

第六章 相关国际标准组织

增强现实 AR 和虚拟现实 VR 的出现使得人们可以利用技术不断地去改进描述世界的方法。人们通过多种多样先进的设备，获得前所未有的“沉浸式体验”。所谓的“沉浸式体验”，是指用户身处于接近真实的虚拟世界，通过音视频技术，使用户完全投入情境当中。

近年来，VR 行业的发展态势良好，许多团队如 HypeVR、NextVR 始终致力于 VR 各环节的技术研究，Facebook、Samsung、HTC 等知名厂商也着眼于 VR，带来了许多先进、便捷的产品。一些体育赛事转播平台 BT Sport、Sky UK 等也已引进 VR 设备，为观众带来 360 度的观看体验。

根据高德纳咨询公司 2017 年度关于新兴科技的调查报告，VR 产业目前正处于复苏期，还需 2-5 年方可达到一个平衡发展的状态。目前的 VR360 视频还存在着分辨率低、头部运动范围小、观看设备庞大等局限性。VR 的最终目标是实现“6 自由度”的完全沉浸式体验，让人感到身临其境，并具有良好的交互感。

目前，多个国际组织正致力于沉浸式媒体的标准制定和研究，各组织致力于 VR 的不同方面，共同推进 VR 产业标准的发展，指导行业相关人员的技术研发。本章将主要介绍这些组织在 VR/AR 方面的近期标准工作进展。



图 6.1 VR360 标准概览

涉及沉浸式媒体标准制定的国际组织非常多，如 MPEG，3GPP，Khronos，IEEE 等。还有一些工业论坛，如 VR-IF，代表相关产业界在产业方向上对沉浸式媒体发展的要求。

其中，MPEG 已经创造并且仍在制作媒体标准，适应工业需求，推动市场的发展。MPEG 将沉浸式媒体方面的工作放在一起统称为 MPEG-I（MPEG Immersive Media）。

6.1 MPEG-I

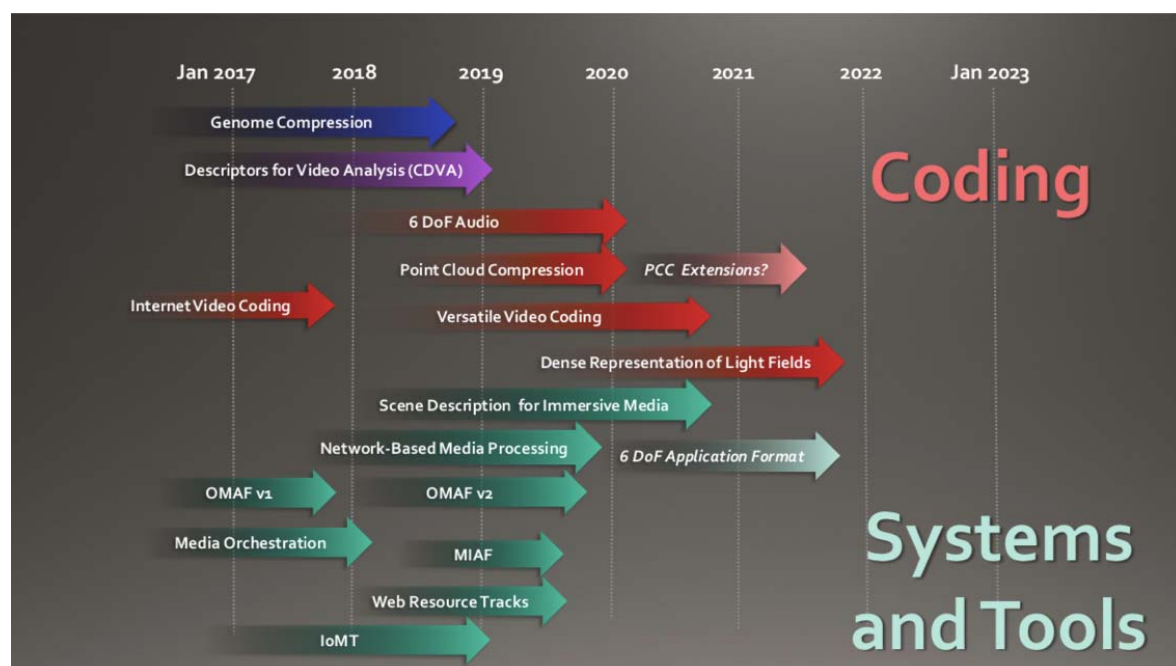


图 6.2 MPEG 对于沉浸式媒体的蓝图规划

MPEG 目前已经公布了其长期规划蓝图（约 5 年），如图 6.2，并收集了各大厂商的反馈和要求。其在沉浸式媒体网络视频编码、应用架构、服务调度等方面都已经有了相应的标准，未来将会把重点放在 6 自由度媒体、内容结合、VR360 直播与点播上。以下是 MPEG-I 标准中最主要的八部分内容：

- 沉浸式媒体架构；
- 全景媒体应用格式（OMAF）；
- 通用视频编码；
- 新型沉浸式音频编码；
- 点云编码；
- 沉浸式服务与应用元数据；
- 沉浸式服务与应用度量指标；
- 基于网络的媒体处理。

OMAF 框架下的 VR360 支持 HEVC、AVC 视频编码标准，MPEG-4 AAC、MPEG-H 音频编码标准，DASH、MMT 等多种流传输协议。在虚拟现实和增强现实的环境下，MPEG 期望给在复杂的，支持交互的场景中的媒体渲染定义规范，就像为 3 个 DoF 媒体定义的 OMAF 一样。

6DoF 应用格式将支持在 MPEG 容器中聚合和打包媒体，用于存储，下载，流媒体和广播分发。可以从场景图/场景描述文件中读取场景，或者可以从内容推断场景。

2018 年 7 月，MPEG-I 报告表示其已经完成了以下的部分工作，并仍需持续推进：

- 3DoF+和 6DoF 测试材料的征集
- 确定并核实 3DoF+和 Windowed 6DoF 的通用测试环境
- 扩展 3DoF+和 6DoF 的参考软件和配置
- Windowed 6DoF 和全景 6DoF 的探索性实验
- 关于密集光场压缩技术的探索性实验

以下将简要介绍 MPEG 构建的沉浸式媒体标准架构，主要作用于 6DoF 应用。

网络组件

控制和管理功能：用于建立支持 MPEG 的客户端和支持 MPEG 的网络元素之间的通信，或多个网络元素之间的通信。预计这种功能并不会在 MPEG 标准范围内，除非有明确的要求。

MPEG 内容来源：托管 MPEG 格式内容并且可以使用 MPEG 定义的协议访问的服务器（集中式/分布式）。

MPEG 媒体网络感知/处理功能：建立 MPEG 内容源或支持 MPEG 的客户端或两者的通信，以支持沉浸式体验。

支持 MPEG 的客户端：提供使用沉浸式 MPEG 内容的所有方法。

设备组件

- 底层（5G, WiFi, IP）
- 媒体访问客户端（DASH 客户端, MMT, 文件系统）
 - 提取基本流和元数据，并使其可供解码器和应用程序使用
- 上行客户端
 - 上行媒体：媒体编码器、内容交付协议
 - 元数据
- 解密
- 媒体解码器
 - 音频
 - 视频
- 应用程序/引擎
 - 场景图
 - 管理
 - 场景描述
 - 脚本
 - 定时跟踪
- 渲染
 - 音频
 - 视觉
- XR 功能
 - XR 输入
 - XR 显示
 - XR 合成
 - 传感器数据
- 捕获
 - 话筒
 - 相机
 - 其他

