



全景媒体的系统架构研究综述

罗莹¹, 宋利¹, 解蓉¹, 罗传飞²

(1. 上海交通大学图像通信与网络工程研究所, 上海 200240;

2. 中国电信股份有限公司上海研究院, 上海 200122)

摘要: 随着虚拟现实技术的不断发展, 全景媒体系统的标准制定与完善逐渐显示出越来越重要的作用。为了规范虚拟现实系统, 研究全景媒体的系统架构具有重要的价值。首先回顾了目前虚拟现实技术的发展以及在实际应用中存在的问题。其次, 介绍了全景多媒体的应用框架的结构和发展趋势, 针对国际标准组织 MPEG OMAF 架构以及 Facebook 和 Google 的全景媒体应用架构进行了简要的介绍。通过对几种系统架构的对比分析, 对全景媒体应用的发展做出了总结和展望。

关键词: 虚拟现实; 全景媒体; OMAF; 应用架构; 标准组织

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2018007

Review of system architecture of omnidirectional media

LUO Ying¹, SONG Li¹, XIE Rong¹, LUO Chuanfei²

1. Image Communication and Network Engineering Institute, Shanghai Jiao Tong University, Shanghai 200240, China

2. Shanghai Research Institute of China Telecom Co., Ltd., Shanghai 200122, China

Abstract: With the development of virtual reality technology, the formulation and perfection of omnidirectional media system standard plays a more important role. In order to standardize the virtual reality system, it is of great value to study the system architecture of omnidirectional media. Firstly, the current development of virtual reality technology and the problems in practical application were reviewed. Secondly, the structure and development trend of the application framework of omnidirectional media were introduced. Then the application architectures of MPEG OMAF, Facebook and Google were introduced separately. Through the comparative analysis of several system architecture schemes, the development of omnidirectional multimedia was summarized and prospected.

Key words: virtual reality, omnidirectional media, OMAF, application architecture, standard organization

1 引言

虚拟现实 (virtual reality, VR) 技术是一种通

过计算机仿真来创建和体验虚拟世界的技术, 其中, 交互式的三维动态视景与用户实体行为的融合大大丰富了用户的观看体验。目前, VR 已成为一

收稿日期: 2017-07-05; 修回日期: 2017-09-27

基金项目: 国家自然科学基金资助项目 (No. 61671296, No. 61521062)

Foundation Items: The National Natural Science Foundation of China (No. 61671296, No. 61521062)

项热门技术,它通过展现 360°的视频为观看者提供了沉浸式的“亲身体验”和“现实生活”^[1]。用户可以交互性地随时切换观看的视角,并且动态地查看他们所期望看到的部分场景。VR 在很多领域都有它独特的应用魅力,如 VR 模拟、VR 游戏、VR 视频直播等。

从用户体验的角度看,VR 技术的最独特之处在于全景视频,也称 360°全景视频或沉浸式视频。全景媒体是指能够根据用户的观看视角进行渲染的图像、视频及其相关联的音频。在拍摄端,全景视频一般由指向不同方向的多个照相机拍摄并拼接而成。由于人的视野有限,因此无法在特定视角看全整体画面,而是通常将注意力放在特定的感兴趣的区域。在渲染端,全景视频播放已经使用许多显示设备实现。但是,VR 服务的核心问题在于如何将全景视频从相机拍摄端向最终的显示端进行传输和存储。全景媒体的技术架构主要由视频拼接与映射、视频编解码、存储与传输等技术构成。目前,已有多家公司提出了视频拼接算法,关于映射方法也有多种模型方案。此外,一些组织和标准正在制定针对全景媒体的编解码和传输的优化算法。同时,全景媒体技术面临很多巨大的挑战。首先,全景视频分辨率是一个技术瓶颈,相机组拼接带来的不同步、变形等因素将严重降低视频质量;其次,全景媒体庞大的数据传输与计算给传输带宽和终端的解码能力提出了巨大的挑战;此外,端到端的时延也是一个影响用户体验的关键参数。

目前,市场上出现的虚拟现实产品标准不一,需要新的行业标准的约束。因此,为了将 VR 技术扩展到更广泛的市场,需要定义一种通用的应用架构标准,可以在不同的 VR 设备之间进行全景视频的存储、管理、交换、编辑和呈现^[2]。

2 全景媒体应用的发展与演进

OMAF (omnidirectional media application for-

mat, 全景媒体的应用格式)最先由 MPEG (moving picture experts group, 动态图像专家组)组织在 2015 年 10 月的 113 届 MPEG 会议上提出^[2],它提出的重要意义在于它为 VR 系统的输入输出接口设定了标准,便于扩展到科学研究和商业领域。此后,很多公司如高通、三星、诺基亚等以及国内外各大高校纷纷参与制定标准的队伍中。与传统平面媒体相比,全景视频在视频获取到播放的过程中,容易形成“碎片化”^[3]。全景媒体的“碎片化”,是由于在视频内容上的空间分块以及端到端的应用设备的性能不统一性造成的,而 MPEG 组织建立 OMAF 标准,正是为了避免和改善全景媒体在内容和应用设备上的碎片化。首先,OMAF 框架可用于将 360°视频与二维图像帧相互转换的映射和渲染;其次,在基本文件格式的基础上,扩充和丰富了 VR 视频存储功能和相关信令的定义;此外,在框架上还增加了动态自适应流的封装和传输;同时它针对全景媒体流提出了更高要求的压缩编码性能。

