

第五章 全景视频 QoE 技术

5.1 主客观评价简介及主观评价的实施案例

沉浸式媒体应用与设备近年来的兴起，在一定程度上导致了 MPEG、3GPP、WebVR 以及其他相关领域标准化的推进。就目前的沉浸式媒体应用与设备而言，如何评价其体验质量并量化形成对比是十分必要的。就此，MPEG 已经出台了相应文件，大意上是要求沉浸式媒体添加与质量评估有关的额外输入以便进行对比和评价。

5.1.1 编码与数据流方案中的质量评估

如今，网络上每天产生的流媒体音频与视频数不胜数，占比也越来越大，因而编码与数据流在更多情况下是被绑定在一起的，且此趋势还会因 360 度视频等沉浸式媒体需要更多数据的情况而日益加深。

目前，沉浸式媒体内容的编码采用 HEVC 标准，因其可以降低通过 HTTP 协议传输动态自适应数据流时存储和带宽的要求，被认为是最先进最完善的编码方式。就此，诺基亚科技团队利用该标准在两种分辨率条件下储存同一全景视频，当向测试设备传输视频数据时，根据测试者的当前角度，部分 tile 通过高分辨率传输，剩余部分则利用低分辨率数据代替，类似方案在之前章节中也已提到过，其大致示意图如下所示。

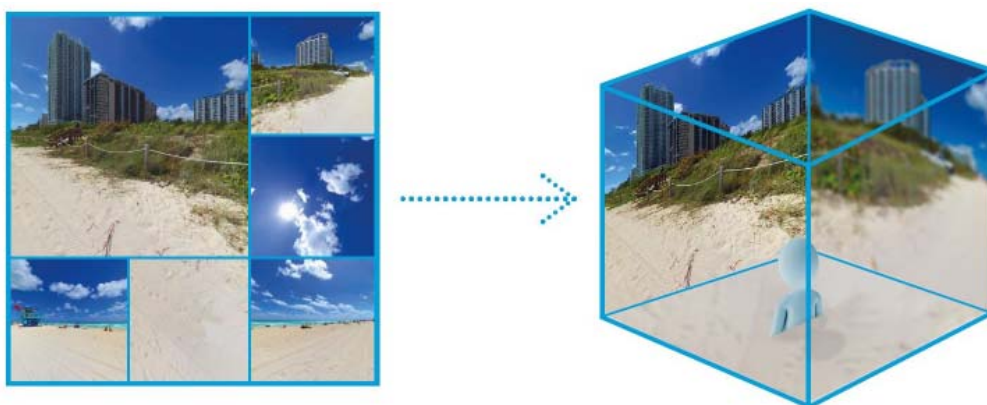


图 5.1 混合分辨率全景视频传输

同时为满足前述选择性传输的随机性，图像集中各 tile 均独立编码以便后续的解码，最终根据传输性能和压缩损失两方面的综合表现找到了一些列较优方案，并在这些可选方案中得到结论：相比于高分辨率传输所有视频内容，混合传输的方式可以降低 30%-40% 的比特率。此外，也有相关研究团队利用相同传输方式，对相同尺寸样式的图像块集在理想环境和实际环境下的脚本进行了详细的评估，包括比特率开销、带宽要求、峰值信噪比（PSNR）等多方面因素，部分结果如表 5.1 所示。实验结果与诺基亚团队所得结论相符合，并得到了更为全面的结论：对于实际的音频视频脚本，可以利用基于图像块的编码与数据流，节省至多 40% 的比特率；而对于理想脚本，在蜂窝网络中传输可降低近 80% 的比特率。

表 5.1 混合分辨率传输结果

Head Movements	Resolution	Tiling	BD-BR [%]	
			Tiles Monolithic	Tiles With Full Delivery Basic
User 1	1920x960	3x2	30.538	-9.008
User 1	1920x960	5x3	34.732	-35.427
User 1	1920x960	6x4	38.680	-35.433
User 1	1920x960	8x5	45.682	-35.360
User 1	3840x1920	6x4	25.874	-38.982
User 2	1920x960	3x2	30.779	-15.075
User 2	1920x960	5x3	34.513	-28.976
User 2	1920x960	6x4	38.501	-40.896
User 2	1920x960	8x5	45.748	-29.970
User 3	1920x960	3x2	31.042	-11.317
User 3	1920x960	5x3	34.926	-31.786
User 3	1920x960	6x4	38.884	-38.389
User 3	1920x960	8x5	46.439	-32.282

类似的 tile 集合也被用于提升媒体的交互性,以及媒体服务质量指标计算的多个方面。法国 IMT Atlantique 团队于 2017 年 5 月发表的文章中提出了一种视角自适应 360 度视频传输的解决方案,该方案需要服务器端提供同一内容视频的多种呈现方式,也就是备有不同质量、不同分辨率的多种图像集,而客户端设备根据用户视角向服务器请求合适的传输带宽,此过程如下图所示。

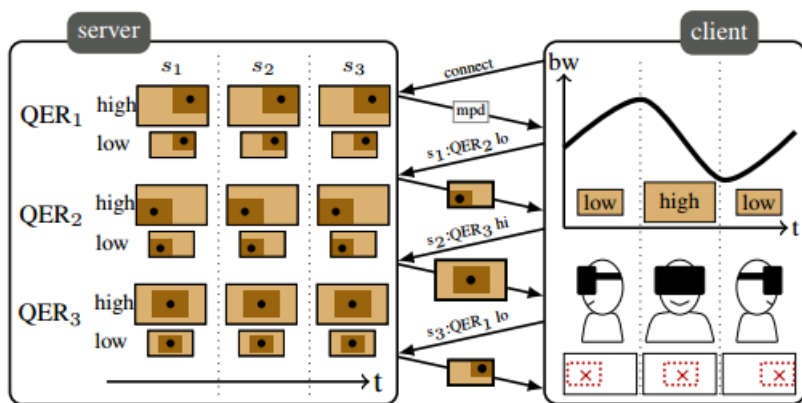


图 5.2 视角自适应的 360 度视频传输

同时,该团队通过 PSNR 和图像质量评估算法 MS-SSIM 得到采用不同图像质量分布策略的数据流和图像投影方式对于最终效果的影响。此外,也有文章提到在基于图像块的流传输时,通过可变化的 IDR 帧的呈现方式可以减少传输拥塞的现象,以提高传输质量。