传统/第 1 阶段架构

第一阶段的架构基于 OMAF，如图 6.3 所示。

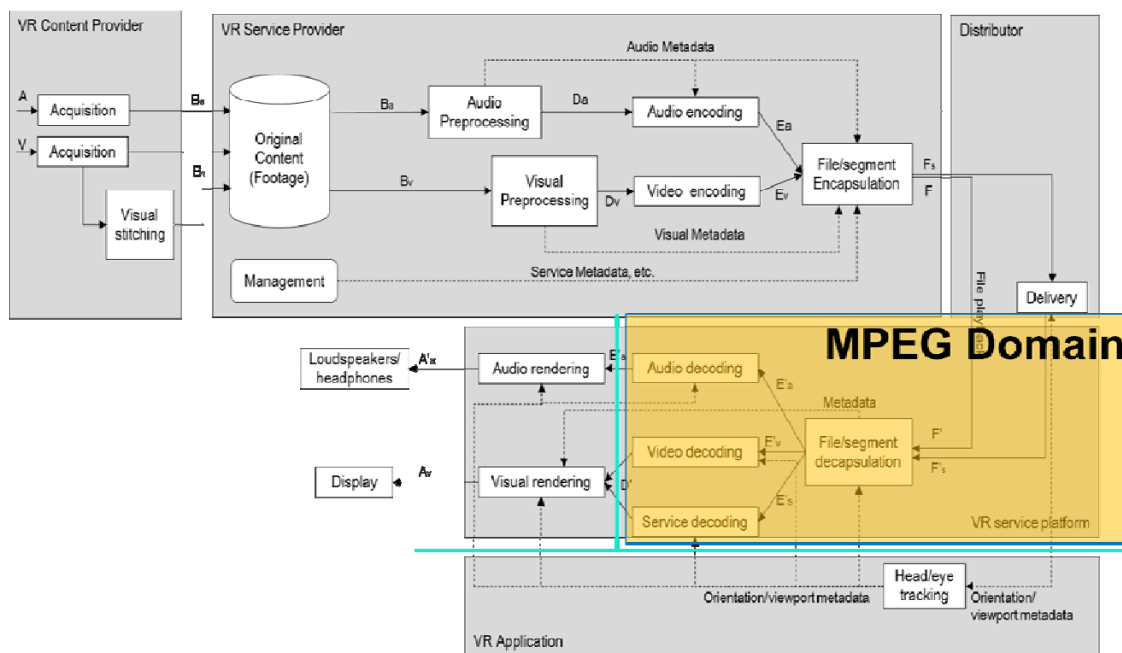


图 6.3 OMAF 架构和关键的 MPEG 域定义

从 MPEG 传统架构和系统模型开始，如 MPEG-2 TS, DASH, ISO BMFF 以及 MMT，系统标准主要提供：

- 1) 不同媒体流的定时和同步
- 2) 多路复用或基础媒体流的组合
- 3) 使用诸如随机访问内容的系统功能
- 4) 所需性能的描述
- 5) 流属性的描述
- 6) 能够在不同的网络场景下提供内容：适应，切换，错误恢复
- 7) 其他系统功能，如加密

用户/界面与传统内容“交互”的能力仅限于随机访问，搜索和组件选择。这种选择/交互不仅影响媒体播放，还影响媒体解码和传送。

在 MPEG-I 第 1 阶段中，通过提供多维视点，扩展了应用程序与媒体堆栈交互的体系结构。这种相互作用导致额外的动态和优化过程。除此之外，体系结构相对不变。

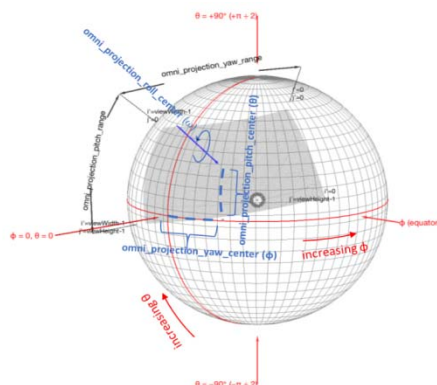


图 6.4 将视点相关的传感器数据添加到媒体消耗中

第 2 阶段架构

该架构预见到客户端通过不同的模式进行输入，以考虑 3DoF / 6DoF 甚至 AR / MR 等应

用。AR/MR 应用通常从本地摄像机和麦克风输入。

渲染由图形渲染引擎和 2D/3D 音频渲染引擎执行，其将解码的媒体资源合成在一起并呈现给用户。

该架构支持 MPEG 和非 MPEG 媒体资源。容器解析器提取有关资源和媒体时间线的信息以及任何嵌入或引用的媒体资源，并使它们在相关引擎中可用。架构支持多同形式的内容使用，例如简化的 2D 版本可以在简单的客户端中呈现，客户端也可以使用 3DoF，3DoF+或 6DoF 版本。

为了描述展示的场景，可以使用场景图，也可利用替代格式提供场景图。例如，通过容器格式描述基本的呈现操作，可以简单客户端。其他场景描述文件也可以包含在容器中。

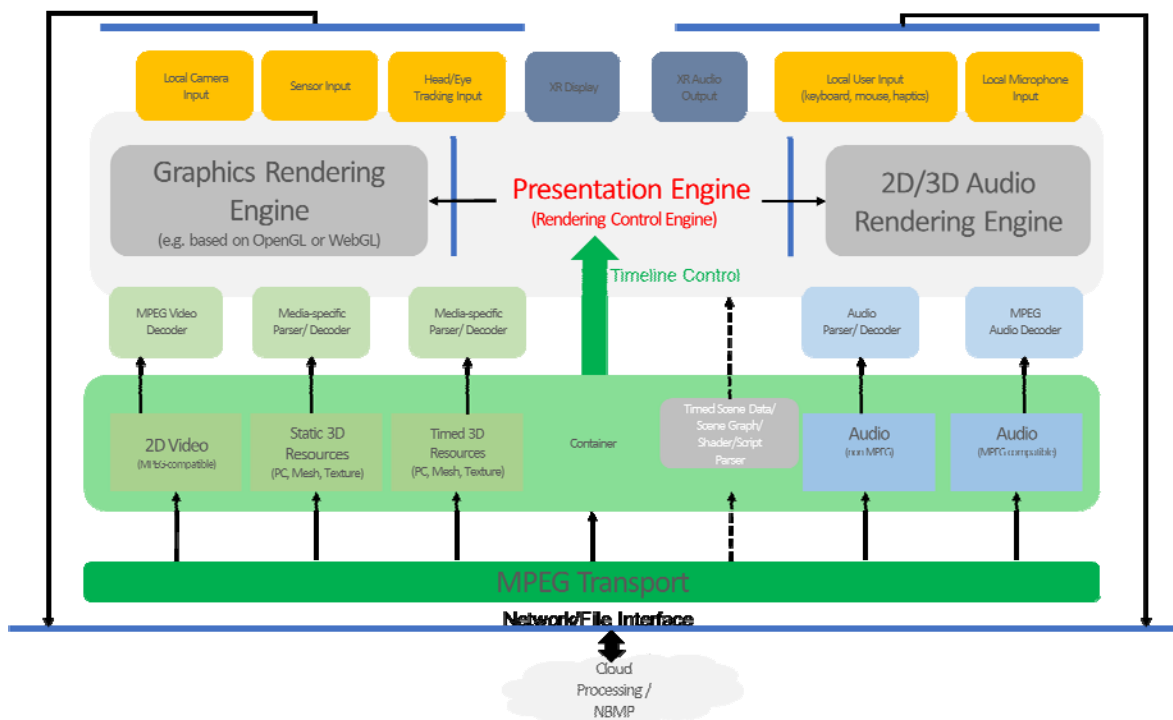


图 6.5 起草的第 2 阶段参考架构

以上为第 2 阶段的一般架构。目前 MPEG 已经考虑从网络、VR 社交性等方面形成一些针对性的替代方案。以下将展开讲述一种 MPEG 提出的以传输为首要目标的媒体转变思想。