MPEG 组织在 2015 年底就初步形成了较为粗略的应用框图,涵盖了图像拼接和映射、视频编解码以及视频在球面的渲染模块。2016 年 6 月,Byeongdoo 等人^[3]在各种 MPEG 会议的提案基础上,对 OMAF 可支持的映射方式进行了比较和归纳。在编码方面,Ridge 等人^[4]提出了下一代 VR 编码的趋势,对编解码器的实际应用提出了挑战性的需求。在存储方面,Wang 等人^[5]提出了一个支持存储全景媒体内容的文件格式的标准,定义了多个 VR 相关的 BOX 类型。在传输方面,Franck 等人^[6]提出了一个新的 DASH (dynamic adaptive streaming over HTTP, 基于 HTTP 的动态自适应流)描述符来帮助 DASH 客户端利用和管理 VR 内容。图 1 所示为 OMAF 标准草案的一个发展阶段路线,从目前的研究趋势可以看出,从 2017 年 1 月开始就有成规范体系的关于 OMAF 框架的 MPEG 会议输出文档。

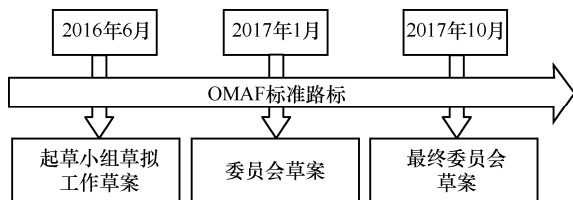


图1 OMAF 标准草案的发展阶段路线

图2所示为关于MPEG OMAF的未来发展路线。从图2中可以看到,在2017年底,三自由度的全景VR系统架构的标准制定完毕,到2020年底,六自由度全景VR系统也将会发布,届时对编解码、传输、文件格式等有全新的标准支持。

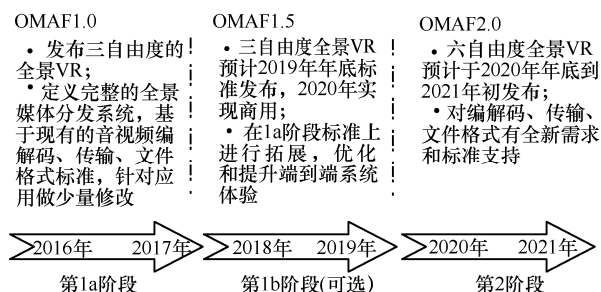


图2 MPEG 组织的 OMAF 标准未来发展路线

关于全景媒体应用的架构,AVS(audio video coding standard,数字音视频编解码技术标准)工作组在2015年下半年也启动了VR全景视频应用工作计划,其任务和目标着重围绕视频编码和传输,定义全景视频紧凑表示方法和编码工具以及系统传输标准,提升全景视频压缩效率^[7]。在第一阶段(2015年年底—2017年3月),AVS组织致力于研究全景视频编码与现有平面视频编码标准的兼容性;在第二阶段(2017年3月—2018年3月),AVS组织将重点放在定义新的全景视频编码工具上;在第三阶段(2017年3月—2020年3月),AVS组织将实现六自由度全景视频的编码。目前,AVS组织在第一阶段定义了10种不同的高效映射模型,面向不同的应用场景,且将主流平面视频编码标准应用到全景视频编码中来。

IEEE虚拟现实与增强现实标准工作组正在制定八项标准(IEEE P.2048),其中涉及全景多媒

体架构的包括沉浸式视频分类和质量标准及沉浸式视频文件和流格式这两个标准。截至2017年4月,全球共有接近200个企业和机构的专家参与该标准的制定工作,成为VR标准化的主要推动力量。

3 全景媒体的应用架构

全景媒体架构对于编解码、封装、传输等提出了更高的性能要求。同时在应用需求上,还需要映射和渲染模块的支持。由于全景媒体的交互性特点,观看者视角也需考虑进入整体架构中。这些性能与应用上的需求构成了全景媒体架构的关键元素,基于这些思想,各大组织和企业正在研究和发展中,形成了逐渐完善的、更为细化的架构。

3.1 MPEG OMAF 下的全景应用框架

为了支持全景媒体的应用,基于OMAF的全景媒体的系统框架应运而生,如图3所示^[8]。图3中实线表示音视频的数据流向,虚线表示OMAF和用户视角的信令流向。在客户端的模块均受视角信令的控制,体现出OMAF框架的用户交互性;另一方面,全景媒体的映射等信息也通过OMAF信令在封装和渲染模块之间传递。对于全景媒体应用的各个模块,MPEG组织经过不断地研究和讨论,提出了丰富的解决思路和算法。

3.1.1 全景媒体的获取

对于全景视频的采集,理论上可以通过7维全光函数来实现。而全光函数所要求的信息量过大,采集、传输和显示等技术问题短期难以获得突破。全景视频是全光函数的近似,它将7维表达简化到4维。早期的全景成像系统使用的是兼有反射折射作用的摄像机^[9],但由于其结构特点,在相机顶部会存在盲区,无法捕获高质量高分辨率的全景视频。目前,较为常用的全景媒体获取的方法是使用包含具有重叠视野的多个鱼眼相机组成的系统。

