTV5 APP 客户端进行直播，如图 30 所示。

此次 5G+VR 赛事直播是继总台首次成功实现 5G 网络 VR 实时制作传输测试后，持续优化技术流程，并结合具体节目需求，实现 5G 新技术应用层面的又一次重大突破。

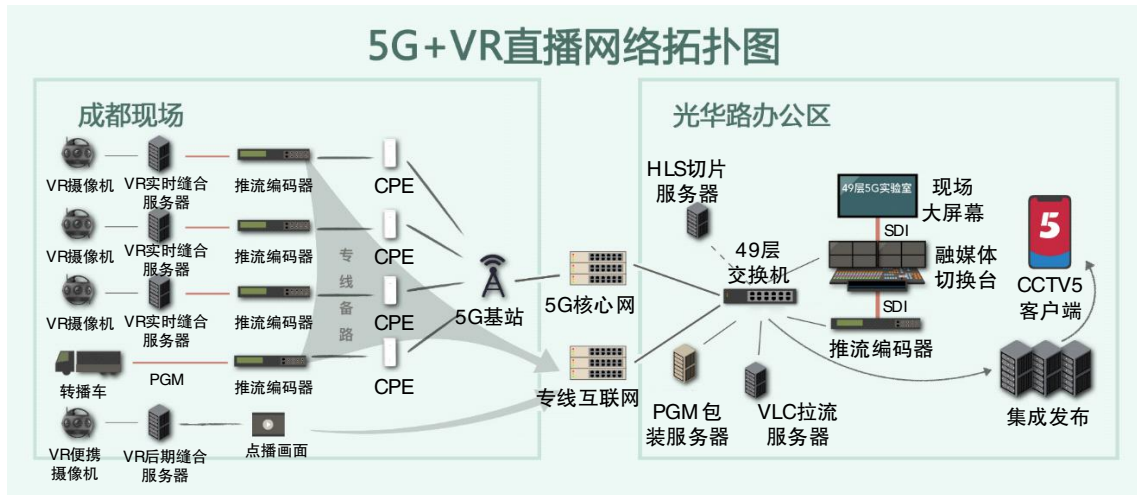


图 30 5G+VR 直播制作系统示意图

(5) 庆祝新中国成立 70 周年活动 VR 视频直播技术的应用

结合庆祝新中国成立 70 周年活动的转播需求，中央广播电视总台首次将 VR 直播系统应用在国庆活动的宣传报道中，通过在天安门广场前方部署的两台 VR 摄像机直播拍摄。回传至总台光华路办公区 5G 实验室的 VR 直播制作系统，进行场景选择切换制作。最后将 VR 直播信号传输至集成发布平台，在客户端等 VR 直播平台播出，为观众提供沉浸式直播 VR 视频全景体验。

在实施过程中，为了保障信号拍摄的安全，总台的 VR 技术团队进行了多次不同型号的拍摄设备试验，从需求调研入手，遍历可能出现的应用场景，最终确定前场四台 VR 摄像机的选型，并从冗余设计原则出发分为两组采集单元，一主、一备进行拍摄回传。为了简化前场系统的配置难度，经过反复测试，确保在 VR 摄像机内进行实时缝合拼接成 3840 像素分辨率的全景视频，并高码率回传至总台光华路办公区 5G 媒体应用实验室的后方直播切换系统进行高质量信号制作。

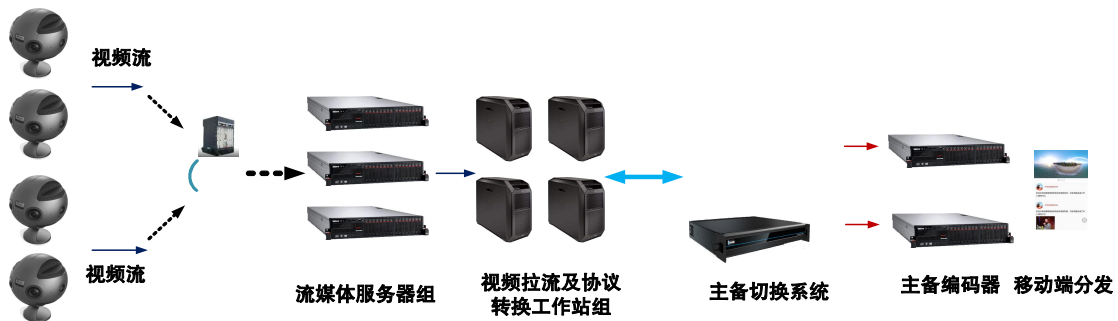
相比较之前的 VR 制作，此次活动在整体传输链路和制作环节的每一个节点都需要充分考虑如何确保系统实施的安全，传输链路通过主备专线将信号传输至后方流媒体服务器，其中一路传输北侧主机位和南侧备机位信号，另外一路传输南侧主机位和北侧备机位信号，以确保直播过程中一路信号出现问题，可以在切换台切换至另一路信号，继续进行直播，直至故障排除。

为了合理分配后方的切换制作系统的输入协议转换资源，技术团队搭建了两台流媒体服务器转发四路采集信号，再由四台协议转换服务器进行协议转换和录制。将 RTMP 视频流转换为 NDI 协议。由协议转换服务器 1 转换北侧主路 VR 视频信号，

由协议转换服务器 2 转换南侧主路 VR 视频信号，由协议转换服务器 3 转换北、南侧备路 VR 视频信号，由协议转换服务器 4 录制北、南侧主路 VR 视频信号，以此确保信号传输在逻辑和功能配置上的高效、安全。

在切换制作单元实现了全 IP 化的系统配置，由网络切换台在后方完成 3840 分辨率配置，高质量远程切换制作。切换原则以北侧机位为主要场景，南侧机位为次要场景。同时在后方工作区还可以通过 IP 链路实时传输节目信号至全景触摸屏进行裸眼监看现场情况。