德国弗劳恩霍夫应用研究促进协会近期的研究中提出了一项时空活跃性指标,目的在于快速、简洁地计算出基于人们感兴趣区域 (ROI) 的视频传输方案。都柏林圣三一学院团队则基于 HTTP 标准和视频观看者的视角对动态自适应的数据流进行分割、整合,产生良好的虚拟现实效果,该团队也利用了 PSNR 及 SSIM 计算并验证了该传输方案相比已有方案,更能切合用户的需求和期望。

除此之外,MPEG 文件中收录的与编译码、数据流相关的论文主要着重于比特率优化以及球面域率失真优化的改良。比特率优化主要是指因全景视频的球面峰值信噪比 (S-PSNR) 与典型 PSNR 计算方式的不同而作出的基于比特率层面的优化,使得传输字节合理分配至不同编码块时,仍保持着可观的 S-PSNR。沉浸式视频相比于传统视频而言,主要会在图像于二维/三维空间转换时产生失真,而此类失真会使得数据传输时的率失真优化过程在一定条件下的作用微乎其微。对于该问题,作出的改良是通过分析球面域失真对率失真优化的影响,

依此寻找出最优方案，下图为实现过程。

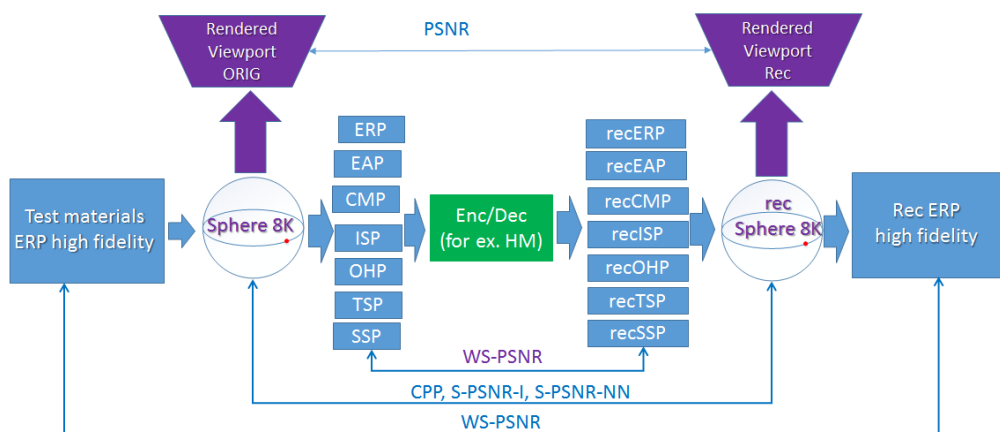


图 5.3 球面域率失真优化的改良过程

5.1.2 客观评价

上节众多方案大多采用的是客观质量评价，然而，许多度量沉浸式媒体质量的客观指标采用了衡量传统媒体的指标，或是略微修改指标的定义以满足 360 度图像或视频的特性，例如自适应性和视角认知度指标。但这样的做法目前存在的问题是，这类指标呈现出的媒体质量往往与人们的主观感受不匹配。就此，有研究人员认为，客观意义上的指标虽不能与人类体验达到完全一致，但仍会存在一定的规律。基于此，其团队利用全景图像样本，得到了大量的主观评分和客观结果，挖掘其中的统计规律，试图让单个客观指标具有类似于人的主观能动性。采用类似方法的基于视频样本的客观指标研究也已存在。当然，对于图像和视频两种表现形式而言，同种指标的计算也可能存在区别，例如图像与视频 PSNR 的计算，前者还需利用 SSIM、VIFP 等算法才能得到严谨的结果。

如上述对于特定指标的研究已有许多，然而目前还缺乏对包含多种编码传输方式，需计算多项指标的大范围样本的研究或评价。

5.1.3 主观评价

相比而言，对于沉浸式媒体主观指标的研究还不是很多。就目前来讲，360 度视频卡顿指标的计算已有研究，且得出的计算方式也可用于传统媒体，如电视、电脑、手机，其中电视与 VR 在相同条件下的卡顿程度对比如下图所示。同时该研究也对之后的卡顿研究提出了诸多建议，是相关领域中首次对头戴式设备如何预知卡顿的问题提出的见解。

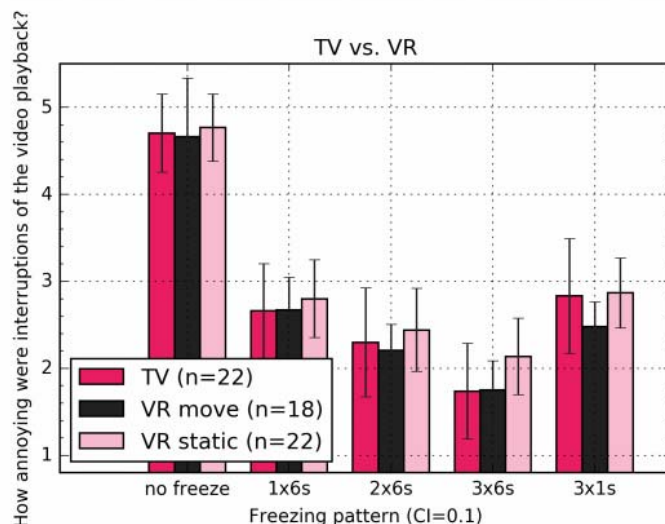


图 5.4 电视与 VR 的卡顿指标对比

另有一项关于 VR 视觉效果评价的研究通过主观测试以及秩相关系数检验 (SRCC) 来衡量 VR 内容质量以及各方向物体的一致性, 并综合两方面表现提出了两种质量检测的主观指标: 0-DMOS 和 V-DMOS。