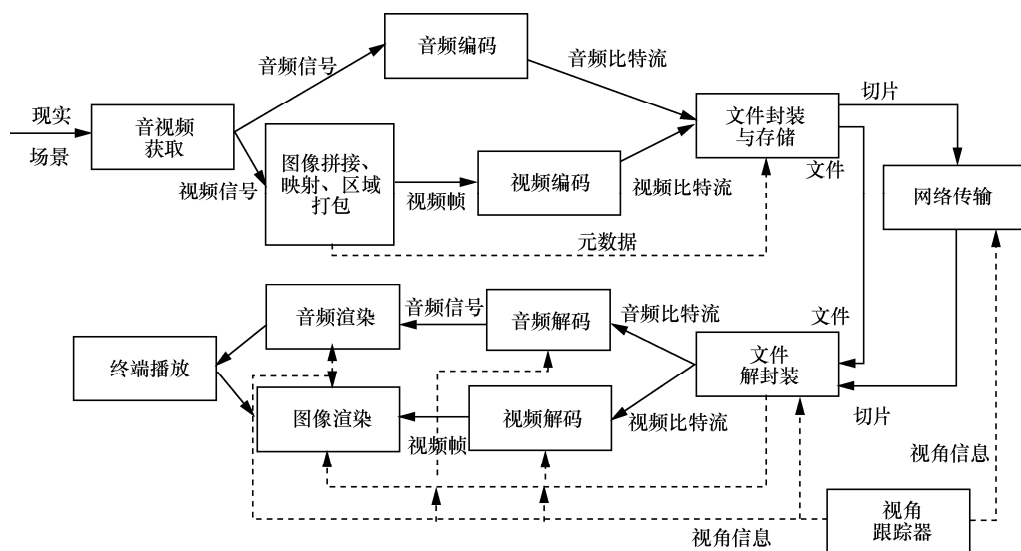


图3 基于 OMAF 的全景媒体的系统框架

在 MPEG 第 115 次会议上，高通和 LG 公司提出鱼眼相机视频相关的文件格式语法和语义^[10]，在 MPEG 的 OMAF 标准中，鱼眼相机的两个有关拼接和渲染的参数——光学变形校正和镜头阴影补偿，被纳入了 OMAF 的信令中，以提高图像渲染的质量。

3.1.2 映射格式

自 2015 年开始，对映射方式的讨论和研究一直很有关注度，先后有十几种不同的映射结构被提出。图 4 所示为经纬图模型映射变换的原理。

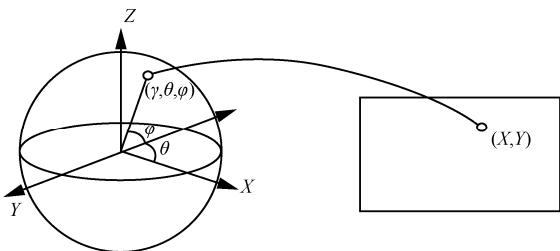


图4 经纬图模型映射变换的原理

由于映射方式的多样性，在 MPEG 第 116 次会议上提出，映射格式需要通过相应测试标准^[11]，才可以被 OMAF 框架最终采用。在 2017 年 4 月的第 118 次 MPEG 会议上，公布了 8 种常用的映射格式，如图 5 所示。这些映射格式可以分为两种：视角依赖的映射格式和视角不依赖的映射格式^[12]。

视角依赖的映射格式	视角不依赖的映射格式
截断方形金字塔结构 (TSP)	立方体结构 (CMP)
基于视角的金字塔结构 (VPP)	分片球体结构 (SSP)
	三角正多面体结构 (TPP)
	调整型立方体结构 (ACMP)
	赤道圆柱结构 (ECP)
	旋转球体结构 (RSP)

图5 两种分类的映射格式

视角依赖的映射结构能获取当前用户视角，并且将视角区域以高分辨率投影到二维平面，其余背景区域则以较低分辨率进行映射变换，这样能够在同等带宽的情况下，提高视角区域的观看质量。但由于视角依赖的映射结构需要从网络上交互式地获取视点内的数据，可能由于网络时延导致视野内数据更新不及时，从而导致用户观看延时。视角不依赖的映射结构不受视角信息的控制，其映射变换结构较为简单，无需获知视角范围的数据。

3.1.3 编解码方案

目前，传统视频编码技术对全景视频仍然有效，新一代视频编码标准 (H.266) 预期能适度减少一半视频码率，可部分缓解全景视频传输的带宽压力。然而，不同于传统视频，全景视频有其



画面的独特性,针对这些特点,一些创新的编码优化思想应运而生。Budagavi^[13]在2015年提出在编码之前加入区域自适应的平滑模块,解决经纬图模型映射导致两级区域过采样问题,节省大量码率。同年,Jin等人^[14]提出了一种扭曲运动补偿方法来解决鱼眼镜头拍摄造成的运动变形。

在2017年4月,MPEG组织提出了适用于全景视频的几种基于视角的编解码方案^[8],如子图像多流编解码法、感兴趣区域增强层法、同分辨率HEVC(high efficiency video coding, 高效视频编码)分块编码法。同分辨率HEVC分块编码法是基于运动约束分块集(MCTS)的编码方法,是将HEVC流按照相同分辨率进行不同质量(假设红色为高质量,黑色为低质量)和比特率的编码,并在接收端根据视角信息解码产生混合质量的图像,如图6所示。

基于视角的全景视频分块编解码能够在固定

带宽情况下,根据观看质量合理地分配分辨率和码率,有效地契合了全景视频的特点,提升了终端用户的观看体验^[15]。

3.1.4 传输机制

在传输中,全景视频的超大分辨率对于带宽和实时性的要求提出了高难度的挑战。在OMAF标准中,提出了两套传输方案:DASH方案和MMT(MPEG media transport, MPEG媒体传输)方案。在全景视频中,根据用户视角进行动态切换主视点码流,则能去除“视角”冗余,减少带宽压力^[16]。