制作完成的信号输出由两台专用编码器进行编码，传送至集成发布平台发布至不同的新闻新媒体客户端以及 HTML5 页面。直播技术架构图如图 31 所示。



4.2.2 有线电视网络/IPTV

(1) 大连天途

大连天途云 VR 平台依托视频网络资源，与云 VR 技术平台相结合，完成了云+端的方式提供 VR 业务。VR 应用在云端运行，通过视频流将 VR 应用的画面传送到终端机顶盒并通过 VR 眼镜进行展现。如图 32 所示。



图 32 天途云 VR 端到端技术架构

在编码环节，VR 视频分辨率为 3840×1920 ，使用 H.264 编码，量化精度 8bit、帧率 25Hz、码率 45Mbps。

传输环节,云 VR 平台将 VR 应用的实时视频推送到用户机顶盒,并将操作指令、头部运动信息等数据回传到云端服务器。

终端接收显示采用的是 VR 专用机顶盒,配置如下:Hi3798CV200 芯片、2G DDR3、4G Flash、HDMI 2.0 接口。

大连天途云 VR 产品采用专用机顶盒加外设的形式,在互动业务连通的环境下,通过云 VR 产品专用的机顶盒对 VR 内容进行解码,由 VR 外设进行头显和操作。外设目前主要为 VR 眼镜,支持高端设备 HTC VIVE VR 眼镜和中低端品牌小派 4K VR 眼镜两款。

云 VR 产品内容包括 VR 游戏、VR 视频、行业应用三部分,游戏类首批上线 20 款优质游戏,每月更新 2 款;视频包含风光类、体育类、记录片类、演唱会以及自制短视频,共计 100 部左右,总时长超过 100 小时,以及 10 款行业应用。

(2) 新疆天山云 VR

新疆天山云 VR 依托现有广电互动点播系统,用户可通过该系统在天山云二代机顶盒上观看 VR 节目。在视频中加入手机端 HTML5 页面二维码,用户可以通过手机扫码的方式,在手机端进行查看。用户也可以直接通过遥控器操作上下左右,选择不同的观看视角进行观看。

新疆广电网络采用“4K+全景”的方式,将全景业务与 4K 业务整合成统一的产品,用户在首页进入后,可以进行相关内容的收看。内容分类包含:影视、综艺、体育、风景、院线(单点付费)。节目来源主要通过子公司未来媒体输送,更新方式为每周更新,内容更新量 8~10 部不等。

(3) 深圳天威

深圳天威视讯 VR 视频业务点播平台主要包含 VR 管理系统、VR 播发系统、VR 终端系统三部分。系统总体架构图如图 33 所示。

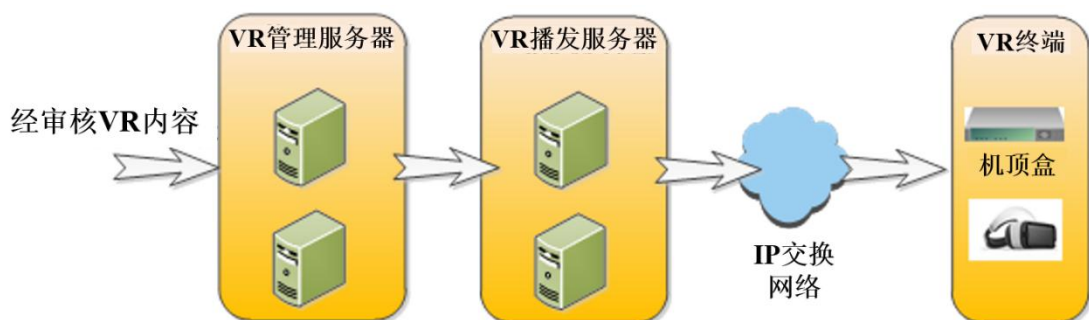


图 33 深圳天威视讯 VR 视频业务点播系统架构

a. 内容审核系统

由市场产品部对 VR 视频内容进行三审一查流程，以保证符合播出要求。其中终审和抽查将在 VR 前端工作站完成。

b. 管理系统

完成对 VR 视频原始内容的注入、存储和管理，提供 VR 业务的内容管理、业务管理、鉴权计费等功能，提供 VR 业务多屏适配及多屏导航。

c. 播发系统

VR 点播播发系统是 VR 播控平台业务能力层，通过 HTTP 方式提供终端视频业务。在试运行过程中，以标准 HLS 的方式向内容分发网络提供视频源。视频源是基于 HTTP 以 IP 的方式传输到终端系统。

d. VR 终端

通过在 1100 台 4K 机顶盒上集成独立的 VR 应用，可从数字电视网络中接收前端传送的视频流，对其进行解码、渲染等一系列处理，将视频内容展示给用户。同时还可以通过 Wi-Fi 将视频流送入移动设备进行观看。

深圳天威视讯 VR 视频业务可通过电视和手机两种载体进行呈现。点播应用播出的是 VBR 15Mbps 码率的 VR 流，该 VR 流用于在电视中呈现；VBR 8Mbps 码率的点播 VR 流用于在手机中呈现。

手机移动端通过扫描机顶盒全景页面的二维码下载安装 VR 客户端 APP，即可在家庭局域网内通过使用手机观看 VR 视频。当手机和机顶盒都连到同一个无线网，点击进入客户端应用，即可看到列出的可点播的 VR 影片资源，点击就可观看 VR 影片。

深圳天威视讯 VR 点播视频业务共包含 60 部短片（二维视频 30 部+三维视频 30 部），全部免费提供给用户观看体验。VR 应用上线后的运营试播时间为期 3 个月。

4.2.3 国内互联网企业

(1) 爱奇艺 VR

爱奇艺 VR 频道是一个 VR 虚拟现实综合媒体平台，提供高品质 VR 视频与行业热点一站式体验。爱奇艺 VR 拥有海量优质 VR 视频在线观看，更有 VR 行业热点资讯与 VR 游戏、硬件测评等综合性内容，一站式入门 VR 行业。

(2) 优酷 VR

优酷 VR 直播平台提供包括演唱会、话剧、音乐会、发布会等内容的 VR 视频，支持跨平台全景格式播放。

(3) UtoVR

UtoVR 提供 VR 视频片源下载，也可在线浏览 VR 视频内容，并提供 VR 视频拍摄、VR 视频拼合、VR 视频制作和 VR 视频解决方案。UtoVR 是 CNTV、第一视频、咪咕视频、浙江联通、天山云 TV、环球世界小姐中国区大赛等 200 多家视频平台及网络电视台全景视频播放技术及 VR 拍摄服务的提供商。