此外, 还有根据主观指标降低视频传输带宽的研究, 而此项研究正是基于 DMOS 指标的基础上进行的, 最终得到在 DMOS 值为 4.5 时, 不同类型的视频传输平均可以降低 44% 的比特率, 从结果上优化带宽的效果已比较可观。

实际上, 在进行主观性能测试时, 我们都假设测试者位于一个特定的环境中, 例如一个常见的 VR 测试环境是让头戴设备的体验者坐在转椅上自由操作, 或者规定其做一些特定的动作, 然而即便如此, 体验者们的感受仍可能迥乎不同, 对于测试结果有着很大的影响。此外, 当屏幕上放映内容, 人们沉浸于其中时, 相关测试设备也需统一。如果测试只关乎某项主观指标的问题, 也只需做到统一就足够了, 但是体验者在沉浸感、交互性以及其他方面的感受就会变差。

接下来, 将对主观评价的具体实施过程进行介绍。

5.1.4 主观评价的实施案例

由于人类的生物学因素和 HMD 的科技限制, 360° 内容中每一时刻只有一小部分可以被测试主体观看和评估。观看全景的 VR 视频时, 人们可以自由地观看 360° 的空间并且可以切换观看方向。这时产生的新的挑战就是如何确定来自同一测试主体或不同测试主体采用不同的评价方式得出的结果是可比的。换句话说, 就是如何去确定测试主体是不是在同一时间对视频的同一个部分做出了评价。

已经有一些论文提出了对高分辨率或全景图像的评估方法。而接下来将介绍一种在数据流系统中主观评价全景视频的测试方法实施案例。

这一方法的基础来自现有的用于 2D 和 3D 视频主观质量评价的 ITU (International Telecommunications Union) 标准。但这一标准并不直接适用于 VR 视频内容, 需要做一些扩展和调整。

主观视频评价实验应包括以下部分: 1) 测试对象知情同意和个人数据收集; 2) 预先筛选测试对象; 3) 指示说明; 4) 预测试问卷; 5) 测试对象培训; 6) 评分部分; 7) 测试后问卷。

5.1.4.1 测试环境及对象

与通过平板显示器观看的传统视频不同，全景 VR 视频的典型观看环境需要 HMD 设备。要求测试对象在全方位空间中自由环顾四周，以评估场景的所有细节。为了避免在测试操作期间测试者被连接线缠绕，应优选无线 HMD 设备。与 2D 视频评价相比，全景视频评价任务具有更高的复杂性，不仅因为显示设备不同，还因为要求测试对象能自由环顾四周。因此，要尽量避免测试对象参与其他操作。评分方式最好采用口述的，比如向测试管理人员口述，由他将评分输入计算机用户界面（UI）。让测试管理人员承担额外工作的作法的原因主要是担心使用自动化的 HMD UI 进行评分可能会导致测试对象的额外疲劳和/或输入错误的状况。为了便于移动，测试对象应该坐在旋转的椅子上或使用站立姿势。当对象头部方向移动时，视角改变并且当前视角的新内容显示给测试对象。

根据所执行测试的类型（初步测试或完整测试），可以组织专家测试对象或非专家测试对象的实验。专家测试对象应为在视频质量或相关领域工作的人员。受试者被筛选后，在初步测试中，建议最低专家数量为 5 人。在非专家测试中，测试对象不应在视频质量或相关领域工作，并且未在至少六个月内参与任何主观视频质量评估的活动。受试者被筛选后，完整测试中，建议最少测试者数量为 15 人。如果测试结果中需要额外的统计可靠性，则建议最少数量为 28 个测试者（筛选后）。

由于测试活动可能产生健康方面的副作用，因此应进行入选前的健康认证。具有以下情况的测试对象应被排除在测试之外：怀孕，癫痫，其他神经系统疾病，强烈（过敏性）流感，强烈宿醉，特殊的短暂睡眠症，严重的一般眩晕症或戴着 HMD 的眩晕症。还应排除需要佩戴非常规 HMD 设备的测试对象。接下来，必须对招募的测试对象进行视力筛查，测试对象应该具有：1）正常远视力和近视力（例如，使用 Snellen 图表或 E 图表测试）；2）正常的色觉（例如，使用 Ishihara 板测试）；3）正常的对比度感知；4）正常的立体视觉（例如，使用随机立体点测试）。如果上述任何一项测试中显示有感知损伤，可以建议他/她不要进行评估实验。如果这一测试对象进行了实验，则应排除相关结果。此外，应记录受试者的瞳孔间距（IPD），因为 IPD 较短的受试者可能会出现更多的视觉不适。一些 HMD 具有补偿 IPD 差异的调整设置，可以使用这一设置进行调整。

5.1.4.2 评价方式

一种选定的评价全景视频的方式是单激励的 ACR-HR，这也是 ITU-T 推荐的方式。利用这种方法，视频测试序列（即已经应用了各种条件的源序列）将逐个地呈现给测试对象，并且在各个类别、等级上独立评分。每个测试序列的参考版本（即最高质量条件下的源序列）将作为其他测试视频序列显示给测试对象并且没有任何特殊标识（隐藏参考条件）。这就要求参考序列需要由视频专家评定为好或优秀。