MPEG-I 项目中的流媒体优先设计

1) 从广播和存储媒体到流媒体

目前，媒体行业的技术创新主要由基于流的应用程序所驱动。对于传统的视频应用，例如视频流服务（Netflix, Hulu, Amazon Prime），情况确实如此，而对于沉浸式视频应用来说更是如此。由于沉浸式体验会占用大量带宽和存储空间，因此内容传递变得越来越个性化、多样化，以使数据流实时地适应消费者的确切需求。

在 MPEG-I 的背景下，新的媒体格式正在被标准化。对于这些格式，预期的应用/服务均是基于流的。此类示例包括 Google Light Fields, Samsung VR, Next VR 等。一对多分发系统（如广播）本身就不太适合这种沉浸式体验，因为单播技术例如 MPEG-DASH 这样的 HTTP 自适应流更适用于此类体验。虽然已经对广播 VR/360 内容进行了一些实验，但随着视频带宽和“个性”的增加，广播变得越来越不可能。当然，也可以想象一种混合模型，其中的一部分内容是一对多发布的，而其余部分仅传送至特定客户端。

这一转变有许多重要的含义：

1. 以发送方为中心的模型正在被以接收方为中心的模型所取代；
2. 很大一部分的传输数据对于接收端而言是唯一的；
3. 传输的数据需要实时适应接收端的动态需求，且必须以尽可能低的延迟提供；
4. 虽然总数据量增加，但最关键的点在于“基础可传输数据单元”的大小需要减小。

这种转变要求 MPEG 的设计原则也有所改变，这就需要考虑高效的流传输策略，或者压缩数据量。然而，目前 MPEG-I 的开发过程中缺乏对流媒体策略的考虑，同时对流媒体尽早的关注将大大增加采用这些未来 MPEG 技术的机会。因此，MPEG-I 提出改变设计原则。

2) 当前方法：从编码到传输的标准化

MPEG 最初是从编码格式的标准化开始的，之后定义了内容的存储，最后指定了传输格式。近年来，媒体已经由基于文件的应用程序和一对多传送驱动。上述传统方法在传输数据适配明确的带宽约束方面非常成功。

然而对于流应用，带宽约束不再是先验、已知的，并且不一定在小范围内，而是具有从例如 2Mbit/s 到 100Mbit/s 的连续频谱，甚至更多。此外，客户端设备很可能会受到传送网络质量的影响。而近年来出现的新型自适应技术一定程度上解决了这个问题，例如 MPEG-DASH，其中接收端负责在给定网络条件的情况下最大化体验质量。

传统的方法是首先考虑压缩，然后再进行传输，这可能导致解决方案不是最优，因为：

- 在存储步骤中做出的决定可能会禁止某些流传输策略
- 流传输步骤可以简化为存储格式的线性传输
- 设计时可能不会考虑延迟等网络因素，然而这些因素对于任何实际部署都是至关重要的

3) 改进方法：流媒体优先的标准化

“流媒体优先的设计”是在考虑封装之前，着眼于媒体数据的流媒体策略。目的是确保文件格式设计能够实现沉浸式体验的高效流传输而不产生不必要的开销，这一原则允许在延迟和效率之间进行权衡。图 6.6 描述了这种标准化过程的三个主要步骤。

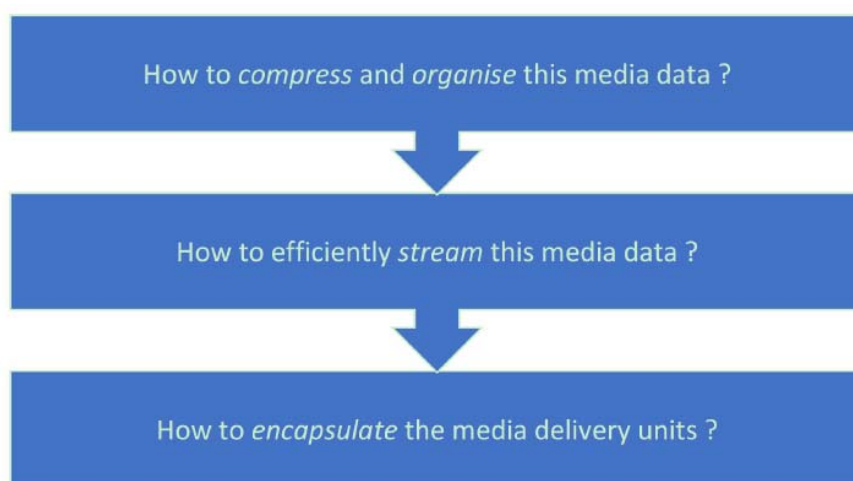


图 6.6 流媒体优化原则

流传输步骤将重点关注：

- 渲染媒体体验时的适应维度是什么？（例如，带宽，FoV，空间位置等）
- 相关的传输协议是什么以及如何充分利用它们？（例如，HTTP/1.1，HTTP/2，QUIC，WebRTC 等）
- 我们应遵循哪种策略以传输最少的数据并最大化 QoE？（例如，基于视角的流传输

等)

- 如何应对网络延迟? (例如, 预提取)
- 如何根据新的 MPEG-I 编解码器实现契合直播体验的端到端延迟?

目前, 流媒体优先标准化的拟议范围包括:

- 短期
 - 提取 OMAF WD 的 Annex D 进行技术研究和整理
 - 为 Annex D 准备参考软件
 - 定义测试条件并进行实验以测量示例流条件下的传输比特率。
- 中长期
 - 高效利用 MPEG-I 媒体资源的现代网络协议的研究
 - 研究利用 MPEG-I 流媒体资源的当前 MPEG 技术 (例如, ISOBMFF, DASH, MMT) 的局限性, 并可能在必要时启动新的流式格式

这种工作的最终目标是优化流式 MPEG-I 内容的 QoE, 因而, 可以期待一个独立于 OMAF 的规范、标准生成。

6.2 IEEE P2048

目前, IEEE 标准协会正在为虚拟和增强现实开发一系列标准。下表提供了关于 IEEE P2048 系列标准工作组的简要描述。在 2017 上半年, 该工作组宣布了针对虚拟现实和增强现实的 8 个 IEEE 标准项目, 目前已经拓展到 12 个相关项目。

IEEE P2048 部分	简要描述
P2048.1 设备分类及定义	此标准规定了关于 VR,AR 设备的分类和定义
P2048.2 沉浸式视频分类和质量度量 (P)	此标准规定了沉浸式视频的分类和质量度量, 包括几个方面: 360 度或 180 度, 是否立体, 观察点是否可移动, 焦点是否可调整等等
P2048.3 沉浸式视频文件和流格式	此标准规定了沉浸式视频文件和流格式, 以及格式所支持的功能和交互。
P2048.4 个人身份	此标准规定了在 VR 中验证 (识别) 人的身份的要求和方法。
P2048.5 环境安全	本标准规定了虚拟现实 (VR) 工作站和内容消费环境的建议。
P2048.6 沉浸式用户界面	此标准规定了在 VR 应用中启用沉浸式用户界面的要求和方法, 以及沉浸式用户界面提供的功能和交互。
P2048.7 真实世界中的虚拟对象的映射	此标准规定了针对 AR,MR 应用的要求, 系统, 方法, 测试和验证来创造和使用真实世界中的虚拟对象的映射
P2048.8 虚拟物体与现实世界的互操作性	此标准对于虚拟物体与现实世界的互操作性, 在 AR,MR 应用中规定了要求, 系统, 方法, 测试和验证
P2048.9 沉浸式音频的分类和质量度量	此标准规定了沉浸式音频的分类和质量度量。
P2048.10 沉浸式音频文件和流格式	此标准规定了沉浸式音频文件和流, 以及格式支持的功能和交互。
P2048.11 车内的增强现实	此标准定义了协助车辆司机或乘客的增强现实 (AR) 的总体框架系统。
P2048.12 内容评级和描述	此标准定义的对于 VR, AR 和 MR 的内容评级和描述