应用于OMAF中的DASH方案传承其基本思想,它通过牺牲存储空间来提高带宽利用率^[17]。在这一方案中,在DASH服务器上,每个视角都存储多份不同码率的视频流,同一时刻根据客户端的视角信息来传输较高码率的主视角切片流和较低码率的其他视角切片流,是码率和视角自适应的动态流传输技术,它的技术框架如图7所示^[8]。

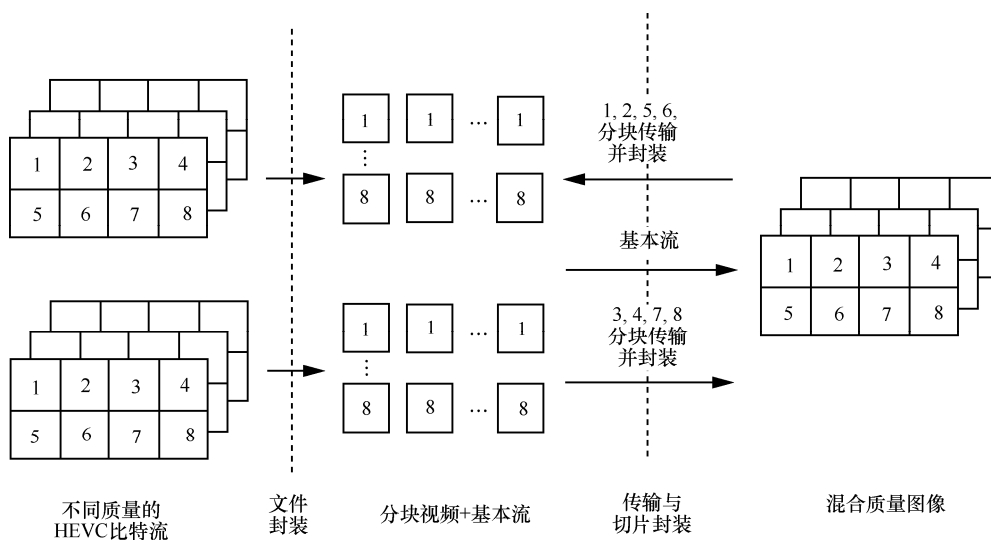


图6 相同分辨率HEVC序列法流程

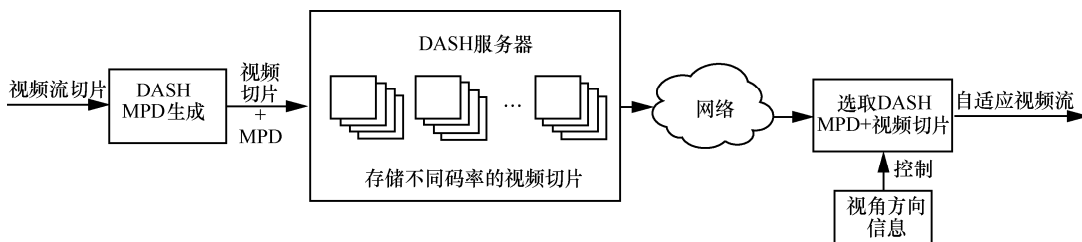


图7 OMAF中的DASH自适应动态流传输框架

基于分块以及视角切换等思想,传输方案的设计和编解码方案一脉相承^[18]。例如图6所示的HEVC运动约束分块集(MCTS)法,在编码端将全景图像划分为多个分块,且编码为不同质量的码流,根据用户视角信息在网络传输中动态切换不同分辨率和码流的媒体流,并在解码端组合成高质量主视角和低质量背景的混合图像^[19]。除此之外,在OMAF中,还提出了使用SRD(spatial relationship descriptor,空间关系描述符)来进行基于用户视角的流式传输^[6]。

另一种传输方案MMT也可作为OMAF应用架构传输模块的候选。MMT和DASH同是MPEG组织标准下的传输协议,二者除了传输架构不同之外,在OMAF架构下,MMT与DASH方案相同,在全景视频的传输中,需要根据当前视角方向,传递全景视频的主视角流,可以根据客户端指定当前视口,也可由发送端的服务器来选择。在传输系统设计中,需要兼并权衡存储、带宽、时延等各因素最大化用户体验和空间、带宽利用率。目前,已经有学者通过优化空间分块的动态流传输方式来提升全景媒体应用中的用户主观质量感受^[20]。

3.1.5 存储格式

为了在基本文件格式中支持OMAF作为媒体存储和封装格式的应用,目前主流的实现方案是在基本文件的基础上增加多个视频track,并在track层次上,添加更多VR信息来支持OMAF这一格式。

为了在track层次来表达VR视频的信息,Wang等人^[3]提出后解码需求机制,它是通过对方案信息box(scheme information box)中信息的添加和修改实现的。

OMAF标准在原有的基本文件格式标准基础上,加入了映射、打包和鱼眼视频等相关box的表达,有关OMAF的新增信息如图8所示。其中,POVD(projected omnidirectional video box,映射