测试过程中，主要向测试对象收集三种不同类型的评价结果：1）视频质量分数；2）视觉舒适度分数；3）测试后的视觉疲劳程度。

视频质量分数的评价方式与普通视频评价测试一样，但如 5.1.4.1 中所述，分数收集是由特殊的口述报告完成的。视觉舒适度指的是观看立体图像时是否舒适的主观感受。当全景视频无法正确地捕获和/或显示时，测试者给出的结果将是一定程度上的不舒适。要注意的是，在独立的评价中，必须使用用于收集视频质量分数的相同测试序列来收集视觉舒适度分数。同时，一般需要在每个测试对象实验结束时填写测试后视觉疲劳度的问卷，以收集有关可能副作用的有价值信息。

对于前两个分数，使用 5 级量表（具有形容词类别判断）来收集结果。视觉质量：1-非常不好，2-不好，3-一般，4-好，5-非常好。视觉舒适度：1-非常不舒服，2-不舒服，3-中等舒适，4-舒适，5-非常舒适。尽管 ITU 建议在需要更大区分度的情况下使用超过五个级

别（例如，对于低比特率序列）来收集结果，但更新的标准中还未要求增加级别数，因为 MOS 的准确性不会增加。如果需要更大的区分度，我们建议使用带小数的 5 级量表。

5.1.4.3 主观评价过程

视频质量评价实验的最后三部分（指示说明，评分和测试后问卷）构成了主观评价过程。如图 5.5 所示。推荐的实验总时间不应超过 1 小时，其中佩戴 HMD 设备的测试时间应为 25-35 分钟。这个时间还包括 2-5 分钟的休息时间，在此期间测试对象可以移除 HMD 设备以放松。对于 180° 的视频内容，每个视频序列的建议长度为 20 秒，对于 360° 的视频内容，每个视频序列的推荐长度为 30 秒。

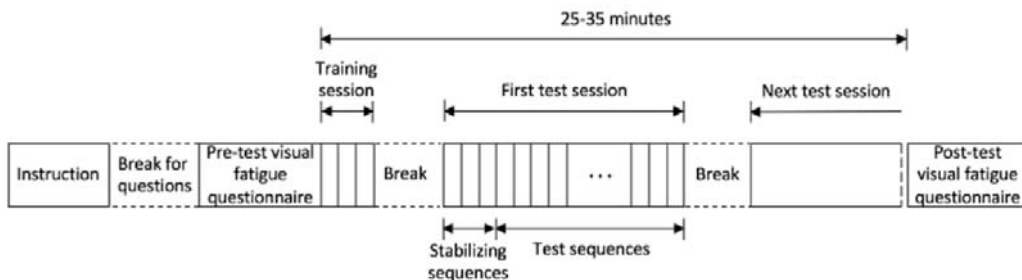


图 5.5 主观评价过程

指示说明

以下是从相关的主观评价标准规范中提取得到，并将其扩展为支持全景视频数据流评价的一组说明：

必须告知受试者可能由测试活动引起的负面副作用。必须进一步告知他们可以随时拒绝或中止他们的测试且不会受到处罚，并且如果他们出现以下任何症状，应停止测试：眼睛疲劳，注意力不集中，一般疲劳，头痛，身体功能恶化，过度紧张，晕动症。要注意的是，如果这些症状在休息后消失，可以继续进行测试。

测试对象不会被告知测试中出现的视频损坏或损坏的位置类型。而以下内容应该告知测试对象：

- 正在测试的用例场景；
- 评估的目标；
- 整体评估程序；
- 每个测试部分中的任务。受试者必须环顾 360 度空间并探索（可选地，同时遵循特定的运动模式），避免停留在静止位置；
- 评估方法（他们将要看到什么，他们需要评估什么，例如，质量，绝对质量，视觉舒适度等方面的差异；）；
- 评分方式以及如何评分；
- 测试序列的数量和类型及其总持续时间；
- 不要将播放内容创建的程序、过程视为视频质量的事故（例如，SW 崩溃，偶尔闪烁，缝合伪像等）。这些不会影响评价的质量排名；
- 播放事故，例如长时间的加载和缓冲，它们不是测试评价的目标应当忽略；
- 不要让视频序列的审美程度和主题影响评分排名（例如，给定的视频包含无聊的内容）。

此外，在测试期间看到的最差质量不对应于评分量表上的最低主观等级。最后，要求测试对象填写测试前视觉疲劳度问卷，以引起他们对某些因素影响程度的注意：近视困难，夜视困难，双视，聚焦，在某个时刻的视觉质量，眼睛干燥或湿润，头后部疼痛，颈部僵硬，

流感或宿醉或无法睡眠，眩晕（例如，观看 VR 视频时乘坐汽车船舶，乘坐游乐项目的情况）。

测试对象培训

对测试对象进行培训的目的是让受试者：1）熟悉评分程序；2）列出现有和待评估的全部范围（包括隐藏参考条件）的损伤类型，使他们的评分标准化；3）鼓励提出有关实验的问题。这也是让受试者习惯于在全景环境中观看四周的机会（通过根据诸如下图中的特定模式环顾四周）。

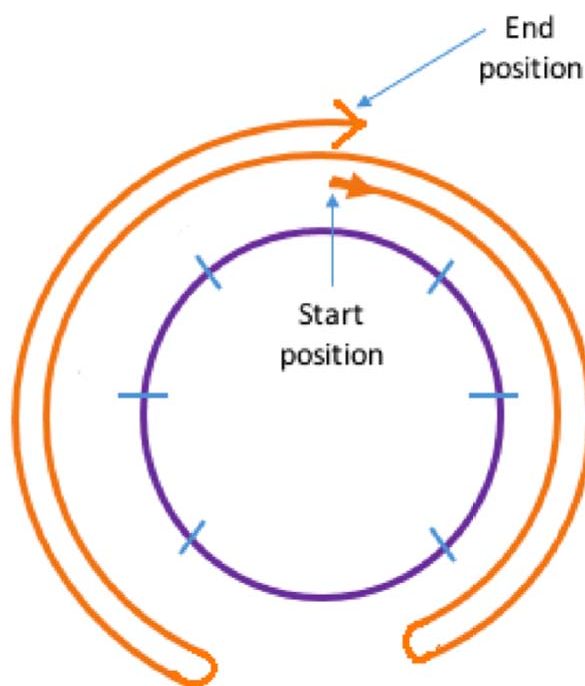


图 5.6 测试对象移动方式

我们建议使用至少五个视频序列进行培训，这些序列不会出现在实验中，但质量和持续时间与测试序列相当。这些序列的质量应涵盖实验中使用的所有测试序列的质量范围。在最终结果处理中不考虑培训测试中给出的评分。