图 6.7 IEEE P2048 12 个标准项目

表中项目将由 IEEE 虚拟现实和增强现实工作组负责, 参与者包括设备制造商、内容提供商、服务提供商、技术开发人员和政府机构等等。目前, VR/AR 相关的技术正在以非常快的速度发展, 而现有的开发标准仅涵盖了 VR/AR 领域的小部分, 该工作组正全方面努力扩展, 以达到满足技术领域的标准化需求。

6.3 IEEE P3333.3

IEEE P3333.3 标准主要的关注点在于头戴式显示器 (HMD) 的 3D 内容运动失真问题。该标准组织目前正通过对焦点失真的视觉反应、对镜片材料的视觉反应、对镜片折射率的视觉反应、对帧率的视觉反应等四个方面的研究提供一个用于解决“基于 HMD 的 3D 内容运动失真引起的虚拟现实 (VR) 失真”的技术指导。



图 6.8 IEEE P3333.3 WG 结构

6.4 Khronos



图 6.9

在 2017 年 GDC 大会期间，Khronos Group（科纳斯组织）公布 VR/AR 标准 OpenXR，为 VR 和 AR 应用程序定义了一个 API。在 OpenXR 中，应用程序和引擎使用标准化接口来询问和驱动设备，这样设备可以自我集成到一个标准化的驱动程序界面。同时，标准化的硬件/软件接口减少了碎片化，同时保留了实施细节，以鼓励行业创新。

由于不同的公司的系统和设备不同，这就导致了 VR/AR 设备和平台出现各种“分裂”的情况。OpenXR 的目的便是减少开发者们为支持不同设备的 API 而产生的麻烦。如果没有设立标准的 API 接口，各平台间的对接将会非常的繁琐和复杂。有了 OpenXR 标作为接口标准后，开发者只用写一次代码便能对接各个平台。Khronos 标准可用于增强用户界面，以及智能手机的游戏和应用中 3D 图形 API 的 OpenGL ES，用于异构并行计算的 OpenCL 以及用于 HTML5 的 3D 图形的 WebGL。

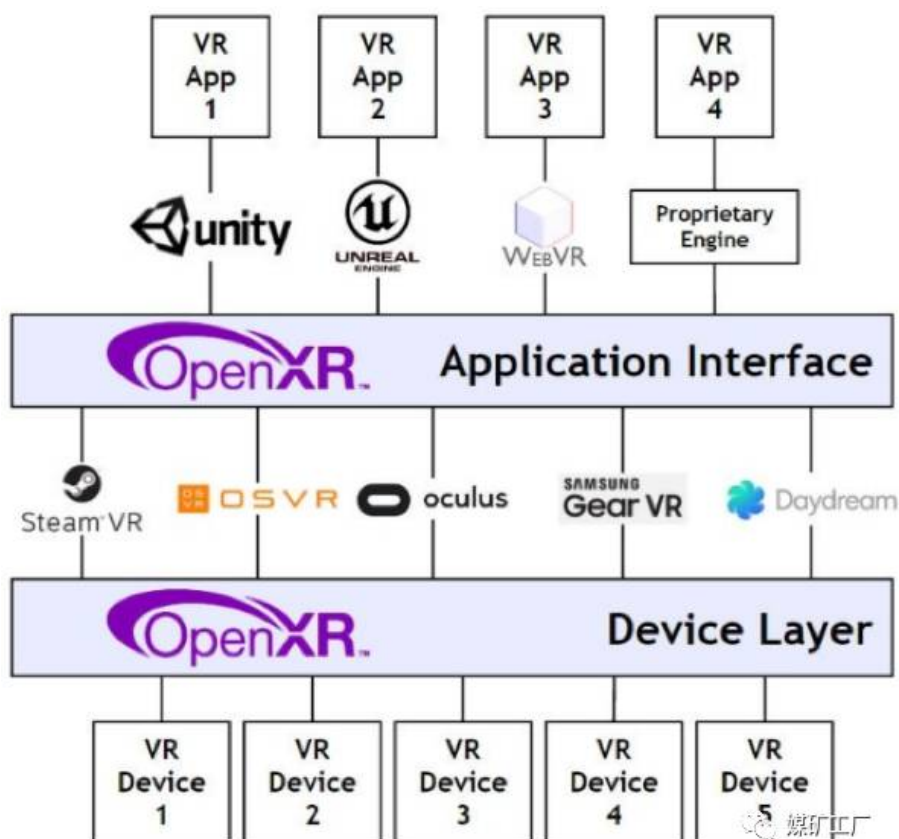


图 6.10 OpenXR 通用接口设计

6.5 WebVR

目前 VR 市场产品众多，无论在硬件还是内容服务上尚无法形成统一标准。而 WebVR 的出现，能够让诸多 VR 头显设备或是 VR 手机在获取内容的方式上统一。WebVR 提出了一个关于 VR 网络应用的开放性标准，即用户可以直接通过浏览器观看 VR 内容。在 2017 年 2 月，Google 已提出在 Chrome 浏览器上植入 WebVR，让 VR 体验更加便捷，图 6.11 为 Google 的 WebVR 实验室的设计截图。

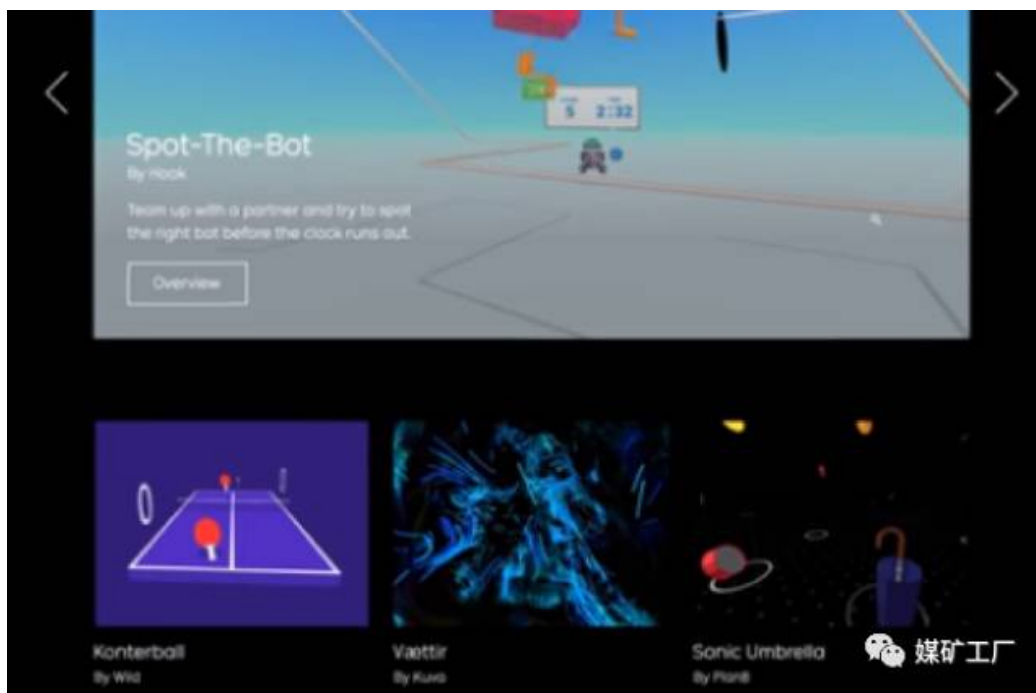


图 6.11 Google 的 WebVR 实验室的设计截图

Web XR 设备 API 规范提供 VR 和 AR 硬件的接口，使开发人员能够在 Web 上创建有趣舒适的 VR / AR 体验。它旨在完成后完全取代传统的 WebVR 规范。WebVR 的优势显而易见。它的开发门槛更低，一个普通 web 前端工程师就可以参与到 VR 应用开发中。它的跨平台性更强，可以跨设备终端、跨操作系统以及跨 APP 载体。其次，WebVR 开发快速、维护方便，可以随时进行调整，并且传播更加便捷。最重要的是，使用浏览器即可体验 VR，无需安装其他应用。2018 年 7 月，WebVR 提供了最新的 VR / AR 硬件接口。