全景视频 box)中有许多映射方式的信息,例如映射结构、方向等视频信息;FOVD(fisheye omnidirectional video box,鱼眼全景视频 box)包含了鱼眼全景视频的参数信息,且用于指示视频轨道中的样本包含了由鱼眼相机捕获的图像信息;RWPk(region-wise packing box,区域打包 box)包含了视频帧的宽高信息,并且表示图像映射后通过了区域打包模块,并且需要在渲染之前进行解包处理。其中,在POVD内部,PROR(projection orientation box,映射方向 box)包含了映射模型的坐标系方向信息,如方向角度、偏转信息;COVI(coverage information box,覆盖信息 box)可以根据POVD的信息以及映射的视频帧信息来表征全景球体表面的相关参数与信息,若POVD中不包含COVI,则指示映射的视频帧可用于表示整个球面。

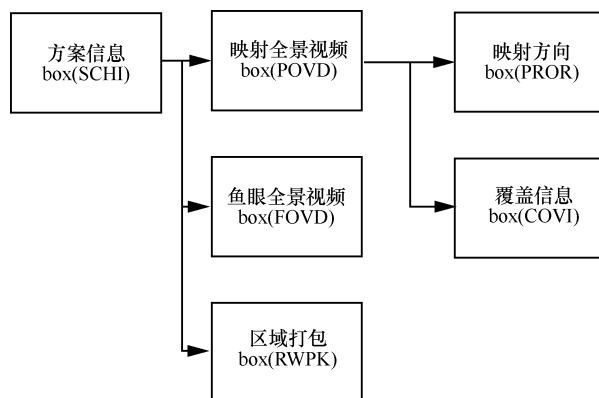


图8 OMAF在方案信息box中的新增信息

除了加入全景媒体的相关box的表达,在OMAF的文件存储格式领域,还有一些目前正在研究的技术,例如,子图像的多路track技术和区域视频质量排名技术等。

3.2 Facebook全景媒体应用方案

VR技术是目前最受关注的前沿科技之一,受到了各家互联网公司的青睐,越来越多的大型科技公司开始涉足VR领域。无论是VR社交还是VR游戏,这些仅仅是Facebook的VR展现形式而已,而支撑VR应用的核心是一个支持全景媒

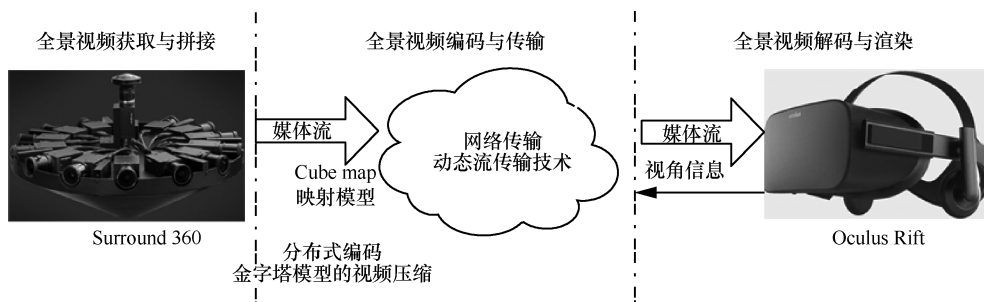


图9 Facebook 的全景视频关键技术示意

体的数字通信架构。这与 MPEG OMAF 架构类似, Facebook 在全景媒体的应用架构中从媒体获取到渲染播放端的关键技术如图 9 所示。

3.2.1 基于 Surround 360 相机的视频获取与拼接

全景媒体应用框架的输入是对 360 视频的获取模块, Facebook 在 2016 年发布了 Surround 360 摄像机, 并且将硬件设计和图像拼接代码开源到网上。Surround 360 由环绕 360° 的 17 台摄像机组成, 它将拍摄到的多路视频进行拼接, 并将其转换成适合于 VR 观看的立体 360° 全景。

在拼接模块, Surround 360 采用了基于光流的算法, 用光流来计算左右眼立体视差, 对左眼和右眼分别合成对应视角方向的虚拟摄像机的新视图, 然后再将左右眼的视图重新组合。这种方法可以捕捉场景的运动, 以达到远胜于普通拼接的无缝立体效果, 拼接后的输出可为每只眼睛提供最大 8 kHz 的视频。

3.2.2 正六面体映射方式

Surround 360 以经纬图映射的方式输出, 这种映射模型会在顶上与底下两部分包含较多的冗余信息, 且呈现效果较为扭曲, 不符合人眼视觉习惯。Facebook 在映射方式上选择了正六面体的方法, 将经纬图重新布局到正六面体上。正六面体映射方法有很多优点, 比如在立方体的每一个面上没有任何失真, 每一面的映射都是相对独立的。其次, 视频编解码器中运动矢量为直线形式^[21], 正六面体不会像经纬图方法那样将图像扭曲, 因

此这种映射方式对编解码器非常友好。此外, 它的像素点分布较为均匀, 不包含冗余信息。

在 Facebook 的方案中, 为了实现从使用经纬图方法显示到立方体映射的转换, 它创建了一个自定义的视频过滤器, 使用多点投影的方式进行二者之间的像素点切换。这套方案将经纬图顶部的 25% 转换为一个立方体面, 将底部的 25% 转换为另一个面, 中间的 50% 转换为 4 个面。通过这样的映射转换, 每帧的像素数量减少了 25%, 从而提高了空间的效率^[22]。

3.2.3 金字塔模型

金字塔模型是一个与视角依赖的立体映射模型, 它的底部为用户视角区域的全分辨率视频, 随着金字塔高度的上升, 在金字塔其他面上的视频压缩率逐渐增加。当用户切换视角时, 用户看到的不是该金字塔其他表面的低质量视频, 而是切换到以下一视角为底部全分辨率的金字塔模型。Facebook 采用金字塔映射模型可以对文件进行压缩, 能够将全景视频文件压缩到原来的 20%^[22]。