(4) Insta360

Insta360 创办于 2014 年，产品包括 Insta360 ONE、Insta360 Nano、Insta360 Air 等消费级全景摄像机以及 Insta360 Pro 专业级全景摄像机。Insta360 旗下产品覆盖 VR 全景影像专业级与消费级市场，针对政府、企业和普通消费者形成了多条产品线和解决方案。广泛应用于旅游拍摄、新闻直播、赛事纪录、活动拍摄等多个领域。

(5) 腾讯 VR

腾讯视频云与多媒体实验室推出的 VR 端到端解决方案，提供 8K 全景实时采集拼接 SDK，通过云上加速节点实现稳定一键推流，配合云端投影转换、转码、多机位切换、视频存储备份、导播、转播等功能，以及腾讯云全球优化 CDN 分发节点，提供高质量 VR 流媒体视频服务。终端提供跨平台 SDK，支持多种投影格式及切块技术的 VR 互动播放。

5. VR 视频端到端解决方案

5.1 VR 视频端到端解决方案系统架构

VR 视频端到端解决方案系统架构包括 VR 视频制播平台、VR 视频业务平台、传输网络、VR 视频终端，如图 34 所示。

VR 视频制播可分为直播场景、录播和点播场景两类。对于直播场景，现场采集的 VR 视音频信号经过 VR 视频实时制作，完成播出分发。对于录播和点播场景，VR 视音频素材经过剪辑、包装、校色等处理后输出成片。

VR 视频业务平台包括 VR 视频直播系统和 VR 视频点播系统两部分。在直播管理系统和点播管理系统的基础上，VR 直播流和 VR 视频源通过转码、切片后形成播出流，随后推送至传输网络。

VR 视频传输网络包括：5G 网络、有线电视网络、IPTV 和互联网等。

VR 视频流传送至家庭网关/机顶盒后，通过 VR 视频终端的 Wi-Fi 模块进行接收，随后进行解码、视频帧处理和头部动态信息操作控制等，最终完成 VR 视频的显示。

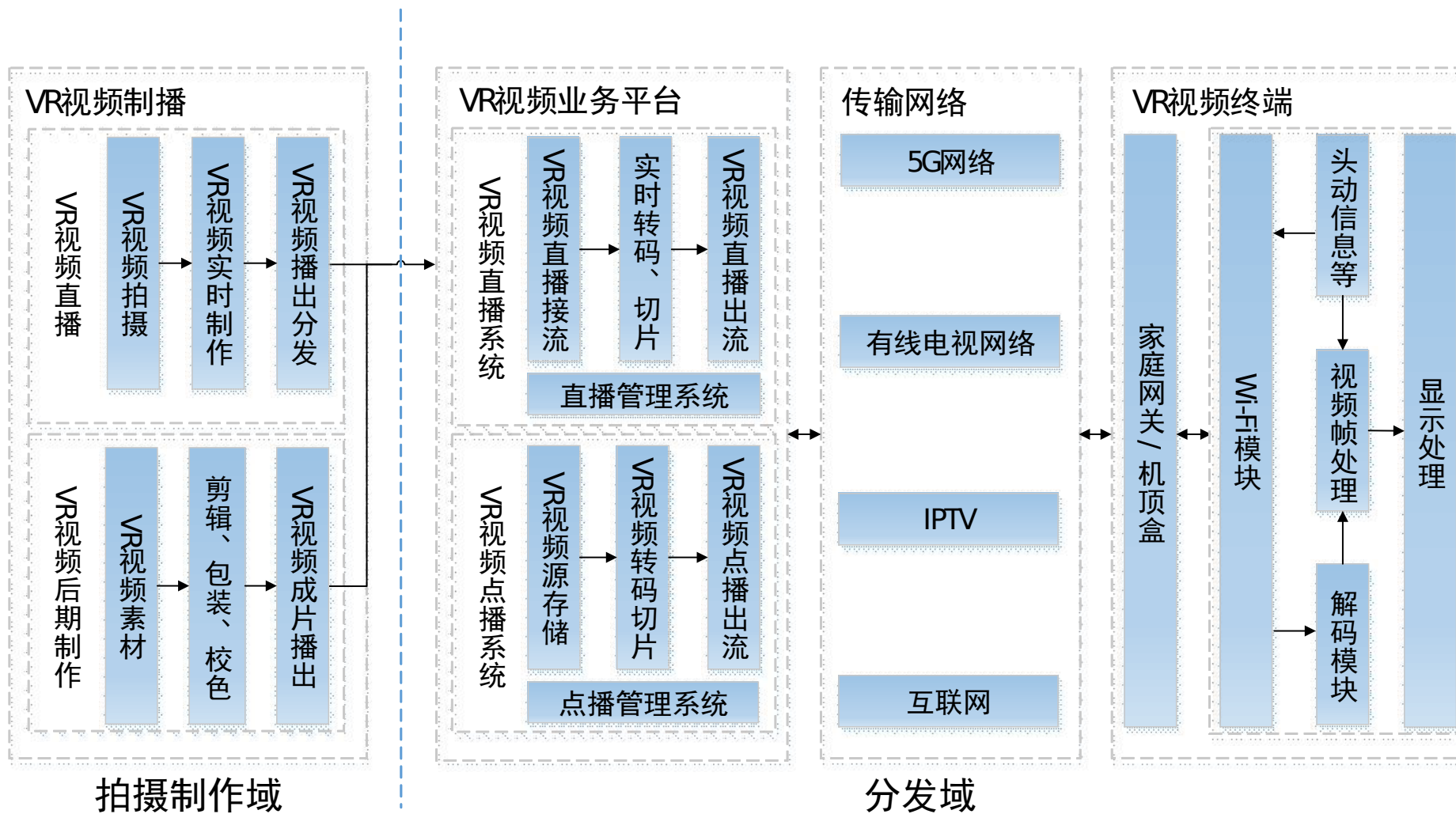


图 34 VR 视频端到端解决方案系统架构

5.2 VR 视频拍摄制作域解决方案

5.2.1 VR 视频拍摄制作工艺流程

通过提炼总结 VR 视频拍摄制作过程的关键工作节点,形成 VR 视频拍摄制作工艺流程。当前 VR 视频业务主要分为直播和录播、点播两类场景,两种场景分别对应不同的工艺流程。

