

引用格式:魏娜,郭晓强.“机顶盒+VR终端”的广播级VR视频业务方案研究与分析[J].中国传媒大学学报(自然科学版),2022,29(05):36-42.
文章编号:1673-4793(2022)05-0036-07

“机顶盒+VR终端”的广播级VR视频业务方案研究与分析

魏娜*,郭晓强

(国家广播电视总局广播电视科学研究院,北京,100866)

摘要:目前,VR/AR技术被视为进入元宇宙的重要接口型技术,VR视频也越来越多的应用于重大活动、体育赛事等场景中,但目前的VR视频类业务尚不能提供给用户高品质的沉浸式视听观看。本文首先对VR视频技术发展的五个维度进行了详细分析,随后研究提出了广播电视机顶盒结合VR终端实现广播级VR视频业务的3种方案,并对各方案优缺点进行了分析比较。此外,本文还研究了广播级VR视频业务在多个重要环节的技术要求,并对代表性VR视频业务实例进行了分析。最后进行了总结梳理,并对广播级VR视频业务进行了展望。

关键词:VR视频;机顶盒;VR终端;广播级;沉浸式

中图分类号:TN94 文献标识码:A

The research and analysis of broadcast-level VR video service scheme based on 'set-top box+VR terminal'

WEI Na, GUO Xiaoqiang

(Radio and Television Research Institute, Beijing100866, China)

Abstract: At present, VR/AR technology is regarded as an important interface technology to enter the Metaverse, and VR video is increasingly applied in major events, sports events and other scenes. However, the current VR video services cannot provide users with high-quality immersive audio-visual viewing. This paper first analyzes the five dimensions of VR video technology development in detail, and then studies and proposes three schemes to achieve broadcast-level VR video service by combining radio and TV set-top box with VR terminal, and analyzes and compares the advantages and disadvantages of each scheme. In addition, this paper also studies the technical requirements of broadcast-level VR video service in many important links, and analyzes the representative VR video service examples. At the end of this paper, a summary is made and a prospect is made for broadcast-grade VR video service.

Keywords: VR video; set-top box; VR terminal; broadcast-level; immersion

1 引言

VR技术是指通过动态环境建模、实时三维图形生成、立体显示观看、实时交互等,生成仿真现实的三

维模拟环境,用户借助交互设备与虚拟场景中的对象相互作用、相互影响并沉浸其中,获得身临其境的感受。随着“元宇宙”、“数字孪生”等概念和技术的进一

基金项目:国家广播电视总局广播电视科学研究院基本科研业务费课题“VR视频终端技术指标研究”(项目编号:JBKY20210270);国家广播电视总局广播电视科学研究院基本科研业务费课题“3DoF VR视频图像质量评价关键技术研究”(项目编号:JBKY20220050)

作者简介(*为通讯作者):魏娜(1988-),女,硕士研究生,工程师,主要从事VR/AR、台内制播、图像质量等领域技术研究。Email:weina@abs.ac.cn;
郭晓强(1977-),男,博士研究生,正高级工程师,主要从事超高清、VR/AR、视音频编解码等技术研究。Email:guoxiaoqiang@abs.ac.cn

步发展,VR/AR等终端设备也被认为是目前接入元宇宙的重要入口,VR视频也越来越多地被应用在重大活动、体育赛事等场景中。

本文首先对VR视频技术发展的五个维度进行了详细分析,为达到终极沉浸式视听观感,VR视频技术在各维度上都需大幅提升。随后本文在研究VR终端渲染显示流程的基础上,设计提出了广播电视机顶盒结合VR终端实现VR视频业务的3种方案,并对各方方案优缺点进行了分析比较。此外,本文还研究了广播级VR视频业务在多个重要环节的技术要求,并对现有代表性广播电视VR视频业务实例进行分析。最后,本文对VR视频技术现状及问题进行了总结梳理,并对广播级VR视频业务进行了展望。

2 VR视频技术发展的五个维度

VR视频是指能够在纯虚拟空间中,为用户提供第一人视角身临其境的高清晰度观看、沉浸式听音、自然交互等超仿真视听体验。基于VR视频的概念和终极呈现目标,VR视频技术将面向高画质、立体感、沉浸声、自由度、交互度等5个维度全向发展,如图1~2所示。