测试方式

我们建议每次测试不应使用超过 20 个视频序列（包括稳定序列）。在每个测试开始之前，必须调整 HMD 设备以匹配测试对象的 IPD。可以通过观看测试静止图片来完成第一次调整。应记录下 HMD IPD 的值。

测试序列名称（例如“剪辑编号 1”）将使用扬声器播放和/或在 HMD 中可视地呈现，例如作为灰色矩形图案上的文本呈现。这样的矩形应固定在 HMD 屏幕的中间，对应测试对象在前一测试序列结束时的水平头部方向。这样能使测试对象在观看下一个测试序列之前无需返回到特定方向，能极大减少受试者的疲劳程度和眩晕症状。每一段测试序列播放结束后，提词屏幕上会呈现“请评价这一剪辑”的文字，持续大约 10 秒，在这期间测试对象应该用口述的方式给这一序列评分。

测试前及测试后的视觉疲劳度调查

该调查的目的是核对实验前后可能出现的副作用及其严重程度。对于全景视频的观看，已有的调查问卷应进行适当调整，并建议调查是否存在颈部僵硬或肩部疼痛，头部后部或前部疼痛或太阳穴疼痛，头晕，注意力不集中，恶心，困倦，眼睛刺痛，眼睛疲劳，聚焦困难，

眼睛干燥或湿润，感觉眼睛朝不同方向看，双重视觉的症状，并调查在实验过程中受试者是否必须闭眼才能重建清晰的场景。

5.1.4.4 评价结果分析

该过程需要计算受试者们共同评价同一个测试序列的平均评价得分（MOS）。在本方法中，评分结果衡量了相同序列的两个版本（即参考和测试序列）之间的质量变化。因此，要在每个测试序列与其相应的参考序列之间计算 MOS 之差（DMOS）。报告的结果还应该包括 DMOS 的标准偏差和置信区间。利用这种方法可以从主观分数评价中去除参考序列的影响。

用于分析的一个基本数据集是测试对象的视角朝向数据。必须以足够的速率对空间中的头部朝向进行采样，以解决头部快速移动的问题。这种头部朝向由头部左右移动、上下移动和旋转的度数表示。该数据不仅用于验证同一测试主体是否在不同测试条件下评价 360° 视频序列的同一个部分，也用来确认所有测试主体的测试公平性。这里提出了两种方法来为计算头部运动模式的统计相似性创建输入数据：

- 1) 基于方向的头部运动模式（OHM）：在每个测试序列内以特定间隔对测试对象头部方向（左右、上下、旋转）进行采样；
- 2) 基于视角的头部运动模式（VHM）：在每个测试序列内以特定间隔记录被测试主体正在观看的视角。

由于全景视频内容具有比人类 FoV 宽得多的 FoV，因而如何公平地评价全景视频内容成为了一个难题。更具体地说，应该有一种方式来衡量：

- 1) 测试对象是否一直在观看所有视频序列（在同一时刻）的相同部分。（对象内相关性，即同一对象观看之间的相关性）
- 2) 多个测试对象是否一直在观看相同视频序列测试条件（在同一时刻）的相同部分。（对象之间的相关性，即不同对象观看之间的相关性）

相似环度量（SRM）则是一种目前可以在全景视频主观评价过程中使用的方法。其用于衡量不同测试用例之间观看模式的相似程度。它适用于单个测试对象和多个测试对象。该指标使用 OHM 输入方法收集的数据。基本思想是在 XY 图上绘制重叠曲线，其中 X 轴表示时间（视频序列持续时间），Y 轴表示头部方向。基于测试类型，每条曲线可以表示由单个测试对象观看的不同视频序列（相同的测试对象，不同的测试条件），或者由不同测试对象观看的相同的测试序列（相同的测试条件）。曲线（观看模式）重叠的程度代表了观看模式的相似性。

5.1.5 VQEG 的主观评价实例

VQEG 于 2018 年完成了一项主观评价调查，其采用了如下的 VR 体验调查问卷和模拟器副作用调查问卷：

How would you rate the picture quality? (circle the verbal option)

Bad Poor Fair Good Excellent

How would you rate the responsiveness of the system? (circle the verbal option)

Bad Poor Fair Good Excellent

How would you rate your ability to accomplish your task of loading the logs on the truck? (circle the verbal option)

Bad Poor Fair Good Excellent

How would you rate the immersion of the experience? (circle the verbal option)

Bad Poor Fair Good Excellent

How would you rate your overall experience? (circle the verbal option)

Bad Poor Fair Good Excellent

图 5.7 VR 体验调查问卷

SIMULATOR SICKNESS QUESTIONNAIRE
Kennedy, Lane, Berbaum, & Lilienthal (1993)***

Instructions : Circle how much each symptom below is affecting you right now.