在 Facebook 的方案中, 一个经纬图的输出将被转换为 30 个视角的金字塔模型, 基本能覆盖整个全景视频的各个视角空间。每一个金字塔有 5 种不同的分辨率版本, 因此, 对于一个全景视频, 一共有 150 个不同版本的编码流可以根据情况而选择。

此外, 为了在合理的时间内处理海量数据的

全景视频, Facebook 也使用了分布式编码, 在多台机器上编码不同的视频分块。

3.2.4 动态流传输技术

由于有限的网络带宽与计算能力的限制, 传输超大数据规模的全景视频会造成缓冲或者中断等问题^[23]。Facebook 针对这些难题, 提出了基于视角的自适应比特流技术, 在视觉感兴趣的区域提供最高质量的视频, 同时降低外围背景的视频质量。在客户端针对目前的网络条件以及综合分辨率、视频质量、当前视角方向等元素可以考虑数十种不同可能的流来呈现。其次, 在传输中服务端需要频繁地更新网络状况, 以更短的时间估计网络带宽, 这样能保证系统及时调整, 避免缓冲时延的发生。

此外, Facebook 还开发了基于内容的动态流技术, 它主要是依靠人工智能给出的显著图数据来实现的, 它利用显著图的统计数据, 计算出观看者可能的关注点和感兴趣点。

3.3 Google 全景媒体应用方案

2014 年, Google 推出了一款价格低廉的 DIY 设备 Cardboard, 受到市场青睐。而如今, Google 已经凭借其各种高端交互式设备, 站到了 VR 技术领域的最前沿。图 10 所示为 Google 全景媒体的基本架构。

3.3.1 Google Jump 全景摄像装置

在全景媒体应用框架的输入口, Google 推出了 Google Jump, 它是由 16 台 GoPro 镜头组、自动拼接软件和播放平台 3 个部分组成。Jump 拍摄

的原始视频经过 Jump 应用转换后, 会生成非常逼真的 3D 虚拟现实视频, 是 Google 虚拟现实视频内容制作的设备。

3.3.2 ETC2Comp 编码技术与 Draco 压缩库

在编码压缩这一环节, Google 发布了 ETC2Comp 技术, 它是一款用于游戏和 VR 开发的编码器。在编码 360°视频的过程中, ETC2Comp 通过部署一些优化技术, 能够以更快的速度获得高质量的视觉效果。在优化策略中, ETC2Comp 通过“定向块”的搜索方式有针对性地获得给定块的最佳编码方式, 这种压缩方式比使用暴力法快得多。在代码方面, 基于每个视频块可以进行独立编码, ETC2Comp 采用了高度多线程。此外, Chrome Media 团队创建了 Draco, 这是一个开源的压缩库, 用于改善 3D 图像的存储和传输性能。Draco 压缩库提高了 3D 图像的压缩效率而不会影响视觉保真度。

3.3.3 空间化音频技术

除了对全景视频方面的处理, Google 还在整个全景媒体框架中引入了空间化音频技术, 将空间音频引入网页, 将浏览器转换成一个 VR 媒体播放器, 用户可以听到浏览器上的 360°环绕声, 其流程如图 11 所示。解码器能记录包含 4 通道的音频, 然后将其解码成任意的扬声器设置。此外, 还使用 8 个虚拟扬声器代替实际的扬声器, 阵列以双向呈现最终音频流。这种双耳呈现的音频可以传达空间感, 为沉浸式的 VR 体验发挥了关键作用。