对于直播场景,通过 VR 摄像机采集拍摄全景画面,同时完成实时的音频嵌入和全景视频的拼接缝合,再将视音频信号同步进行适配于网络传输的压缩编码,然后通过固定网络或移动网络将高码率的视音频流信号回传至 VR 切换制作系统,经过解码转换为可适配视频制作系统的信号,在后方的制作系统内完成 VR 信号收录、播放、在线包装以及必要的实时调色等处理,最终通过播出设备将制作好的信号进行面向客户端的分发播出。面向直播的 VR 视频拍摄制作工艺流程如图 35 所示。

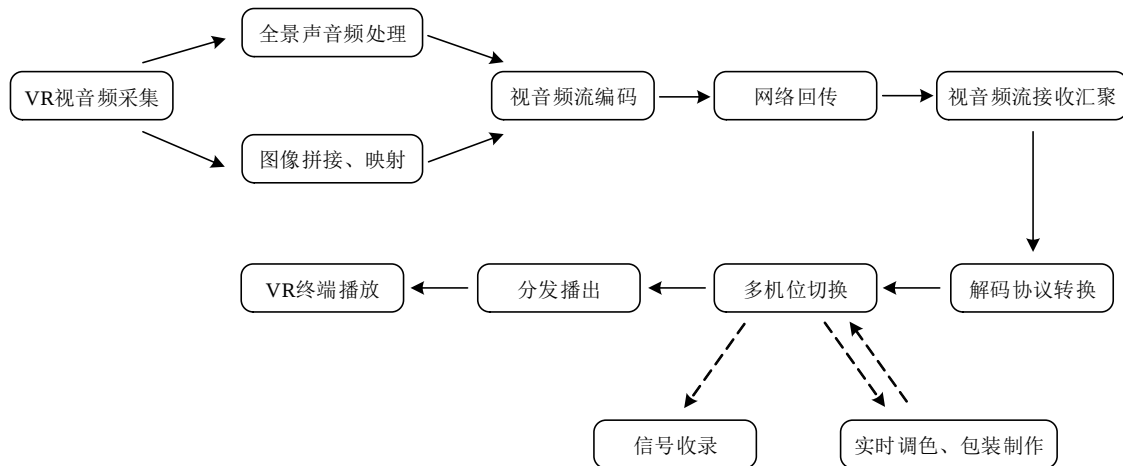


图 35 VR 视频直播制作流程示意图

对于录播和点播场景,VR 视音频采集后对画面进行一级调色,调色后的 VR 图像经过拼接、映射、移动镜头增稳等处理后,依次进行剪辑、包装和二级调色,最后输出成片,用于后续录播和点播使用。面向录播和点播的 VR 视频拍摄制作工艺流程如图 36 所示。

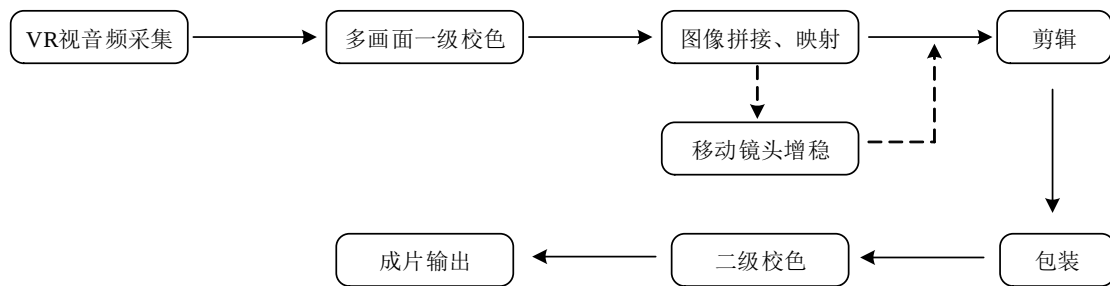


图 36 用于录播和点播的 VR 视频制作流程示意图

在 VR 视频拍摄制作流程中，各重要节点均应有专业的技术人员进行技术指标和艺术效果监看，如前端 VR 摄像机拍摄/拼接实时预览，剪辑调色输出信号实时监看等。

5.2.2 VR 视频拍摄方式

根据拍摄位置的不同，VR 视频的拍摄一般分为地面拍摄、航拍、水下拍摄等方式。

(1) 地面拍摄

a. 固定机位拍摄

VR 视频固定机位拍摄，是指将全景摄像机固定在一个位置，所有内容围绕摄像机进行布置，并以摄像机为原点进行采集。固定机位拍摄的技术要点如下：

- 考虑摄像机的固定点。通常将摄像机架在一个较为空旷的位置，从而使整个空间分布比较均匀辽阔，减少因放置不当造成的空间紧张感和局促感，但有时为了特定的创作意图和情绪表达，也会选择在较狭窄或近距离的空间拍摄。摄像机固定机位一般可以选择在具有图片意义的地方，比如风景的制高点、谷底等。此外，摄像机固定点选择时应优先考虑摄像机的安全距离，因为在安全距离以内拍摄的视频在后期拼接时容易出现拍摄对象畸变、缝合不完整等情况。目前这个问题还无法有效解决，只能通过前期拍摄、后期拼接来减少拼接接缝处的瑕疵。因此，无论双目摄像头，还是四目摄像头，在实际拍摄中都应尽量避免将重要的拍摄对象置于镜头交界处。一般设备交界处的安全距离在 1m~2m 之间。
- 考虑拍摄的高度。VR 视频的拍摄高度代表用户的观看高度，在布置固定机位时，需要考虑拍摄内容呈现给用户怎样的观看视角，从而根据预设脚本的用户视角选择固定摄像机的高度，以确保用户的最佳观影效果。