图1 VR视频技术发展的五个维度

VR视频技术发展的终极目标就是能够提供视网膜级3D立体观看、真实沉浸声、六自由度运动、自然体验式交互的视听触嗅体验,并尽可能的降低观看晕动症、眼疲劳等不适感。目前VR视频技术还处于初中级发展阶段,仅可向用户提供一定程度的沉浸式视听感受。

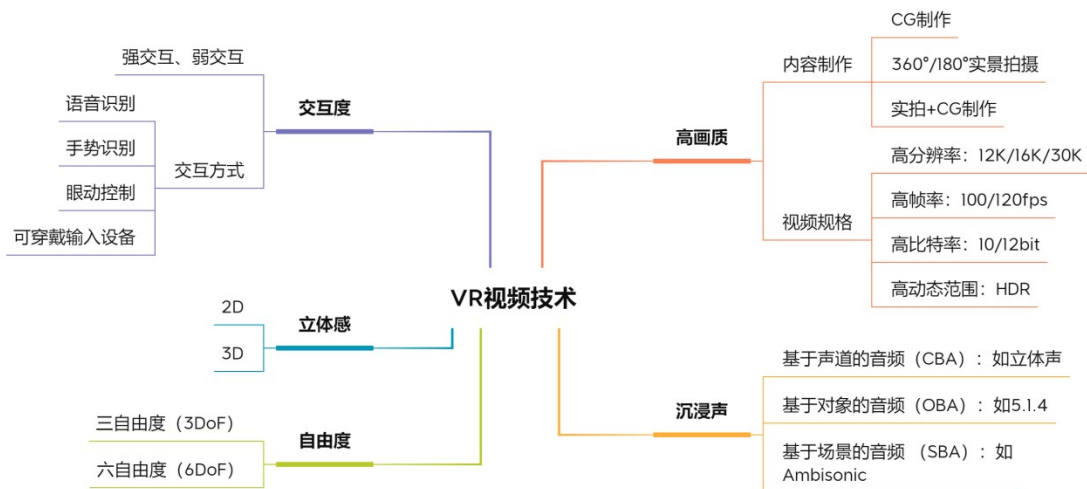


图2 VR视频技术五个维度的发展方向

(1)高画质

VR视频从内容制作的角度可分为CG制作、360°/180°实景拍摄、实际拍摄+CG相结合等类型。目前VR视频业务主要面向个人和行业两类用户,包括VR直播、VR游戏、VR教育、VR医疗等。VR视频与不同领域的特点和需求相结合,制作相应的视频内容。VR视频内容技术规格将在高分辨率(12K/16K/

30K)、高帧率(100/120fps)、高比特率(10/12bit)、高动态范围(HDR)等方面进行逐步提升。

(2)立体感

人眼双目观看的真实世界是3D立体式的,为了在VR视频中能够体验到更加真实的沉浸观看感受,未来成熟的VR视频将呈现舒适的3D观看效果。现阶段观看3D VR视频时会进一步增加人体的眩晕感,

但随着VR终端设备性能的不断迭代,VR视频内容技术规格的不断提升,3D VR将会逐步突破瓶颈。

(3) 沉浸声

真实世界的声音来自四面八方,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。目前VR视频中大都选用双声道/立体声作为音频的呈现方式。为再现真实世界的听音效果,VR视频未来将更多使用基于对象(OBA, Object-based Audio)和基于场景的音频(SBA, Scene-based Audio)技术,如5.1.4、Ambisonic等,进行沉浸音的还原和重放。

(4) 自由度

基于用户观看自由度,VR视频可分为三自由度(3DoF)VR视频和六自由度(6DoF)VR视频。典型的3DoF VR应用场景是坐在椅子上的用户(固定位置)通过HMD观看360°全景视频内容。目前VR视频正面向6DoF发展,用户可以在物理空间内任何位置、任何方向自由的观看节目素材。