6.6 3GPP



图 6.12 3GPP

3GPP SA4 关于虚拟现实（FS_VR）的研究项目结果被记录在技术报告 TR 26.918 中。该技术报告中主要包含以下七点：

- a) 现有的 VR 音频和视频内容制作工作流程和格式以及渲染；
- b) 一系列涵盖流媒体、会话、用户生成和遗留内容消费 VR 服务的用例；
- c) 双耳听音条件的音频质量评估；
- d) 视口独立和相关 VR 流媒体的视频质量评估；
- e) 延迟和 VR 音频/视频同步；
- f) 网络，内容和设备对 VR QoE 的影响；
- g) 差距分析和候选解决方案。

同时，3GPP SA4 启动了以下工作和研究项目：

(1) 流媒体虚拟现实配置文件（VRStream）的工作项目，其目的是为 VR Streaming 用例集定义相关的媒体和协议启用者。

(2) 沉浸式语音和音频服务（IVAS_Codec）的 EVS 编解码器扩展工作项目，总体目标

是开发用于身临其境的 4G 和 5G 服务和应用的单一通用音频编解码器。

(3) Live Uplink Streaming (FLUS) 框架工作项目，其目的是为点对点定义上行链路直播流媒体（例如 360 视频，VR，UHD，多声道音频）的框架。

(4) 针对 VR 音频的 3GPP 编解码器的研究项目 (FS_CODVRA)，其目标是评估现有 3GPP 音频编码能力是否适合启用 VR 服务，并提供如何使用和配置现有编解码器以提供最佳 VR QoE 的推荐。

(5) 关于 VR 的 QoE 指标的研究项目 (FS_QoE_VR)，其目标是调查可能需要由客户端向网络报告以评估 VR 用户体验的 QoE 参数和指标。

VR Streaming

3GPP 最终确定了 TS 26.118 (VR Streaming 的媒体配置文件) 的技术工作。最新版本 TS26.118 1.1.0 可用。

- 该规范预计将于 2018 年 9 月获得批准
- 定义视频和音频的操作点（基本流和渲染要求）以及媒体配置文件（包括文件格式和 DASH 约束）
- 视频
 - 操作点
 - 基本 H.264 / AVC: 带有 ERP 的 H.264 / AVC HP@L5.1
 - 主要 H.265 / HEVC: H.265 / HEVC MP10@L5.1, ERP, RWP, 立体视觉;从 OMAF 中选择工具
 - 灵活 H.265 / HEVC: H.265 / HEVC MP10@L5.1, 增加了立方体投影和 HDR; 高达 120 fps
 - 媒体资料
 - 基础视频: 基于 H.264 / AVC OP, 单流 HEVC, 无视点优化
 - 主视频: 基于 H.265 / HEVC OP, 样本入口 hvc1, 单个或多个独立的自适应集, 单个流呈现
 - 高级视频: 基于灵活 H.265 / HEVC OP, 样本入口 hvc1/hvc1, 提供单个或多个相关的自适应集, 单个和多个流呈现, 允许 tiling 等。
 - 所有配置文件都可以作为 OMAF 视频配置文件的子集被提供。
- 音频
 - 测试了 4 种候选方案;具体结果记录在技术报告中
 - 同意纳入兼容 OMAF 的 MPEG-H 音频
- 元数据: 支持在 2D 屏幕上渲染 360 体验。
- 在系统级别, 完成 PSS 和 MBMS 服务的集成。这意味着广播 VR 是可能的。

6.7 DVB

DVB(数字视频广播)是电视、广播和技术公司的联盟,旨在为数字电视和其他广播技术创建一套开放的技术标准。在虚拟现实(VR)发布报告后,DVB 为 VR 内容设定了新的标准。在 2017 年年中,DVB VR 活动从 VR (CM-VR-SMG)的商业模块 (CM)研究任务被推广到了 CM-VR 官方团队。CM-VR 的总体目标是提供商业需求,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。



图 6.13 DVB

CM-VR 首先将重点放在系统上，针对 DVB 集成式接收器/解码器（IRD）以及下一代 IP 连接设备，利用 DVB 宽带和广播网络向广播公司提供 VR 内容。考虑到现有的技术和现实的部署方案，CM-VR 将打算提供“全景/ 3DOF +”的视听体验。

目前，DVB 定义了 VR 在商业上获得成功的三大要素：技术、运动晕眩以及内容，其中技术方面包括制作、传输以及 VR 显示。

由于 VR 技术的规模和复杂性，DVB 在第一阶段将研究 VR 和 360 度 VR 设备的体验，第二阶段将考察更为复杂的 VR 设备领域。CM-VR 的目标是能够获得 CES2018 规定的可用的稳定要求，然后将这些标准提交 DVB 商业模块和指导委员会批准。无论如何，DVB 显然都会考虑到其他 VR 标准化组织和行业组织所做的工作，特别是 MPEG，VR 行业论坛和 3GPP。CM-VR 未来将协调不同组织的工作而做出自己的贡献。

一项调查问卷于 2018 年 2 月 26 日启动，重点是 VR / 360 / 3DOF 内容，目标是：

- 确保存在商业需求
- 关注使用情况，根据用户兴趣确定优先级

到目前为止，已收到并处理了 7 份答案：2 家技术提供商，1 家广播公司，1 家广播网络运营商，1 家专业设备制造商，2 家消费电子制造商

然而在 2018 年 7 月，考虑到对 CM-VR 组的支持程度，SB 决定暂停 VR 工作六个月，并在 2019 年 2 月在 SB91 上重新考虑该主题。

6.8 VRIF



图 6.14 VRIF

VRIF 建立的目的是为 VR 提供一个广阔的市场，维护消费者、内容设备制造商、服务提供者、广告公司等多方利益，进一步广泛提供高质量的音像虚拟现实体验。目前的重点是为分配音频和视频内容服务实现高质量，可互操作的体验。VRIF 最新发行的 VR 指南已在 CES 2018 期间公布，讲述了 VR 内容创作、传输、安全性、交互性等技术细节，进行了少量更正和其他说明。

- 即将推出的版本
 - 关于直播的工作已经开始
 - 开始着眼于 VR 和 HDR 的结合
 - 将 Presentation API 添加到指南（渲染）

- 开始解决文本和字体问题
 - 安全性：VR 内容中的水印
 - 测试和交互操作：正在逐步推进
- VRIF 指南的初始版本侧重于具有三个自由度的 360° 视频传输系统（3DOF），并包含：
- 基于 ISO MPEG 全景媒体格式（OMAF）的跨行业互操作点文档
 - VR360 内容的最佳实现方案，重点是人体因素，如晕动症
 - VR360 流媒体的安全注意事项，侧重于内容保护，同时也关注用户隐私。

6.9 QoE: QUALINET, VQEG, ITU-T

ITU 在 QoE 方面的一些工作在之前章节中已经有所介绍。目前，ITU-R WP 6C 工作组通过工作草案描述了一种原型 HMD，其空间分辨率为 $8K \times 4K$ ，完整 360° 图像的空间分辨率为 $30K \times 15K$ 。此外，他们已经开始研究高级沉浸式视听（AIAV）系统的参数值，例如图像分辨率，投影映射方法以及线性 360° AIAV 程序等。ITU-T SG12 Q13/12 目前给出了与沉浸式媒体相关的以下工作项目：G. QoE-5G（5G 网络新型服务的 QoE），G. QoE-AR（增强现实（AR）的 QoE），G. QoE-VR（VR 服务的 QoE）和 P. 360-VR（或 G. 360-VR，用于 HMD 上 360 视频的主观测试方法）。接下来将对后两个项目进行简要介绍。