- 考虑光源的问题。VR 视频的拍摄无法使用传统打光设备进行补光。因为打光设备在空间中任何位置放置，都会出现在摄像机镜头中。可采取的解决方案有：在拍摄空间增加几个常规灯源，或者在拍山洞或者溶洞等场景时，可在三脚架下藏一个光源，但要提前进行拍摄测试以避免出现光晕和光圈。
- 考虑现场拍摄人员和设备所处位置。由于 VR 视频是 360° 全景拍摄的，因此现场环境的人和事物都会出现在摄像机镜头下。为避免穿帮，应提前考虑拍摄人员和设备隐藏的位置，既能随时看到摄像机的拍摄内容又不影响拍摄，比如某个建筑后面，或者为导演和摄像师提前安排好一定的角色。最好先拍摄小样，手机端预览效果调整到最佳再进行实际拍摄。

b. 移动机位拍摄

全景摄像机的摄像头是定焦的，没有传统视频拍摄的特写镜头和走步移位。因此，在一些 VR 视频拍摄过程中，会出现随着位置的改变切换镜头的情况，此时需要移动机位拍摄技术。通常，移动机位拍摄需要使用平衡车，头戴拍摄设备、车辆、现场的索道等均可作为固定摄像机的设备，还要事先设计好摄像机的移动轨迹，既要避免穿帮又要方便后期制作。在拍摄过程中，移动设备的速度最好低于 5km/h，以防因为移动设备的抖动造成后期视频抖动，引起用户观影的不适。

(2) 航拍

在特定拍摄需求下，会使用无人机进行航空拍摄。航拍过程多采用无人机带 VR 全景摄像机的方式进行，在拍摄过程中，需要根据所使用 VR 全景摄像机的重量、大小等选择对应的无人机，确保整个拍摄的安全裕度。此外，高空 VR 拍摄的稳定性对最终呈现的画质有很大影响，需要针对不同的航拍飞行器采用可适配的方式，挂载舵机平衡仪，实现空中机械增稳，在后期的多画面缝合编辑阶段辅以软件追踪增稳，这是实现 VR 航拍前后期配合的稳定方案。

(3) 水下拍摄

VR 视频的广播级水下拍摄与传统水下拍摄有很大不同，传统设备的机身防水壳一般不会影响到镜头拍摄，而 VR 视频拍摄设备因多目镜头都为超广角镜头，所以防水壳需要紧贴设备机身，且镜头处的透镜玻璃需要有较高的光学反畸变能力。目前 VR 视频拍摄设备的防水模式主要通过特定产品的特殊防水外壳进行适配，在

设计上一般通过 VR 摄像机配合光学反畸变透镜的方法实现，通常可在水中下潜拍摄 10m 深度的范围。

5.3 VR 视频业务平台方案

VR 视频业务平台的总体目标是以 VR 视频业务为中心，在 5G、有线电视网络、IPTV、互联网等网络条件下，以直播/点播平台为基础，依托视频处理算法和云计算服务以及专业的直播技术，面向用户提供高质量、优质体验的 VR 视频服务。

5.3.1 VR 视频直播/点播平台

(1) VR 视频直播/点播平台基本架构

VR 视频直播/点播平台基本架构以云计算的三层架构来表示。VR 视频直播/点播平台方案涉及对前端采集制作设备的支持、直播、点播、转码、存储、分发和播放等具体功能，如图 37 所示。

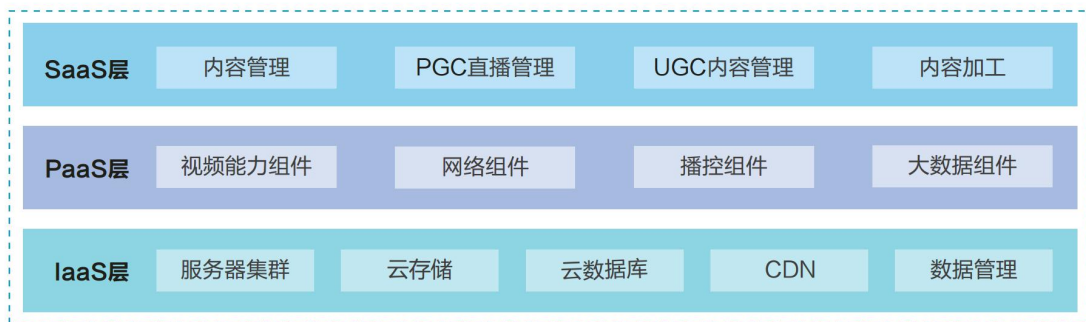


图 37 VR 视频直播/点播平台基本架构

IaaS 层主要包含服务器集群、云存储、云数据库、CDN 与数据管理，实现 VR 视频直播/点播业务存储、网络、服务器以及相应处理能力的自动部署，具备高度的动态可扩展性。

PaaS 层主要包含视频能力、网络、播控和大数据组件，其中视频能力组件包括在线转码、直播录播等功能；网络组件包括多 CDN 支持、DNS 智能调度等功能；播控组件包括图像鉴定、视频比对等功能；大数据组件包括播放数据分析反馈、CDN 分析等功能。

SaaS 层主要包含内容管理、PGC 直播管理、UGC 内容管理、内容加工等组件，其中内容管理包括上传管理、编目管理等；PGC 直播管理包括频道管理、编码管理等功能；UGC 内容管理包括实时监控、内容检测等功能；内容加工组件包括逻辑打

点、在线快编、云剪辑等功能。

(2) VR 视频直播/点播平台提供服务

VR 视频直播/点播平台提供 VR 视频直播、VR 视频点播等具体服务。

a. VR 视频直播

VR 视频直播/点播平台将在 VR 视频直播场景下提供下述功能：

- VR 视频直播源接收
- 多格式在线转码
- 用于直播内容的多协议输出播出
- 直播推流
- 提供 RTMP 直播推流地址
- 对 VR 视频直播进行实时监控
- 支持多分辨率录制并存储直播视频流

b. VR 视频点播

VR 视频直播/点播平台将在 VR 视频点播场景下提供下述功能：

- VR 视频文件上传
- 多格式在线转码
- 自定义编码格式和编码参数
- 选择 VR 视频回看
- VR 视频内容标注

5.3.2 VR 视频直播/点播平台技术指标

VR 视频直播/点播平台技术指标包括 VR 视频节目制作与交换中的视频参数指标、VR 视频编码指标、传输协议与播放控制等，各模块具体参数指标如下：

(1) VR 视频节目制作与交换中的视频参数

高标准、高质量的 VR 视频节目内容将带给用户更真实、更舒畅的观看体验，参照 2.2 节，映射后图像有效像素数至少为 7680×3840 (8K \times 4K)，比特深度至少为 10bit。