(5) 交互度

VR视频的交互依赖于传感器技术的不断发展,用户可以通过触点反馈装置模拟触觉体验,通过语音系统与虚拟现实世界进行听、说交互,并通过手势识别仪器、眼动控制器、可穿戴输入设备等与VR系统进行实时交互。目前VR视频的交互通常分为强交互(如VR游戏)和弱交互(如VR视频直播)两类。

3 VR视频终端渲染及显示

VR视频终端一般包括接收、解码、渲染、显示等模块,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。

VR视频软件在终端渲染显示每一帧画面时,通过追踪用户的头部姿势,保证用户观看的画面与用户位置和头部姿态一致,实时完成追踪、渲染、显示等流程。用户头部发生运动后新的一帧图像渲染显示在VR HMD的渲染流程如图3所示。

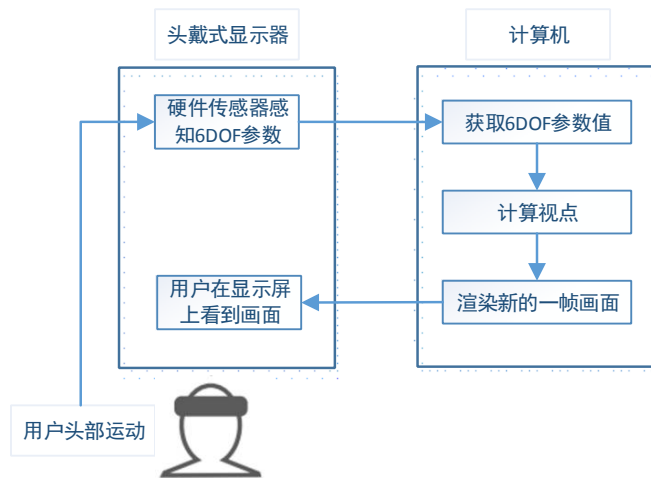


图3 VR视频帧图像渲染流程图

(1) 渲染

渲染处理是将三维虚拟空间物体投影到平面,形成双目视觉的处理过程。渲染为用户提供高画质、流畅低时延的VR视频体验。良好用户体验的VR视频需要极强的终端渲染能力,因此减少计算开销、降低渲染时延成为渲染技术的发展趋势,如云渲染、异步时间扭曲渲染、多分辨率着色、Multi-View渲染等。

(2) 显示

人眼的可视角度范围一般在90°至110°之间,人眼视网膜每角度内可分辨的极限像素数大约为60个。通过计算可知,在终端获得视网膜级的观看清晰度体

验,拍摄端需保证30K×15K的超高分辨率。因此,VR终端设备的显示分辨率、硬解码处理等能力均需大幅度提升。

4 广播电视机顶盒+VR终端的方案设计

VR视频流通过互联网传输至VR终端,接收处理后供用户观看的整体技术方案已基本成熟,并可通过云VR+5G的方式,进一步降低终端计算复杂度及传输时延。Meta、优酷、爱奇艺等互联网公司的VR视频业务均基于此类技术方案,但互联网传输方案中,大多业务受限于传输带宽、终端处理能力等因素,最终

呈现出的VR视频内容质量较低,沉浸感不佳,用户体验较差。

近年来,通过广播电视网络传输VR视频类业务也逐步增多,面向广播电视网络中提供高质量VR视频业务的实际需求,本文研究提出了基于广播电视机顶盒+不同类型VR终端在用户侧接收处理的技术方案:有线电视/IPTV网络传输的VR视频流,可通过机顶盒接收、解码处理后推送至电视机或外接式VR终端观看,也可

以通过机顶盒接收、转发至一体式VR终端解码渲染和观看。此外,本文还对广播级VR视频业务的端到端各环节技术要求进行了分析,详述如下。

(1)方案1:广播电视网+机顶盒+电视机

有线电视/IPTV网络传输压缩的VR视频流至机顶盒,机顶盒解码处理后,将VR视频通过HDMI等方式输出至电视机等大屏,用户通过机顶盒遥控器操作观看VR视频,技术方案如图4所示。