1. General discomfort	<u>None</u>	<u>Slight</u>	<u>Moderate</u>	<u>Severe</u>
2. Fatigue	<u>None</u>	<u>Slight</u>	<u>Moderate</u>	<u>Severe</u>
3. Headache	<u>None</u>	<u>Slight</u>	<u>Moderate</u>	<u>Severe</u>
4. Eye strain	<u>None</u>	<u>Slight</u>	<u>Moderate</u>	<u>Severe</u>
5. Difficulty focusing	<u>None</u>	<u>Slight</u>	<u>Moderate</u>	<u>Severe</u>
6. Salivation increasing	<u>None</u>	<u>Slight</u>	<u>Moderate</u>	<u>Severe</u>
7. Sweating	<u>None</u>	<u>Slight</u>	<u>Moderate</u>	<u>Severe</u>
8. Nausea	<u>None</u>	<u>Slight</u>	<u>Moderate</u>	<u>Severe</u>
9. Difficulty concentrating	<u>None</u>	<u>Slight</u>	<u>Moderate</u>	<u>Severe</u>
10. « Fullness of the Head »	<u>None</u>	<u>Slight</u>	<u>Moderate</u>	<u>Severe</u>
11. Blurred vision	<u>None</u>	<u>Slight</u>	<u>Moderate</u>	<u>Severe</u>
12. Dizziness with eyes open	<u>None</u>	<u>Slight</u>	<u>Moderate</u>	<u>Severe</u>
13. Dizziness with eyes closed	<u>None</u>	<u>Slight</u>	<u>Moderate</u>	<u>Severe</u>
14. *Vertigo	<u>None</u>	<u>Slight</u>	<u>Moderate</u>	<u>Severe</u>
15. **Stomach awareness	<u>None</u>	<u>Slight</u>	<u>Moderate</u>	<u>Severe</u>
16. Burping	<u>None</u>	<u>Slight</u>	<u>Moderate</u>	<u>Severe</u>

图 5.8 模拟器副作用调查问卷

以下是根据 17 名测试者得到的 5 项 VR 体验质量的 MOS 值：

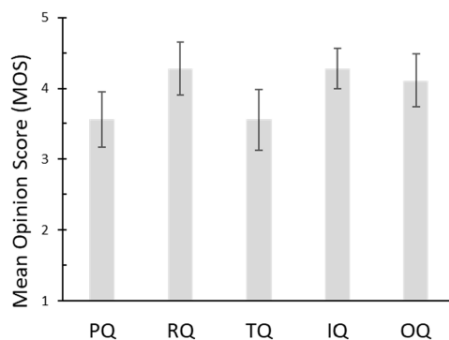


图 5.9 视频体验调查结果

除了基础调查外，VQEG 通过调整屏幕和操作杆配置，设置了 10 种不同的延迟情况。测试结果（图 5.10）显示，总体体验质量大致上是随着延迟的升高而降低的。但对于某些细化质量，如画面质量、任务完成度等，低延迟并不意味着数值的上升。

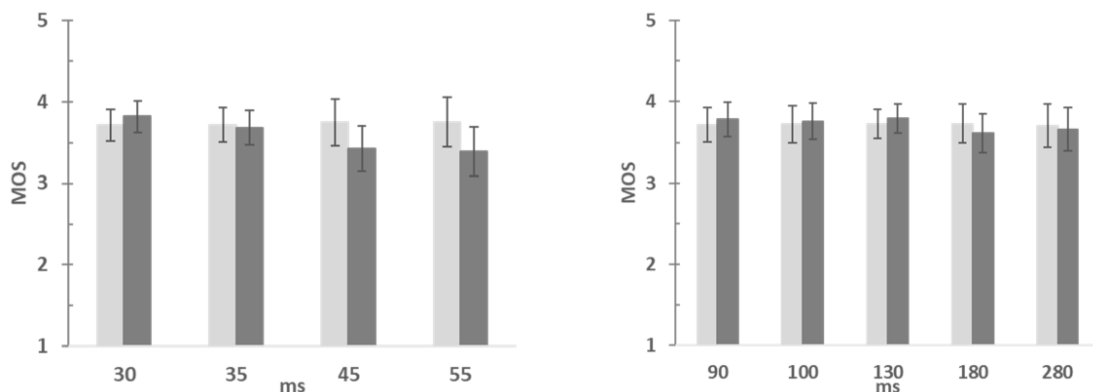


图 5.9 延迟调查-总体质量评分

5.2 典型的 QoE 要素

每个新技术的出现以及它的发展都会遇到很多问题，VR 在发展的道路上也会遇到许多瓶颈，这些瓶颈又会或直接或间接的影响用户的体验。就目前来说，不论是在 Oculus 还是 DayDream 平台，越来越多的消费者将使用 VR 设备的时间用在了观影上。这个数据也代表了用户的真实核心诉求，因为 VR 游戏从内容本身的体验上，还不足以满足大多数人们的要求，或许等到将来 VR 产业链和技术储备都成熟后，这种情况可以得到较大的完善。

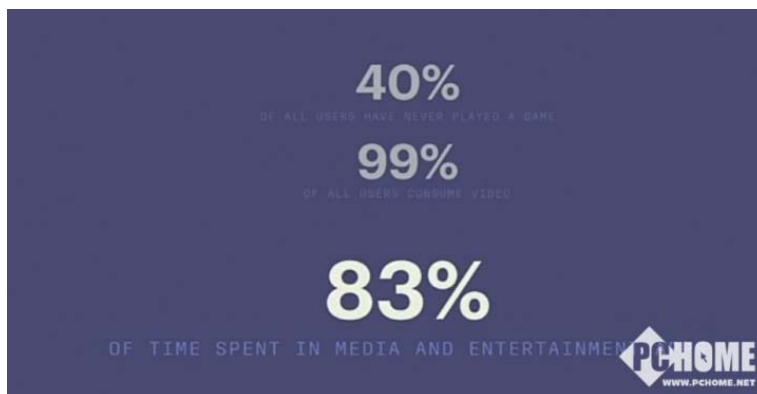


图 5.10 用户分布（来自 2018 Facebook8）

下面就介绍一些在 VR 发展过程中影响用户体验的要素和瓶颈。

5.2.1 视觉保真度

视觉保真度即是用户在使用设备观看内容时，所呈现的内容与原本的内容的相似程度。造成这种差别的根本原因就是数字设备呈现的画面是许多像素拼接而成的，这就必然会造成一定的失真情况。视觉保真度中有三个关键要素：纱窗效应，Mura(即亮度/颜色不均匀)和混叠。