G. QoE-VR

在该项目中，华为贡献了一份“G. QoE-VR 基准的更新版本”。主要增加的是关于 VR QoE 内容指标的新部分，并在 Krakow Q13/12 会议上与 VQEG 一起讨论过。

讨论总结如下：

- 来自 TU-Ilmenau 的代表认为这份文件包含的技术水平过高，建议修改它的结构，这一点得到了一致的认可。
- 之后再次提交了新的更新版本并得到同意，成为新的基准。

该文件中介绍了当前主要的 VR 应用场景：直播、点播、游戏、社交与购物。而影响 VR 的因素目前被归纳为：

- 人为影响因素
 - 视听
 - 模拟器疾病
- 系统影响因素
 - 内容相关因素：空间音频、空间深度（3D）、时空复杂性、类型等
 - 媒体/编码相关因素：音视频编码器、比特率、分辨率、帧率、音频采样率、编码延迟等
 - 网络/传输相关因素：丢包、延迟、带宽、波动等
 - HMD 相关因素：解码器性能、头部跟踪延迟、自由度、FoV、显示分辨率、刷新率等
- 背景影响因素
 - 物理背景
 - 时间背景
 - 社会背景
 - 任务前后关系

G. 360-VR

- 华为同样提交了一份“G. 360-VR 基准文件”，并收到了 TU-Ilmenau 的一些建议。
- TU-Ilmenau 提出了“使用 HMD 进行 360 度视频 QoE 评估的主观测试方法”。
 - 它比较了两种主观评价方法：修正后的 ACR 和 DSIS 用于 360 度视频 QoE 评估。
 - 结果显示 M-ACR 在统计学上更可靠，并且用户不太容易发生模拟器疾病。
 - 因此建议将 M-ACR 用于评估 360 度视频，特别是短序列，如 10s。
 - 该建议已被同意纳入基准。

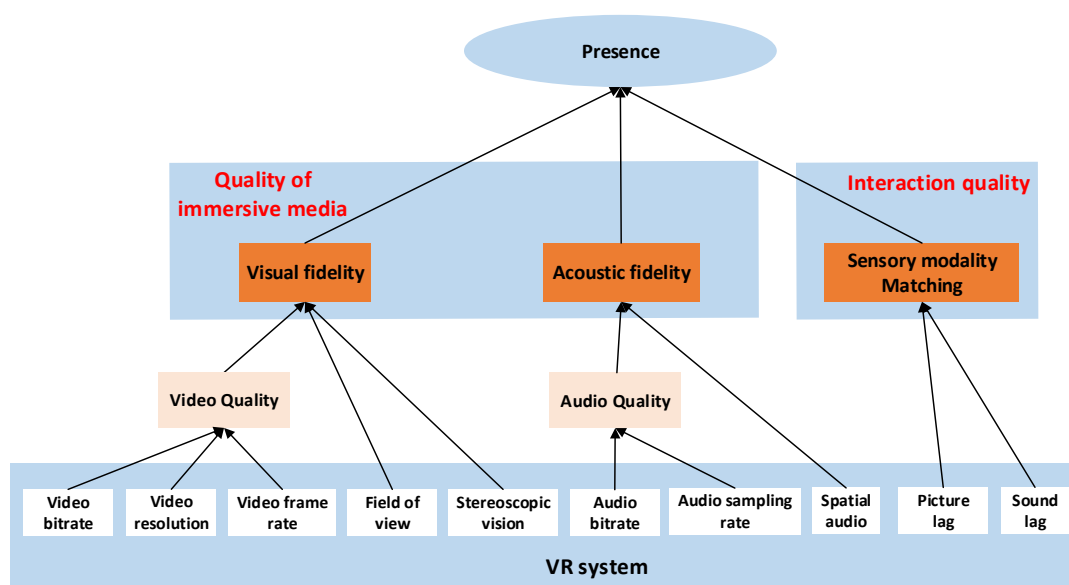


图 6.15 G. 360-VR 的框架

G. 360-VR 框架

- 源端条件
 - 视频性能
 - 应允许单视场和立体内容。
 - SRC 的质量也应尽可能相同。
 - 建议使用 4K 或更高分辨率的视频序列，以避免在 VR 显示器上放大的原始 VR 内容的低分辨率导致体验极度不佳的情况。
 - 音频性能
 - 可以使用立体声和空间音频。
 - 需要更多调查。
 - 互动性
 - 360 VR 应用是弱交互式 VR，用户在虚拟环境中被动地体验预先拍摄的内容。
 - 360 VR 的交互体验主要体现在 MTP 延迟上。
 - 持续时间
 - 使用持续时间为 10 秒至 5 分钟的序列。
 - 由于 10 秒至 20 秒非常短，因此建议使用如 M-ACR 的方法。

- 测试方法
 - 选项 1 - ACR
 - 选项 2 - M-ACR
 - 选项 3 - Double Stimulus Impairment Scale 测试法
 - 其他方法

M-ACR 和 DSIS 测试法流程如下所示：



图 6.16 M-ACR 法流程



图 6.16 DSIS 法流程

- 测试环境

测试时大致分为两种环境：

 - 受控环境
 - 公共环境

受控环境应代表一种不会分散人注意力的环境，在这种环境中，测试者可以合理地使用测试设备。该条件下，测试应在噪声隔离环境中进行，因为噪声是影响测试的主要环境因素。

公共环境应代表一种会让人分心的环境，该环境具有可能影响用户的噪声或其他因素。

2018 年 7 月，QUALINET 与 QoMEX 共同开展了年度会议，更新了沉浸式媒体的相关部署，特别是关于“沉浸式媒体 Qualinet-VQEG 联合团队（JQVIM）”和“沉浸式媒体体验（IMEx）”的任务。

此外，VQEG 启动了沉浸式媒体工作组（IMG，Immersive Media Group）。IMG 整理了来自各个研究团队的可用全景图像/视频集，如表 6.1、6.2 所示。

表 6.1 360 图像集

数据集/ 提供者	图片 数量	分辨率	格式	附加材料	（资料）注释	URL
Salient360 Université de Nantes / Technicolor	98	从 5376 × 2688 至 18332 × 9166	经纬图 投影	眼部/头部 跟踪数据	Check: Y. Rai, J. Gutiérrez, and P. Le Callet, “A dataset of head and eye movements for 360 degree images,” in Proceedings of the 8th ACM Multimedia Systems Conference	salient360 @univ-nant es.fr

					(MMSys), Jun. 2017.	
Stanford University, Universidad de Zaragoza, University of California Berkeley	22	8192 × 4096	立方体投影	眼部/头部跟踪数据	V. Sitzmann <i>et al.</i> , “Saliency in VR: How do people explore virtual environments?,” <i>IEEE Trans. Vis. Comput. Graph</i> , 2018.	https://vsitzmann.github.io/vr-saliency/
Berkeley V-Sense, Trinity College Dublin	22	4096x2048	经纬图投影		A. De Abreu, C. Ozcinar, and A. Smolic, “Look around you: Saliency maps for omnidirectional images in VR applications,” in <i>2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)</i> , 2017, pp. 1–6.	https://v-sense.scss.tcd.ie/?tag=saliency-maps
Nanjing University, Vision Lab, Immersive images	15	4096 × 2160			MOS	http://vision.nju.edu.cn/index.php/database/item/64-im-images
SUN360 MIT	67, 583 (来自英特网)	多种分辨率, 最高为: 9104x4552	经纬图投影		J. Xiao, K. A. Ehinger, A. Oliva, and A. Torralba, “Recognizing scene viewpoint using panoramic place representation,” <i>Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.</i> , pp. 2695–2702, 2012.	http://people.csail.mit.edu/jxiao/SUN360/main.html