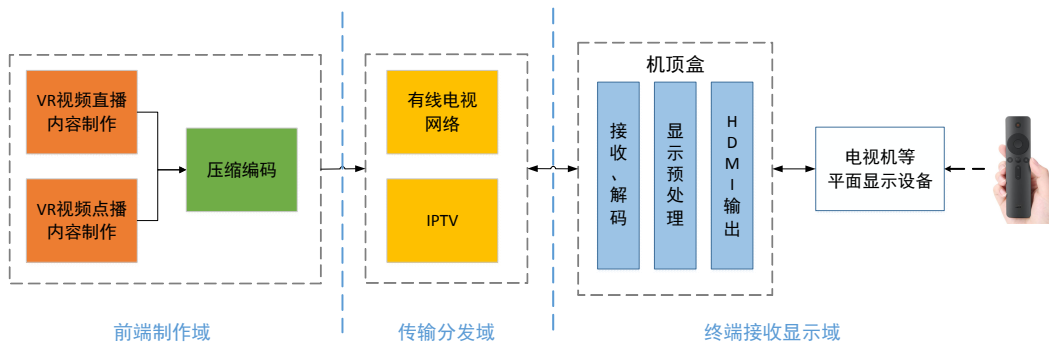


图4 方案1:广播电视网+机顶盒+电视机

本方案中,用户侧不需要VR终端设备,仅在电视机等平面终端设备上观看映射处理后的VR视频。用户可自行选择全景模式和VR模式观看。当选择VR模式时,用户可通过机顶盒遥控器的“上/下/左/右”键来切换视角观看VR视频、按“确认”键回到默认主视角等。

基于本方案,用户在电视机上观看到的VR视频图像均存在一定程度的拉伸和变形,且不能体验360°沉浸感,不能算作真正的“VR视频”。但本方案的优点是终端接收处理复杂度较低,用户接入便捷度高。本方案仅需修改4K/8K机顶盒的底层代码,并开发相应的VR视频APP/SDK即可。

(2)方案2:广播电视网+机顶盒+外接式VR终端

外接式VR终端通常包括解码处理、显示、定位等多个独立模块。以HTC VIVE Pro2设备为例,解码处理在主机完成,显示模块为头戴式HMD,头戴设备和计算主机间通过“数据线+串流盒”进行连接。手柄和头戴设备的定位、交互渲染通过两个光学无线定位器完成。

方案2中,有线电视/IPTV网络传输VR视频流至机顶盒,机顶盒解码处理后,将VR视频通过HDMI等方式通过串流盒输出至外接式VR终端显示,外接式VR终端的空间位置和姿态等数据也通过串流盒回传至机顶盒,用于计算新的头部姿态对应的实时渲染显示画面,技术方案如图5所示。

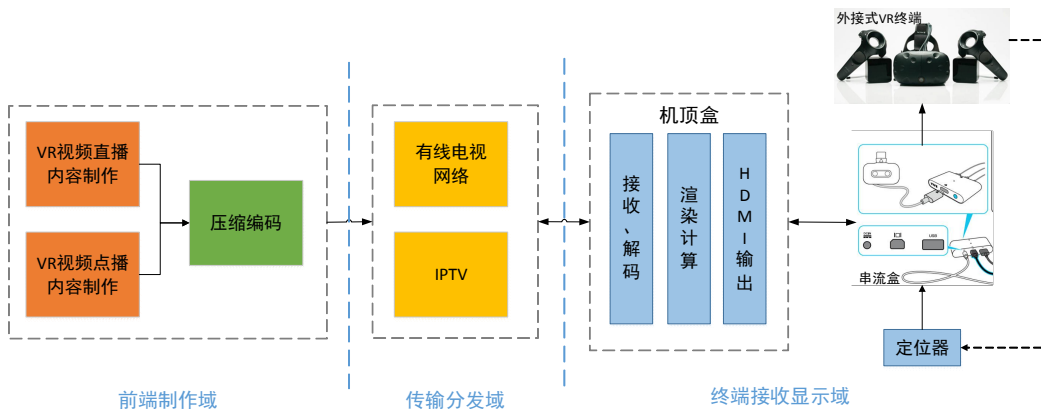


图5 方案2:广播电视网+机顶盒+外接式VR终端

本方案中,机顶盒充当了外接式VR终端的解码处理主机,完成解码处理的工作。同时观看物理空间中(如家庭客厅)的定位器需和串联盒一并连接至机顶盒上,收集机顶盒接收头显和手柄的位置姿态等数据并进行计算和处理。本方案中除修改4K/8K机顶盒的底层代码、开发相应的VR视频APP/SDK外,还需设计改造机顶盒的硬件处理模块及接口,连接定位器、串联盒等设备,方案相对复杂度较高。