纱窗效应

在许多初代的 VR 头显中，如 Rift 和 Vive 等，纱窗效应可能是最为明显的伪影。从技术来说，纱窗效应是“低填充系数”的结果，而“纱窗”的命名是因为这看起来就像是透过纱窗这样的精细网格来感知影像的。

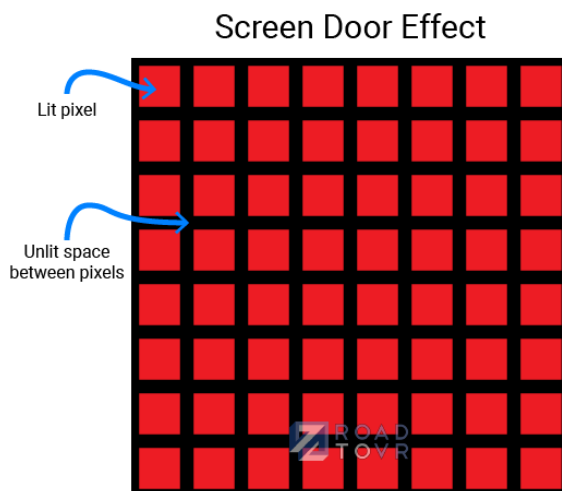


图 5.11 纱窗效应

像素是而单独点亮的元素，它们通过阵列排放以建立显示。由于各种原因，像素有时难以紧密地打包在一起，而这导致它们之间的间隙没有点亮。显示屏的“填充系数”描述了实际点亮面积与非点亮面积之间的比例。在低填充系数的显示器上，使用者容易感知像素之间的非点亮空间，从而导致纱窗效应的出现。

此外，造成这种现象的原因还有 VR 眼镜的低分辨率，而且 VR 眼镜中是有放大镜的，这就更加容易导致纱窗效应的出现，就像本来是一张光滑的白纸，但通过放大镜来观察的话也会出现坑坑洼洼的现象。总的来说，纱窗效应是在像素不足的情况下，实时渲染引发的细线条舞动、高对比度边缘出现分离式闪烁的现象。

要想真正地消除纱窗效应，屏幕所呈现的内容一定需要有很高的分辨率。

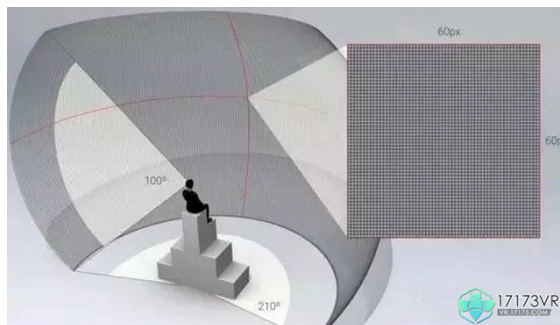


图 5.12 人的视场角

一般来说，每个人的视场角是水平 210 度左右，垂直 100 度左右，而要想看不到像素颗

粒感，视场角中的每一度要看到 60 个像素，即水平和垂直两个方向为 60*60 个像素(低于 60 个像素人类的视网膜就能分辨出像素粒)。所以水平的 210 度里就有 12600 个像素，垂直的 100 度里有 6000 个像素，理想情况下需要 12600*6000 的画面才能真正消除纱窗效应。但是即使 4K 的分辨率也才只有 4096*2160，在移动端都很难实现支持，距离理想情况有很大差距。因而结合目前沉浸式媒体的发展情况，现在需要的是硬件设备上的技术突破，而且手机端的计算处理能力也必须跟得上才有可能彻底消除纱窗效应的影响。

Mura(即亮度/颜色不均匀)

由于设备等原因，即使是由单一颜色值组成的一帧，从计算机输出到显示器时，也很难实现每个像素显示完全相同的颜色。Mura 便是像素之间颜色和亮度不均匀的结果。如下图所示，显示器的所有像素都设定为一个颜色值，但由于设计和制作中的不完善，设备能力不足等原因，实际的颜色输出并不一致。

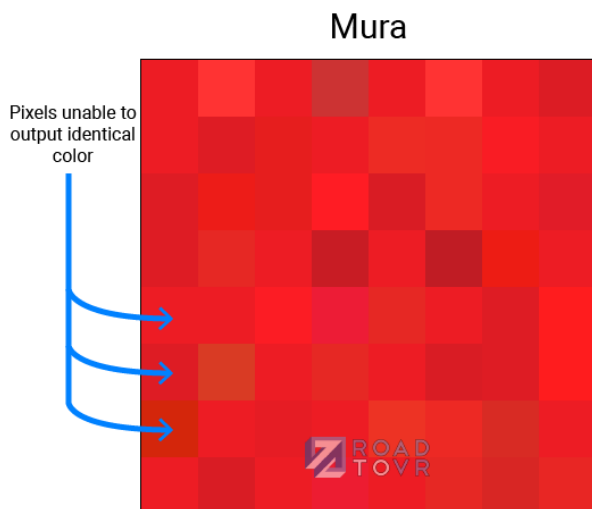


图 5.13 Mura(亮度/颜色不均匀)