Laboratory for Image & Video Engineering	10	4096×2160	经纬图投影		“To understand the joint impact of compression and spatial resolution”	http://live.ece.utexas.edu/research/quality/immersive_images/
Omnidirectional HDR consumer camera dataset, MMSPG EPFL	43	5476x2688	经纬图投影	速率、跟踪和切换信息	A.-F. Perrin, C. Bist, R. Cozot, and T. Ebrahimi, “Measuring quality of omnidirectional high dynamic range content,” <i>in Applications of Digital Image Processing XL</i> , 2017, p. 38.	https://mmsp.epfl.ch/360hdr-consumercamera

表 6.2 360 视频集

提供者	视频数量	分辨率/帧率	时长	格式	附加材料	(资料) 注释	URL
IMT Atlantique (from Youtube)	7	3840×2048 / 25, 29.97, 30, 60 fps	70 s	经纬图投影	头部数据	X. Corbillon <i>et al.</i> “360-Degree Video Head Movement Dataset,” <i>ACM MMSys’ 17</i> .	http://dash.ipv6.enstb.fr/headMovements/
National Tsing Hua University (from Youtube)	10	4K	1 min	经纬图投影	头部数据	W.-C. Lo <i>et al.</i> , “360 Video Viewing Dataset in Head-Mounted Virtual Reality” <i>ACM MMSys’ 17</i>	https://dl.acm.org/citation.cfm?id=3192927
Tsinghua University	18	多种分辨率: 4K, 2560x1440, ...	2-8 min	经纬图投影	头部数据	C. Wu <i>et al.</i> , “A Dataset for Exploring User Behaviors in VR Spherical Video Streaming”, <i>ACM</i>	https://wuchlei-thu.github.io/

						<i>MMSys' 17.</i>	
Stanford University - Virtual Human Interaction Lab (from Youtube)	73		29-668s		头部数据唤醒速率	B. Li, et al.. "A Public Database of Immersive VR Videos with Corresponding Ratings of Arousal, Valence, and Correlations between Head Movements and Self Report Measures", <i>Frontiers in Psychology</i> , Dec. 2017	http://vhil.stanford.edu/360-video-database/
Universidad de Zaragoza, Stanford University	216	4K	6+6 s (2 shots)	经纬图投影	眼部/头部跟踪数据	A. Serrano et al., "Movie editing and cognitive event segmentation in virtual reality video", <i>ACM TOG</i> 2017.	http://webdiis.unizar.es/~aserrano/projects/VR-cinematography
Nanjing University, Vision Lab,	2	8K					http://vision.nju.edu.cn/index.php/database/item/83-pa-videos
Inter-Digital	6	8K	60s (4x), 300s (2x)	经纬图投影		MPEG internal document but content license allows usage for research and publications but not for commercial use	

Wuhan University	48	从 2880×1440 至 7680×3840	从 20s 至 60s		头部数据	M. Xu, C. Li, Z. Wang, and Z. Chen, “Visual Quality Assessment of Panoramic Video,” pp. 1–12, Sep. 2017.	https://github.com/Archer-Tatsu/head-tracking
Technische Universität Ilmenau	20	4K	20s	经纬图投影	头部数据 + SSQ 结果	S. Fremerey et al., “AVTrack360: An open Dataset and Software recording people’s Head Rotations watching 360° Videos on an HMD.” , <i>ACM MMSys 2018</i> .	https://github.com/Telecommunication-Telemedia-Assessment/AVTrack360 SRCs: 请联系: stephan.fremerey@tu-ilmenau.de

而在 2018 年召开的 VQEG 会议上, IMG 同意:

- (重新) 开始评估沉浸式媒体质量的联合工作。
- 不仅关注“娱乐”360 视频的视听质量, 还将范围扩大到:
 - “带有任务的视频”
 - “具有互动性的视频(或 VR)”
 - 不仅仅评估“视频质量”

据此, IMG 对其工作范围和具体应用场景进行了初步的归纳和总结。应用场景为:

- 单向应用场景(360 视频)
 - 娱乐
 - 培训(面向任务)
- 双向应用场景(交互式)
 - 机械控制(人与机器)
 - 私人交流(人与人)

结合实际约束条件, 细化的应用实例如下表所示:

表 6.3 IMG 应用实例总结

应用实例		自由运动	语义导航	带有任务的评估	交互性
单向	娱乐	有	+/-	无	无
	培训	有	有	有	无
双向	机械控制	有	有	有	有限制的
	私人交流	有	有	有	自由交谈

而其工作内容包括:

- 创建源内容的参考数据库
- 表征/分类源内容
- 定义要衡量的质量因素以及具体做法（主观方法）
- 建立测试条件/视频损坏对于 QoE 影响的评估方法
- 生成上述 4 点主观结果的数据集。

对于质量评估，IMG 初步考虑从视听质量、呈现效果、晕眩程度、效率等方面进行考量。测试条件包括视频捕获、压缩与传输、终端显示、延迟等。QoE 则包括沉浸感，参与感，存在感等。

第 1 阶段：全向视频的视听质量

IMG 第 1 阶段的目标是在 VR360 的视听质量方面，以传统视频指标如 PSNR、VQM、VMAF 为参考，达到与传统视频相同的成熟度。具体目标有以下几点：

- 定义评估 VR360 视听质量的主观方法，以某种方式等同于 ITU-T P.910，ITU-R BT.500 等对传统视频质量的评价方法。
- 收集一组“要求严格但不过分严格”的参考序列。
- 基于一系列合理的视频损坏条件，共同创建一组主观评估序列，用作可重复研究的参考，以及客观指标的训练集。
- 提出客观指标来模拟 360VR 视频的视听质量。

IMG 要求参考序列持续时间必须为 60 秒（可能不是，但应为一较长的时间）。原始序列必须具备高质量：

- 采用中高端相机系统录制。
- 视频可以是：
 - 单视场序列。
 - 将左眼和右眼视图独立编码为文件的两个完整流。
 - 将左眼和右眼视图编码为单个数据帧的 Top-bottom 流。
 - 无缝拼接的全景视频。
 - 高分辨率的经纬图投影视频：8K，8K UHD，6K，4K，UHD，3K，HD。
 - 高帧率。用于测试的高质量视频应支持不同的帧速率。预期样本应高于每秒 30 帧：30 fps，60 fps，90 fps，200 fps。
 - 定向音频。可能为：四，五，七，八扬声器配置，高保真度立体声配置。
 - 未压缩或具有感知无损压缩：MPEG-2 TS，MP4，原始视频，OpenEXR 文件。
 - 具有与视频捕获相关的其他元数据信息：深度映射视频轨道，其他传感器信息。

而在序列组中，视频损坏条件应涵盖视频数据流的预期情况，同时这种损坏必须是各向同性的。这样，视频缺陷和用户行为之间的交互就将减少。

要考虑的损坏包括：

- 几何变化：投影，分辨率等。
- 切换到单视场视频。
- 视频压缩。
- 自适应流式传输的影响（如卡顿）。

最终视频并在 3 DoF 消费设备（手机）上播放，因为这是 VR360 视频较为常规也是预期的应用场景。

此阶段工作的最终目的是利用该评估方法检测来自不同实例的代表性内容：广告，电影，体育，新闻，纪录片，教育等。

第 2 阶段：沉浸式媒体的体验质量

第二阶段应侧重于建模和度量沉浸式媒体，以更好的方式来表征参与度，沉浸感，不适感和其他影响。

具体而言就是研发一种标准化的衡量方法，需要考虑以下方面：

- 基于任务的评估，可用于评估交互性，如 ITU-T P. 805 或 P. 920。
- 基于行为的评估，替代基于问卷的评估。
- 提供对测量内容的精确定义，以便能够为其研发方法。