(3)方案3:广播电视网+机顶盒+一体式VR终端

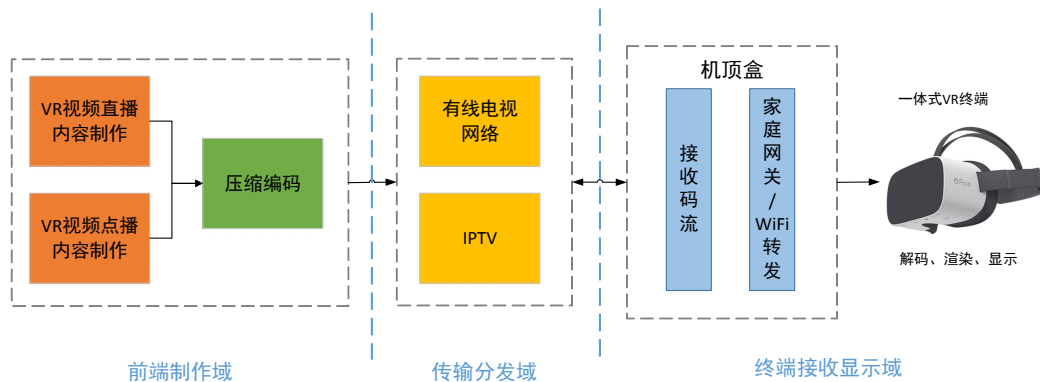


图6 方案3:广播电视网+机顶盒+一体式VR终端

本方案中,机顶盒通过WiFi等协议只完成码流的转发工作,VR视频的解码、渲染、显示和交互均由一体式VR终端完成。同时,还可以通过家庭网关/局域网将一体式VR终端内观看的VR视频画面投屏到电视机上,多人共同分享观看内容。本方案中机顶盒需设计实现局域网内码流的无线转发功能。

(4)各方案比较

多地广播电视网中都首选方案1开展了相关的VR体验业务。方案1需在终端开发相应的VR视频解码渲染软件,用户可在已有机顶盒和电视机上体验VR视频业务,用户侧系统基本无改造,因此最适宜于VR视频最初的推广阶段。但方案1不能提供真正的360°沉浸式感受,因此也会对用户观看VR视频产生“错误”的引导。

方案2所需的终端用户侧系统改造最为复杂,对机顶盒的软硬件能力要求也较高,但可为用户提供较好体验的VR视频业务。但由于方案2需要单独定制VR视频机顶盒,在用户侧观看空间内需安装定位器等设备,且外接式VR终端设备近年来发展趋势变慢,方案2的实用性相对较低。

方案3考虑到VR终端设备的发展趋势,以及尽量简化用户侧的系统改造,使用机顶盒转发码流至一

近年来随着一体式VR终端设备的不断迭代更新,由于此类终端具备更高分辨率/大视场角、精准的跟踪和识别、空间音频呈现、轻薄低功耗、AR/VR一体化等优势,也越来越受到用户的喜爱,成为VR终端产品最重要的发展方向。

在方案3中,有线电视/IPTV网络传输VR视频流至机顶盒,机顶盒接收后,将VR视频通过家庭网关/局域网转发至一体式VR终端,VR终端进行解码和显示,技术方案如图6所示。

一体式VR终端的接收方式,并结合广播电视与网络视听优质内容,可以便捷的为用户提供弱交互模式下的高品质VR视频业务。但目前在实际应用中案例相对较少。

5 广播级VR视频业务重要环节技术要求分析

5.1 前端制作域

(1)拍摄

VR视频的拍摄一般分为单机位拍摄和多机位拍摄两类,与4K/8K超高清拍摄相比,由于受限于拍摄方式和取景美学构图等原因,VR视频拍摄更为复杂。VR视频拍摄在构图、拍摄高度/距离、布光、视频缝合、一致性调整、视觉舒适度等方面均需特别注意。如在360°全景视频拍摄录制中没有任何地方可以隐藏,因而在拍摄前需要设计好工作人员和道具的隐藏方法。此外,为防止用户观看时感觉头晕,360°全景视频中镜头要尽量避免快速移动、频繁场景切换等画面镜头。