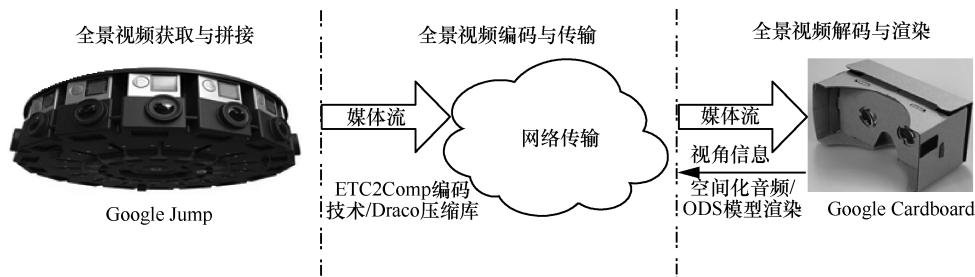


图 10 Google 全景媒体关键技术

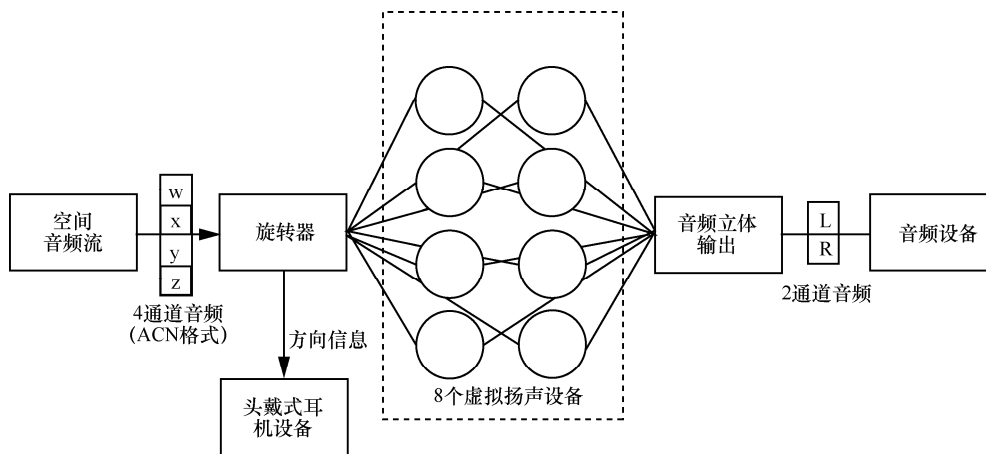


图 11 空间化音频技术的流程

3.3.4 基于 ODS 模型的渲染技术

Google 设计了一个 ODS (omnidirectional stereo, 全景立体视频) 系统, 它能够实现无缝拼接与映射, 还可以采用传统视频格式进行存储。ODS 映射的算法只捕获每个摄像机的中心射线, 并借用周围摄像机的其他射线方向, 因此, 这种光线捕获方法呈现在左右眼的射线几何将会变得更加立体和全方位。此外, ODS 都是采用的预渲染方式, 在头戴式设备中播放都非常流畅。

虽然 Google 加入虚拟现实领域迟于 Facebook, 但发展势头很猛, 相比于 Facebook 有自己社交平台的优势, Google 也致力于以 Android 操作系统为核心的 VR 应用, 同时也在推行一体化 VR 头盔, 即不再需要连接智能手机或电脑。不同于 Facebook 收购 Oculus 公司开发 Rift 头戴

式设备, Google 从硬件到生产再到内容输出正在构建自己的闭环系统和框架, 并且在前后端也继续研究全景媒体的应用。

4 全景媒体应用架构的对比

综合几家全景媒体应用架构的特点, 从多个维度对以下几个框架做对比分析, 具体见表 1。

除了表 1 中所示, 在结构严谨性上, MPEG 的 OMAF 标准中, 在各个模块环节如映射、编码、传输等方面都提出了不同的解决方案。此外, 还进行了模块间的对比性实验, 在很多细节上都进行了完善和优化。在这一点上, Facebook 和 Google 在框架中的大多数模块中都没有平行方案。

从发展进程上而言, Facebook 和 Google 这两个具有代表性的科技公司, 主要都是在走以产品

表 1 多个维度的框架对比分析

	OMAF	AVS	Facebook	Google
获取设备	全景相机		Surround 360	Google Jump
映射	多种映射模型, 如金字塔、立方体等	10 种高效映射模型, 如圆柱体、双环带等	正六面体、金字塔模型映射方式	ODS 映射模型
编码	若干种基于视角编码方案	兼容目前主流的平面视频编码标准	分布式编码	ETC2Comp 编码技术与 Draco 压缩库
传输	基于视角的 DASH 和 MMT 方案	可支持 HLS、DASH 和 MMT 传输	动态流传输技术	兼容传统方案
呈现设备	Gear VR、HTV Vive、Oculus Rift、Google Cardboard 等			

为核心、技术为支持的产业路线，在系统实现中去优化相应的关键技术点，使系统更加完善，更加符合市场需求。而目前，MPEG 和 AVS 的框架中各个模块的细节和采用的算法还在进一步研究和讨论中，因此并未形成行业内的标准，没有投入产业使用。

从行业标准上而言，由于 2016 年虚拟现实行业热潮爆发，各种 VR 产品参差不齐，为了避免 VR 市场分散，需要全景媒体框架的标准化，MPEG OMAF 的领头性是不容小觑的。但是，这一标准并不限制日新月异的 VR 市场创新和少数产品的各异性。而 Facebook 和 Google 作为创新性公司，其架构与 OMAF 框架存在差异，但并不影响其后续的发展与延伸。

5 结束语

本文对全景多媒体的应用架构展开了详细的介绍。首先回顾了目前虚拟现实技术的发展以及存在的问题。其次，引出了 MPEG 下的 OMAF 这一通用格式的作用以及它近期发展和面临的挑战，同时也对比了国内标准组织 AVS 以及在工业界 VR 技术领域占主导地位的两家公司——Facebook 和 Google，在全景媒体应用方面的进展和现状。针对 MPEG OMAF、Facebook 和 Google 的全景媒体系统架构，进行了详细的介绍和分析，并进行了对比总结。

总的来说，全景媒体应用框架的讨论、制定和完善是一个具有挑战性的课题。8K 甚至更大分辨率的全景视频对于网络带宽提出了高难度的需求。随着全景媒体直播技术的发展，时延将是影响用户体验的一个重要参数，而终端显示设备播放的视频质量也决定了用户观看效果。这些关键技术的研究将对今后虚拟现实技术的发展具有十分重要的研究意义和应用价值。尽管目前 VR 技术在游戏、社交等领域发展迅速，但它内部的系统结构仍需要继续细化和完善，全景多媒

体应用的发展还处于起步阶段。如今，越来越多的组织和企业都加入制定 VR 行业标准的队伍中，提供了新的思考和方法，因此值得展开更为深入的研究。

参考文献:

- [1] 董振江, 张东卓, 黄成, 等. 虚拟现实视频处理与传输技术[J]. 电信科学, 2017, 33(8): 45-52.
DONG Z J, ZHANG D Z, HUANG C, et al. Technology of virtual reality video processing and streaming[J]. Telecommunications Science, 2017, 33(8): 45-52.
- [2] ISO/IEC JTC1/SC29/WG11/M37044. Use cases and requirements of omnidirectional media format, MPEG 113[S]. 2015.
- [3] ISO/IEC JTC1/SC29/WG11/N16189. WD on ISO/IEC 23000-20 omnidirectional media application format, MPEG 116[S]. 2016.
- [4] ISO/IEC JTC1/SC29/WG11/M37152. Future video coding requirements on virtual reality, MPEG 113[S]. 2015.
- [5] ISO/IEC JTC1/SC29/WG11/M37914. Support of vr video in iso base media file format, MPEG114[S]. 2016.
- [6] ISO/IEC JTC1/SC29/WG11/M38662. Dash descriptor for omnidirectional media, MPEG 115[S]. 2016.
- [7] WANG R G. VRU (virtual reality union) meeting report, AVS-N2422[R]. 2017.
- [8] ISO/IEC JTC1/SC29/WG11/N16824. Text of iso/iec dis 23090-2 omnidirectional media format, MPEG 118[S]. 2017.
- [9] ISO/IEC JTC1/SC29/WG11/M39469. OMAF: lens shading parameters metadata for omnidirectional video, MPEG 116[S]. 2016.
- [10] ISO/IEC JTC1/SC29/WG11/M39380. OMAF: camera and lens parameter for stereo 360-degree video using fisheye cameras, MPEG 116[S]. 2016.
- [11] ISO/IEC JTC1/SC29/WG11/M40178. Update of OMAF PROJ-VE description, MPEG 117[S]. 2017.
- [12] ISO/IEC JTC1/SC29/WG11/N16828. Projections under considerations for ISO/IEC 23000-20 omnidirectional media application format, MPEG 118[S]. 2017.
- [13] BUDAGAVI M, FURTON J, JIN G, et al. 360 degrees video coding using region adaptive smoothing[C]//IEEE International Conference on Image Processing, Sept 27-30, 2015, Quebec City, Canada. Piscataway: IEEE Press, 2015: 750-754.
- [14] JIN G, SAXENA A, BUDAGAVI M. Motion estimation and compensation for fisheye warped video[C]//IEEE International Conference on Image Processing, Sept 27-30, 2015, Quebec City, Canada. Piscataway: IEEE Press, 2015: 2751-2755.
- [15] SKUPIN R, SANCHEZ Y, HELIGE C, et al. Tile based HEVC video for head mounted displays[C]//IEEE International Symposium on Multimedia, Dec 11-13, 2016, San Jose, USA. Pis-



- cataway: IEEE Press, 2016: 399-400.
- [16] HOSSEINI M, SWAMINATHAN V. Adaptive 360 VR video streaming: divide and conquer[C]//IEEE International Symposium on Multimedia, Dec 11-13, 2016, San Jose, USA. Piscataway: IEEE Press, 2016: 107-110.
- [17] 祝谷乔, 宋皓. MPEG-DASH 与 HLS 流传输技术的比较分析[J]. 电信科学, 2015, 31(4): 23-27.
- ZHU G Q, SONG H. Comparison and analysis of MPEG-DASH and HLS adaptive streaming delivery technology[J]. Telecommunications Science, 2015, 31(4): 23-27.
- [18] SREEDHAR K K, AMINLOU A, HANNUKSELA M M, et al. Viewport-adaptive encoding and streaming of 360-degree video for virtual reality applications[C]//IEEE International Symposium on Multimedia, Dec 11-13, 2016, San Jose, USA. Piscataway: IEEE Press, 2016: 583-586.
- [19] ZARE A, AMINLOU A, HANNUKSELA M M, et al. HEVC-compliant tile-based streaming of panoramic video for virtual reality applications[C]//Proceedings of the 2016 ACM International Conference on Multimedia(MM), October 15-19, 2016, Amsterdam, The Netherlands. New York: ACM Press, 2016: 601-605.
- [20] XIE L, XU Z M, BAN Y X, et al. 360ProbDASH: improving QoE of 360 video streaming using tile-based http adaptive streaming[C]//Proceedings of the 2016 ACM International Conference on Multimedia(MM), October 15-19, 2016, Amsterdam, The Netherlands. New York: ACM Press, 2016: 521-525.
- [21] 周建同, 杨海涛, 刘东, 等. 视频编码的技术基础及发展方向[J]. 电信科学, 2017, 33(8): 16-25.
- ZHOU J T, YANG H T, LIU D, et al. Trends and technologies of video coding[J]. Telecommunications Science, 2017, 33(8): 16-25.
- [22] FAUTIER T. VR video ecosystem for live distribution[J]. IBC 2016 Conference. 2016, 28(13): 32-36.
- [23] QIAN F, JI L S, HAN B, et al. Optimizing 360 video delivery over cellular networks[C]//Proceedings of the 5th Workshop on All Things Cellular: Operations, Applications and Challenges,

October 3-7, 2016, New York, USA. New York: ACM Press, 2016: 1-6.

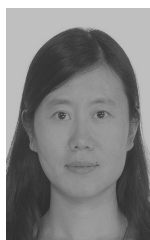
[作者简介]



罗莹 (1993-), 女, 上海交通大学图像通信与网络工程研究所硕士生, 主要研究方向为视频通信与处理、虚拟现实技术。



宋利 (1975-), 男, 上海交通大学图像通信与网络工程研究所研究员, 主要研究方向为图像处理、视频编码。



解蓉 (1975-), 女, 上海交通大学图像通信与网络工程研究所副教授, 主要研究方向为视频编码与通信、视频处理。



罗传飞 (1978-), 男, 中国电信股份有限公司上海研究院高级工程师, 主要研究方向为 IPTV 系统、视频体验技术等。