为开展相关工作，IMG 已为沉浸式媒体质量评估提供了明确的框架和术语定义。其中，沉浸式媒体框架涉及到：

- 硬件组件（显示器，控制器等）
- 信号
- 交互水平
- 运动自由度
- 渲染自由度

部分术语定义如下表所示：

表 6.4 沉浸式媒体术语定义

Binocular Disparity	左右眼看到的图像位置的差异，或者显示器显示的左右视图上的差异。
Earcon	用于表示特定事件或传达其他信息的简短、独特的声音。（来源：维基百科）
Head-Motion Parallax	从不同位置或方向观察物体位置的位移或差异。
Motion-to-High-Quality Latency	头部运动和在头戴式设备中以高质量显示内容所花费的时间。
Overlay	基于 360 度视频内容的视觉媒体渲染。
Transparency	视觉媒体的一个或多个叠加，其中叠加区域在视觉上是同质的，并且叠加和 360 视频内容的可见度可以是任意选择的，并不一定是“二选一”。
Viewing Space	用于观看的 3D 空间，其中可以进行图像/视频渲染和 VR 体验。
Viewpoint	用户观看场景的点；它通常对应于摄像机位置。轻微的头部运动并不意味着不同的观点。
Visual Media	视频，图像和文字。

目前，IMG 已与多个实验室展开合作，相关实验室/联系人及其测试设备如下所示：

- 诺基亚贝尔实验室，PabloPérez
 - HMD：三星 GearVR（带 Galaxy S8 / S9），Oculus Rift，HTC Vive
 - 计算机：具有高处理能力的台式机/笔记本电脑（nVIDIA GPU 等）
 - 生理/行为测量设备：ECG

- COMLAB (Roma TRE 大学), Federica Battisti
 - HMD: HTC Vive, Microsoft Hololens
 - 计算机: 具有高处理能力的台式机/笔记本电脑 (nVIDIA GPU 等)
 - 生理/行为测量设备: 配备传感器的椅子 (正在开发)
- 分布式和互动系统 (DIS), CWI, Francesca De Simone
 - HMD: 2 Oculus Rift, 1 OSVR
 - 计算机: 每个 HMD 一台服务器
 - 采集设备: 8 个相机, 2 个 kinects
- IPI-LS2N (Université de Nantes), Jesús Gutiérrez
 - HMD: 2 个 HTC Vive, 1 个 Hololens, 1 个 Metavision
 - 计算机: 每个 HMD 一台服务器 (nVIDIA GPU 等)
 - 生理/行为测量设备: 眼动仪 (适用于 HTC Vive, Hololens 等), 脑电图, 心电图, GSR,
 - 采集设备: Ricoh Theta, Lytro Illum
- Vaader-IETR (INSA Rennes), Fang-Yi Chao
 - HMD: HTC Vive
 - 计算机: 用于 HMD 的服务器和具有 nVIDIA GPU 的台式机。
 - 生理/行为测量设备: HTC Vive 眼动仪

6.10 DASH-IF

DASH-IF 目前正在研究低延迟 DASH, 这与 Live VR 服务相关, 并且可能包括具有基于 DASH (VR) 分发的常规广播同步的内容。

6.11 CTA

2018 年 7 月, CTA 发布了 CTA-2069, 描述了增强和虚拟现实技术的定义和特征, 介绍了新兴消费级技术的各种术语。

在更早期, CTA 在 AR/VR 方面建立了第一个标准工作组。涉及内容包括:

- 虚拟现实 - 消费者体验和期望, 该研究调查了美国消费者的 VR 体验, 以深入了解消费者所接触的内容及其内容偏好。
- 消费者情感: VR 店内演示: VR HMD 及内容, 本研究将为视频内容的开发和分发策略以及 VR 产品的未来提供行业指导。
- 增强现实与虚拟现实: 消费者情感, 消费者反馈报告可以确定消费者对 AR 和 VR 技术及其各种用例的认知和感知。

6.12 SMPTE

SMPTE VR/AR 研究组于 2018 年 2 月 28 日成立, 旨在研究图像/声音捕获和发布的标准化方法的当前和未来需求, 以创建 VR/AR 的分发和显示系统。另一个目标是研究可能的标准应用场景。

具体来讲, VR 和 AR 内容有许多不同的捕获方法, 文件格式, 显示系统和后期制作方法。该小组要解决的问题是确定是否需要将这些方法标准化, 以实现更简单、容易的替换。小组目前正在研究用于生产和后期制作的 VR 和 AR 系统, 并创建一份报告, 记录当前系统, 相关现有标准和新标准的建议, 并对现有标准和所需标准进行差距分析。

6.13 ETSI

ETSI 于早期启动了增强现实新组织, 特别是增强现实行业规范组 (ARF ISG), “在工作初期阶段, ARF ISG 希望听取一些 AR 行业用例, 部署 (试点) AR 服务时遇到的障碍以及互操作性要求。”

6.14 SVA

流媒体视频联盟 (SVA) 于 2016 年底成立了 VR/360 度视频研究小组。他们目前的工作是记录 360 视频市场的相关技术和经验。除了评估传统视频服务的 CDN 性能外, SVA 还希望包含 VR360 内容, 以便了解延迟因素和 CDN 对 VR360 传输的影响。

小结

“沉浸式”媒体时代与 2D 时代相比, 其带来的信息量是巨大的。采集呈现, 存储与传输, 对于技术实现来说都是很大的挑战。国际标准组织在压缩编码方面还是会发挥很大的作用。OMAF 等标准已经基本实现 3 自由度的视频, 而 6 自由度的视频还需要更多发展的空间。对于 Immersive Media 来说, 提升体验感, 推动更多人用是当下最关键的。这一系列工作都需要标准工作来督促与推进。当下, 越来越多的组织和企业都加入到制定沉浸式媒体活动的标准中, 共同推进这项工作的创新与进步。

本章参考资料:

- [1]Christian Timmerer. Overview of standards activities related to immersive media (v2). ISO/IEC JTC1/SC29/WG11 MPEG2016/N17136. October 2017, Macau, CN.
- [2]<https://mp.weixin.qq.com/s/IPbXbgoFq8U5KgIk9zVNAA>
- [3]State of the Art, State of the Industry and State of the Standards, ACM Multimedia System Conference - Amsterdam, 12 June 2018
- [4]Virtual Reality (VR) media services over 3GPP. (Release 15). 3GPP TR 26.918 V15.0.0. 2017, 09.
- [5]Timmerer C. Immersive Media Delivery: Overview of Ongoing Standardization Activities[J]. 2017, 1(4):71-74.
- [6]<https://lists.aau.at/mailman/listinfo/mpeg-i-visual>
- [7][https://www.itu.int/ifa/t/2017/sg12/exchange/wp3/q13/G.360-VR/WD11Huawei-7-Proposed baseline for G.360-VR-v2.docx](https://www.itu.int/ifa/t/2017/sg12/exchange/wp3/q13/G.360-VR/WD11Huawei-7-Proposed%20baseline%20for%20G.360-VR-v2.docx)
- [8][https://www.itu.int/ifa/t/2017/sg12/exchange/wp3/q13/201802_Interim_meeting_Geneva/WD14- Restructured version of G.QoE-VR baseline.docx](https://www.itu.int/ifa/t/2017/sg12/exchange/wp3/q13/201802_Interim_meeting_Geneva/WD14-Restructured%20version%20of%20G.QoE-VR%20baseline.docx)
- [9]<https://www.its.bldrdoc.gov/vqeg/projects/immersive-media-group.aspx>
- [10]<https://docs.google.com/document/d/1xLxVeXYCegRHfPMWyiIoOELvR00pUsHGdLE6lWYhJOM/edit?usp=sharing>

[11]<https://drive.google.com/drive/folders/0B4K5KVGJNKEpOHd3cUZRRnppSDQ?usp=sharing>

[12]<https://www.its.bldrdoc.gov/vqeg/vqeg-home.aspx>