在视频参数方面,为达到广播级优质的用户观看体验,VR视频在拍摄时需至少达到8K分辨率、50fps以上,VR视频节目拍摄制作视音频基本参数见表1。

表1 VR视频节目拍摄制作视音频基本参数值

序号	参数	数值
1	球面360°图像到矩形二维图像的映射方法	推荐ERP、CMP、EAC
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、60、50
5	扫描模式	逐行
6	色域	SDR(GY/T 307)、HDR(GY/T 315)
7	量化电平(bit)	10、12
8	音频取样率(kHz)	32、44.1、48
9	音频还音方式	立体声、环绕声、沉浸声

(2) 编辑制作

VR视频节目一般分为直播类节目和点播类节目。VR视频直播节目实时制作流程主要包括视频缝合、实时调色、在线包装/字幕制作、实时监看、视音频编码、分发等流程节点。VR视频点播类节目后期制作流程主要包括视频缝合、缝合后素材导入、视频剪辑、视频调色、字幕/特效制作、音频制作、审核修改、合成输出等流程节点。

与传统平面视频编辑方法不同,VR视频节目编辑制作中需特别注意镜头缝合、字幕/特效添加、沉浸式音频制作、交互流程制作等环节。

(3) 映射

映射环节是将360°的球体视频投影在二维平面上的算法过程,有多种映射方法都可将360°视频图像投影成适用于当前视频编码器的二维平面格式。其中最常使用的是ERP和CMP映射,也有金字塔映射、分区域球体映射等其他多种映射模型。复杂度越高的映射模型可以一定程度上节省编码码率和传输带宽,但对VR终端的逆映射处理能力要求也相对较高,且设备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,选择适宜的映射模型。

(4) 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。针对VR视频,MPEG等标准组织也正在研究VVC(Versatile Video Coding)、点云等新型压缩编码算法。此外,基于VR视频观看时真正呈现在人眼前的只是360°全部视频的一小块区域的特性,也可以使用基于FOV(Field of View)可视角的编码传输方法,极大的降低视频码率。

视频码率是影响VR视频在终端接收观看清晰度的重要指标之一,为达到广播级优质的用户观看体验,现阶段8K/50P/10bit的视频采用H.265或AVS2编码时,码率需达到80~100Mbps。

5.2 传输分发域

传输分发中可参考采用MPEG-I提出的OMAF(Omnidirectional Media Format)标准,即MPEG-I(ISO/IEC 23090):Part2。MPEG OMAF适用于3DoF VR,规定了球面坐标系统、映射和矩形区域调整打包方法、全景视频的存储、ISOBMFF元数据、通过DASH/MMT封装和传输全景视频流等内容。

传输带宽很大程度上决定了VR视频业务的质量。现阶段8K VR视频需要80~100Mbps带宽,强交互模式下MTP(Motion To Photons)头动感知时延应小于10ms,弱交互模式下MTP头动感知时延应小于20ms。VR视频对传输带宽码率的具体要求见表2。

表2 VR视频对传输带宽的要求

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
典型带宽(Mbps)	80~100	800~1000
MTP头动感知时延	强交互:小于10ms 弱交互:小于20ms	强交互:小于5ms 弱交互:小于10ms

5.3 终端接收显示域

(1) 视音频解码

由于VR视频具有高分辨率、高帧率、高码率、沉浸式等特点,VR终端需具备相应的视音频解码能力。

在视音频解码方面,VR终端、机顶盒均需支持HEVC、AVS2等主流视频解码方式,能够对符合表1中各参数的VR视频进行正常解码。未来还需能具备环绕声、沉浸声的音频解码还音能力。

(2) 渲染交互

VR终端现阶段大多支持头/手3DoF自由度,支持ERP、CMP等基础逆映射格式,MTP头动感知时延均不超过20ms。为达到更真实的视听、交互体验,并逐步降低观看晕动症等不适感,未来终端设备需面向6DoF自由度、更低的MTP时延发展演进。