(2) VR 视频编码

码率是影响 VR 视频在终端接收观看清晰度的重要指标。VR 视频直播码流支持 AVS2、H. 265、H. 264 等主流视频编码格式。对于 8K/50P/10bit 的 VR 视频，当采

用 H.265 或 AVS2 编码时，码率为 80~100Mbps。

(3) 传输协议

VR 视频直播/点播平台支持 HTTP、HTTP-HLS、HTTPS、RTMP 等网络传输协议。在 VR 视频直播场景下，还需支持 VR 视频直播流透传功能，可实现 RTMP 格式码流和 HLS 格式码流的相互透传。

(4) 播放控制

VR 视频直播/点播平台在视频播放控制方面，需支持通过陀螺仪控制播放角度、通过手指滑动选择观看角度、支持视频平铺/双目显示/单屏显示多种 VR 视频显示方式、支持多码率 HLS 流清晰度切换，以及支持观看焦距调整。

5.3.3 VR 视频直播/点播平台部署

(1) 高性能服务器

VR 视频直播/点播系统需要部署不同的子系统来配合实现 VR 视频的接入、转码、发布、存储以及媒资运营等功能。平台侧需要能提供多个子系统相应的物理机或虚拟机资源，包括：CPU、内存、硬盘、网卡等，以 8 核 CPU 主频 2.5GHz 及以上、内存 32GB 或以上配置进行各子系统的安装部署。

其中 VR 视频直播系统由于对直播流做实时转码，需要配置更高性能服务器，如采用 16 核 CPU、内存 64GB 或以上配置，以及配备专业显卡如 Nvidia Tesla P4，硬盘方案推荐使用 RAID0+1 或 RAID0 等。

(2) 松耦合、可扩展的架构

VR 视频直播/点播系统采用松耦合、可扩展的架构，各子系统模块均可扩展，当其中某一模块成为性能瓶颈时，只需增加运行该模块的服务器数量，而不会影响其他模块。

(3) 系统可靠性

VR 视频直播/点播系统采用冗余容错的架构，将所有模块都冗余部署，使得某一服务器故障不会导致系统服务中断。并部署完善的系统监控功能，无论是硬件还是软件故障，都能够及时告警。

5.4 VR 视频网络传输方案

VR 视频传输主要有两种技术路线：全视角传输方案和基于 FOV 的传输方案。这两种方案可应用于 5G 网络、有线电视网络、IPTV、互联网等传输网络中，其中

基于 FOV 的 VR 视频传输方案要求传输网络具有双向通信能力。

5.4.1 基于有线电视网络/IPTV 的 VR 视频传输方案

面向 VR 视频直播和 VR 视频点播场景，在现有有线电视网络/IPTV 视频服务平台基础上搭建对应的 VR 视频业务支撑系统，并在机顶盒端集成 VR 播放器，适配 VR 一体机来实现 VR 视频服务。

5.4.2 基于 5G 网络的 VR 视频传输方案

针对 VR 视频直播/点播高并发、大流量等业务特点，可在目前网络基础上进行承载优化和体验管理增强，部署 QoS 保障、家庭高性能 5G Wi-Fi、UDP 组播等方案降低网络负载，保障用户体验。

5G 网络将可以随时随地提供 100Mbps 以上大带宽，满足 8K VR 视频分辨率的高清画质体验，同时 5G 网络的 5~8 毫秒级的网络时延将消除 VR 视频观看的眩晕感。因此 5G 网络是 VR 视频传输的优选方案。以下主要描述基于 5G 网络的 VR 视频传输方案。

(1) 业务逻辑架构

我国 5G 商用牌照已发放并逐渐进入规模商用阶段，5G 的大带宽与低时延为 VR 业务高速发展提供了良好的支撑。基于 5G 网络的 VR 视频传输是依托应用流化和云计算技术，采用先进的视音频压缩编码技术将云端的显示输出和声音输出传送给终端的一种云计算方案。通过将云计算和 VR 视频技术有效结合，解决了传统 VR 视频产品架构面临的 VR 视频应用承载和展现的问题。渲染上云的技术方式可以使用户即使在小带宽、低硬件配置下也能体验好的 VR 视频。

基于 5G 网络的 VR 视频传输架构由中心云、5G 网络、边缘云、终端四个部分组成，如下图所示：

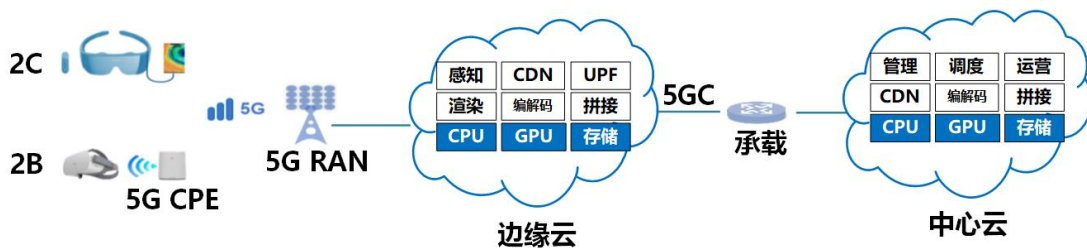


图 38 基于 5G 网络的业务逻辑架构

(2) 中心云

VR 视频业务渲染和计算能力下沉到边缘云后，中心云主要承载 VR 视频业务的用户管理、资源调度、内容运营等管理能力及 CDN 存储与分发能力。

(3) 5G 网络

5G 网络对于 VR 视频系统，一方面可以提供前端采集的无线上行方案，可以大大提升系统部署的灵活度，降低有线部署时对场地的特殊要求和布线难度；另一方面可以提供更大的下行带宽，在下行传输当前 VR 视频观看视点的同时，可以同时打包下发周边多个视点的视频内容，相比 4G 网络下的交互体验更为平滑自然。