(3) 显示

如2.2节所述,VR终端设备的显示能力限制了VR视频的观看质量,严重影响用户的沉浸观感体验。从观看角度出发,为达到广播级的VR视频观看体验,VR终端设备的视场角不能小于90°,单眼显示分辨率不低于HD级别,且为避免观看卡顿感显示刷新率不小于60Hz。

6 现有广播电视VR视频业务实例分析

截至目前,依托广播电视网络开展的VR视频类业务实例还相对较少,且受限源端内容、传输带宽与终端设备能力,目前广播电视网内还未有8K及以上分辨率的VR视频类业务,因此用户在终端观看体验的清晰度、流畅度、沉浸感均有待进一步提高。

(1) 广播电视网+机顶盒+电视机

新疆天山云VR业务依托IPTV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。

该VR视频业务采用了将全景视频平铺形成2D视频,通过传统点播网络推流至用户机顶盒的方式。在视频中加入手机端H5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。VR视频内容分类包含:影视、综艺、美女、体育、风景、院线等。

(2) 广播电视网+机顶盒+外接式VR终端

大连天途云VR平台依托大连天途优质的视频网络资源,与全球领先的云VR技术平台相结合,完成了对云端的方式提供VR业务模式的基础研究。

传输环节中使用有线和IPTV网络两种方式。在有线网中,云VR平台将VR应用的实时视频通过IPQAM推送到用户机顶盒,并通过EOC网络将操作指令、头部

运动信息等数据回传到云端服务器;IPTV网络中,云VR平台将VR应用的实时视频通过IP网推送到用户机顶盒,并通过IP网络将操作指令、头部运动信息等数据回传到云端服务器。终端则定制开发了VR视频专用机顶盒,并匹配HTC VIVE外接式VR眼镜显示观看。

(3) 2022北京冬奥8K VR沉浸式观赛

2022北京冬奥会上,央视频推出了一款基于互联网的8K VR应用,用户通过VR头戴式显示设备可以身临其境的在虚拟空间中,点播观赏冬奥会开幕式和闭幕式,以及冰球、单板滑雪、花样滑冰、自由式滑雪、冰壶、短道速滑六项热门运动。

冬奥VR赛事的8K直播是分别由五台VR 180°摄像机和一台VR 360°摄像机现场采集,以每秒60帧的速度录制。在传输中采用基于FOV/Tile分块技术,让大规模传输超高分辨率和高帧率的VR视频成为可能。此外,央视频VR应用还融合了沉浸式实时3D场景渲染、用户输入与反馈交互等先进技术,并配合高性能VR终端设备,为用户呈现出无与伦比的沉浸式观赛体验。

7 结语

VR视频由于能够提供给用户身临其境的视听体验,越来越多的应用在各个领域,但高品质的VR视频需要高带宽、低延时等保障。目前VR视频类业务大多面临着VR视频源质量低、网络传输带宽不足、终端处理能力不够、交互度低等问题与挑战,因此仅能向用户提供初级的体验型业务,不能达到真正高保真的沉浸式视听体验。VR视频技术也正处在随着元宇宙技术体系的快速推进而不断发展的状态中。

未来,依托广播电视与网络视听的高品质内容优势,利用有线电视网络和IPTV的网络资源,广播级VR视频业务将能够不仅着眼于优化VR视频技术提升沉浸式体验,也将更加注重打造互动性与个性化,VR视频也将逐步成为用户真正喜爱和接受的新型视听节目。

参考文献(References):

- [1] ISO/IEC 23090-2, 信息技术沉浸式媒体编码表示第2部分:全景媒体格式(Information Technology-Coded representation of immersive media-Part 2: Omnidirectional media format) [S]. 2021.
- [2] 国家广播电视总局. GY 356-2021, VR视频系统节目制作和交换用视频参数值[S]. 2021.
- [3] 国家广播电视总局. 关于印发5G高新视频系列技术白皮书的通知[EB/OL]. [2020-08-25]. http://www.nrta.gov.cn/art/2020/8/25/art_113_52661.html.