基于 5G 网络的 VR 视频传输网络包括 RAN、承载和 Core。5G RAN 需要综合考虑信号的覆盖、容量、接入性、移动性和稳定性；5G Core 需要基于 VR 视频业务时延的需求，选择下沉的位置；网络侧通过构建端到端单独的切片，并借助切片资源隔离和差异化 SLA 保障 VR 视频直播/点播业务的需要。

(4) 边缘云

5G 网络的大带宽和低时延不仅为 VR 视频业务的下行分发提供了稳定、可靠的承载管道，其上行的带宽能力也为 VR 视频业务采集提供了更为便捷的方式。将云计算平台部署到汇聚节点以下的位置，边缘云平台由集中化向分布化扩展，把 VR 视频业务所需的图形渲染、视频拼接、转码等业务下沉，距离用户更近，满足云平台 VR 渲染小于 30ms 的时延要求，高质量地应用于 VR 视频直播上行、VR 视频 CDN 分发等业务场景。

VR 视频业务中，边缘云需部署 CPU、GPU、存储等基础资源，具备渲染、编解码、视频拼接等能力，并可提供第三方应用，该方案具有如下优势：

a. 网随流动：满足业务级别的分流，按需保障业务连续性，可基于切片提供端到端安全可靠的保障能力；

b. 异构加速：高性能支持集成多种类型（NPU、GPU）的加速卡，满足媒体加速、渲染加速、AI 等不同场景需要；

c. 极简运维：一站式用户界面，开放平台第三方应用快速上线，云边协同；

d. 安全可信：全云全栈自主可控，多层防护保证网络安全、数据安全，且数据加密传送。

5.5 VR 视频终端

VR 视频终端包括 Wi-Fi 模块、解码、视频帧处理、播放操作控制和显示处理

等模块，具备解码、姿态感知、运动轨迹预测、实时模型渲染和呈现等能力。目前 VR 视频主流分辨率为 4K，但高质量的 VR 视频分辨率需至少达到 8K 以上，更高可达到 30K，因此终端典型的单眼分辨率需达到 2K 及以上，硬件需支持 8K 解码能力，终端头动感知时延应小于 20ms，同时还需具备高性能的通信能力。针对基于 5G 网络的 VR 视频传输方案，终端侧还需与 5G 网络、边缘云和中心云协同。

6. VR 视频应用的挑战与前景

VR 视频的出现给广播电视带来了新的机遇，本章梳理了 VR 视频节目制作、传输分发和终端消费各环节所面临的问题和挑战，解决这些挑战对完全实现身临其境的 VR 视频体验是非常必要的。

6.1 VR 视频节目制作域

6.1.1 拍摄指导

由于许多拍摄技术限制和取景美学构图等原因，与传统节目制作相比，VR 制作往往更加复杂。在 VR 视频拍摄录制中，没有任何地方可以隐藏设备和制作人员，因而在制作期间进行镜头间视频的拼接是不可缺少的；此外，为了防止用户观看时感觉头晕，VR 视频中镜头的快速移动、位置变化、场景切换等通用拍摄手段必须小心使用。在这些约束下，导演必须找到能够引导观众注意力的替代方法。

6.1.2 映射

映射环节是将 360° 的球体视频投影到二维平面上的过程，有多种映射方法都可将 VR 视频图像投影成适用于当前视频编码器的二维平面格式。其中最常使用的映射模式是 ERP，也有其他各种方法如金字塔映射、立方体映射等。但 VR 显示设备（如 HMD 等）一般不支持多种投影映射模型，因此在前端制作和传输分发过程中就需要考虑到各种终端的不同适配情况。

6.1.3 制作流程和创作工具

对于 VR 专业内容创建，自动化生产工作流程对于及时、高质量的内容交付是必要的。线性叙事 VR 视频内容工作流程包括：素材拍摄、格式转换、拼接映射、音频制作、资产创建、渲染、审看、分发和存档等。此工作流程虽然可能类似于 HDR 视频制作，但处理的数据量通常要大得多。而对于非线性叙述 VR 视频内容制作，即基于一些用户交互的 VR 视频内容，工作流程还包括游戏开发的工具。VR 视频内容制作也受益于这类新的创作工具，如声音设计创作工具，使得在通过 HMD 或 360° 视频播放器观看视频内容的同时，可以平移和可视化声源。

目前，VR 视频内容制作工具被集成到广播制作流程和基础设施中的还不太多，

统一的技术标准对于 VR 视频内容的制作、交换、分发等会带来一定工作量的简化。

6.2 VR 视频传输分发域

6.2.1 高传输码率

VR 视频的传输分发对于现在正在使用的传输分发系统是目前最大的挑战。VR 视频的传输码率远远超过当前高质量广播节目的数据码率。许多消费者设备的网络吞吐量和处理能力都不能满足，通常都需要进行数据缓冲来保证连续一致的沉浸式体验。

为了给用户提供高质量的 VR 视频，业界正在寻找各种方法。这些方法大多基于视角相关视频流的动态切换，即用户正在观看的视角接收高质量视频，而其他视角接收较低质量的视频。为防止在头部旋转期间感知到低质量视频内容，从一个视角到另一个视角的无缝切换是至关重要的。但对于广播电视信道传输 VR/AR 视频时，基于视角的解决方案可能行不通，需要一次性将整个视频进行传输。

6.2.2 VR 分发格式

标准化的分发格式对于向用户提供高质量的沉浸式 VR/AR 内容也是非常重要的。目前，MPEG 已发布了第一版全景媒体应用格式 OMAF 标准 ISO/IEC 23090-2，它不仅定义了 VR/360° 内容的媒体格式，还定义了使用 MMT、MPEG-DASH 传送 OMAF 内容的方法。

6.3 VR 视频终端消费域

VR 视频生产和传输过程中的许多技术问题是终端高质量内容消费需求驱动发展出来的，这些需求超过了对传统电视广播的要求。例如许多专家建议，至少需要每秒 60 帧和双眼 4K 的视频图像才能实现高质量 VR 体验。可接受的动作和声音、运动和光的延迟是实现和预防眩晕的重要技术参数。除此之外，调研发现消费者对于 HMD 的重量、外观设计也有一定的要求。VR 视频在不同终端的呈现效果也不一样，因此亟需统一的标准进行规范。

目前创建的 VR 视频大多持续时间不到 15 分钟，这是为了避免更长时间使用 HMD，尤其是观看高速运动内容时可能发生的眩晕感。尽管已进行了多年的研究，但预防、去除眩晕感仍旧是 VR 消费的难题。众所周知，眩晕感是视觉渲染技术参

数（延迟、帧数、显示视图的闪烁、显示宽度）和人体方面（曝光持续时间、个人灵敏度、运动控制、健康、性别、年龄、情绪、焦虑、姿势稳定性等）因素共同影响的。中国电子协会消费电子分会等单位组织了 VR 头戴式显示设备对青少年眼健康影响的实验研究，结果显示，使用 VR 头戴式显示设备会对青少年眼部指标和主观感受产生影响，建议青少年合理分配使用 VR 头戴式显示设备的时间和频率。

6.4 发展前景展望

5G 具有大带宽、低时延的特性，为 VR 视频的发展提供了强有力的支撑，VR 视频不仅在广播电视领域快速发展，也将在新闻、旅游、游戏、电影、科研、教育、医疗等领域得到广泛应用，VR 技术与 AR 技术融合发展也是未来的发展趋势之一。在未来的规划及发展中，VR 视频应用与广电的融合将为广电业务转型升级注入新兴力量，也将为智慧广电新生态建设开辟全新空间。

参考文献

- [1] Report ITU-R BT.2420-0(04/2018) 《Collection of usage scenarios and current statuses of advanced immersive audio-visual systems》.
- [2] Recommendation ITU-R BT.2123-0 (2019) 《Video parameter values for advanced immersive audio-visual systems for production and international programme exchange in broadcasting》.
- [3] Recommendation ITU-R BT.2133-0 (2019) 《Transport of advanced immersive audio-visual content in IP-based broadcasting systems》.
- [4] ISO 11664-2-2007/CIE S 014-2:2006 色度 第二部分：CIE 标准照度 (Colorimetry - Part 2: CIE standard illuminants).
- [5] GY/T 307-2017 《超高清晰度电视系统节目制作和交换参数值》.
- [6] GY/T 315-2018 《高动态范围电视节目制作和交换图像参数值》.
- [7] ISO/IEC 23090-2 (2019) 《Information technology—Coded representation of immersive media — Part 2: Omnidirectional media format》.
- [8] DVB. (11/2016) 《Executive summary DVB study mission on virtual reality》.
- [9] SMPTE 2016 Annual Technical Conference and Exhibition, P. Routhier, 《Virtually perfect: Factors affecting the quality of a VR experience and the need for a VR content quality standard》.
- [10] 3GPP TS 26.118(09/2019) 《Virtual Reality (VR) profiles for streaming applications》.
- [11] 3GPP TR 26.999 《Virtual Reality (VR) Streaming Interoperability and Characterization》.
- [12] IEEE 1857.9 《Standard for Immersive Visual Content Coding》.
- [13] Brown, M., Lowe, D.G. Automatic Panoramic Image Stitching using Invariant Features. Int J Comput Vision 74, 2007, pp. 59-73.

- [14] Zaragoza, J., Chin, T. J., Brown, M. S., Suter, D. : As-projective-as-possible image stitching with moving dlt. In: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2339–2346.
- [15] C. Lin, S. U. Pankanti, K. N. Ramamurthy and A. Y. Aravkin, "Adaptive as-natural-as-possible image stitching," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 1155–1163.
- [16] Chen, Yu Sheng & Chuang, Yung-Yu. (2016). Natural Image Stitching with the Global Similarity Prior. 9909. 186–201. 10.1007/978-3-319-46454-1_12.
- [17] Gao, Junhong & Yu, Li & Chin, Tat-Jun & Brown, Michael. (2013). Seam-Driven Image Stitching.
- [18] F. Zhang and F. Liu, "Parallax-Tolerant Image Stitching," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2014, pp. 3262–3269.
- [19] Ye, Yan & Alshina, Elena & Boyce, Jill. (2018). JVET-F1003: Algorithm descriptions of projection format conversion and video quality metrics in 360Lib.
- [20] ITU-T H430.1(08/2018) Requirements for immersive live experience (ILE) services.
- [21] ITU-T H430.2(08/2018) Architectural framework for immersive live experience (ILE) services.
- [22] ITU-T H430.3(08/2018) Service scenario of immersive live experience (ILE).
- [23] ITU-T H430.4(11/2019) Service configuration, media transport protocols, signaling information of MMT for Immersive Live Experience (ILE) systems.

- [24] M. Wien, J. M. Boyce, T. Stockhammer and W. Peng, "Standardization Status of Immersive Video Coding," in IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol. 9, no. 1, pp. 5-17, March 2019.
- [25] M.-L. Champel, R. Koenen, G. Lafruit, and M. Budagavi, Draft 1.0 of ISO/IEC 23090-1: Technical Report on Architectures for Immersive Media, document N17685, ISO/IEC JTC1 SC29/WG11 MPEG, 123rd Meeting, Ljubljana, Slovenia, Jul. 2018.
- [26] High Efficiency Video Coding, document ITU-T H.265, 5th ed., 2018.
- [27] B. Bross, J. Chen, and S. Liu, Versatile Video Coding (Draft 4), document JVET-M1001 and N17506, JVET ITUT SG 16 WP3 and ISO/IEC JTC1 SC29/WG11, Marrakech, Morocco, 13th Meeting, Jan. 2019.
- [28] 《VR 视频点播业务试播实施方案》，2018.09，深圳市天威视讯股份有限公司.
- [29] 《大连天途有线电视公司云 VR 产品项目介绍》，2018.03，大连天途有线电视公司.