不同显示屏的显示技术在此特性上有所差别，例如 LCD 在减少 Mura 方面往往表现得相当不错，而 OLED 表现就较差一些，需要仔细校准才能够产生良好的性能。

混叠

由于显示器由排列在网格中的方形像素（通常是这样）组成，因此很容易显示与像素网格行对齐的水平线条和垂直线条。

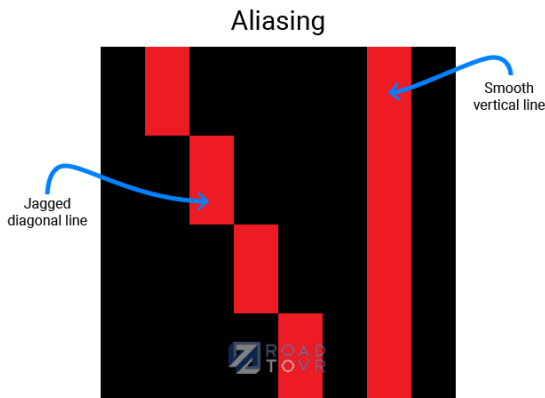


图 5.14 混叠

但在显示对角线或曲线时，就会产生问题：由于显示器只能用沿网格放置的方块来绘制

曲线，这样会导致非完全垂直或者非完全水平的边上出现锯齿。这意味着只有直线才能利用像素和像素网格的基本形状自然地显示出来。但是通过增加显示器的像素密度可以很直接地减少混叠的影响，显示器分辨率的提升能够更好地支持像素更精确地拟合出所需呈现的曲线。当然还可以通过抗锯齿的相关技术较少混叠的影响。

5.2.2 晕眩延迟

目前还有一个影响 VR 体验的大问题，就是晕眩感太强。一般来说，产生 VR 晕眩的根本原因是大脑对视觉和运动的认知不同步，即眼睛看到的内容(画面)与耳朵接收到的信息(位置)不匹配，导致脑负担加大，从而会产生晕眩感。VR 晕眩一般分为硬件和软件引起的晕眩。

硬件晕眩

VR 硬件中，主要有 5 种器件因其性能等原因会导致晕眩：GPU、感应器、显示屏、芯片成像透镜，以及瞳距和距离调整结构。一般来说，硬件晕眩是造成晕眩的主要原因。想要解决硬件晕眩问题，最简单的方法就是使用目前最好的硬件。但就目前来说，好的产品价格仍旧很贵，无法满足一般消费者的要求，所以现在需要做的，一方面是降低硬件的成本，另一方面还需要产业链的供给。

软件晕眩

VR 的软件晕眩一般有以下几大原因：

1) 游戏内容

很多 VR 游戏本身的内容就会产生晕眩。比如在坐“VR 过山车”时，视觉上正处于画面中的状态，在做剧烈的高速运动，但是前庭系统却并没有感知到运动状态，这时就会导致头晕，产生晕眩感。

2) 画面与现实世界的差异

VR 显示器给出的画面变形严重、视野小，这些都与现实世界存在差异，时间长了就会感觉头晕。

3) 画面滞后于动作

VR 中的 MTP 延迟造成时间上的不同步，当人转动视角或是移动的时候，画面呈现的速度会有些慢跟不上头部的运动。在 VR 这样全视角的屏幕中，延迟也是造成晕眩的关键因素。而且现在许多的 VR 设备均是基于 Android 系统，但传统的 Android 系统都是采用了双显示缓冲或者三显示缓冲，这个机制让 VR 应用无法知道制定的图像什么时候能够显示在头盔屏幕上，这就会造成更大的延迟。

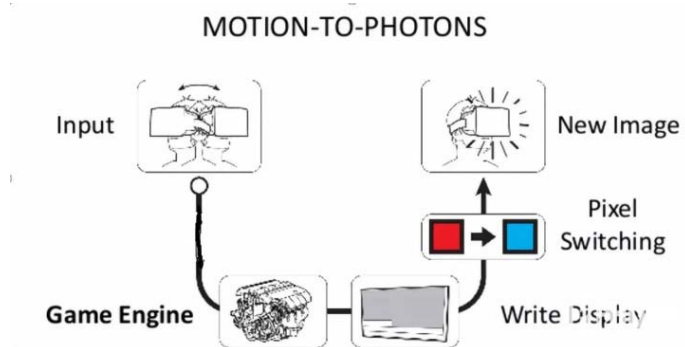


图 5.15 MTP 延时

4) 瞳距不一

由于每个人的瞳距不一，对某些人来说，人眼瞳孔中心、透镜中心、画面中心三点并非

一线，从而出现重影现象，看久了也会非常容易头晕。

5) 景深不同步

景深不同步，也是眩晕的原因之一。比如说，在你的面前，有一张桌子，在桌子上，近处放了一个杯子，远处放了一个玩偶。你看着近处的杯子，按理来说远处的玩偶应该模糊不清，但是现在，远处的玩偶也看的非常清晰。这是因为 VR 设备对现实的模拟还不够真实，无法真正地欺骗到大脑，受到困扰的大脑不堪重负，才会造成眩晕。

所以，如何解决 VR 晕眩问题呢？下面将介绍一些能够解决晕眩问题的方法。

低延迟技术

这里所谓延迟指的是从头部转动到画面转动的延迟。画面延迟在很大程度上又取决于显示屏的刷新率。目前世界上最先进的虚拟现实设备刷新率在 75Hz 左右。研究表明，头部运动和视野的延迟不能超过 20ms，不然就会出现眩晕。20ms 的延迟时间对于 VR 头显而言是一个非常大的挑战。首先设备需要足够精确的办法来测定头部转动的速度、角度和距离，这可以使用惯性陀螺仪（反应灵敏但是精度差）或者光学方法来实现。然后计算机需要及时渲染出画面，显示器也需要及时地显示出画面，这一切都需要在 20ms 以内完成。相应的，如果每一帧显示的时间距离上一帧超过 20ms，那么人眼同样也会感到延迟。所以，VR 头显的画面刷新率应该超过 50FPS，目前来说 60FPS 是一个基准。但是要想做到更好的效果，这个刷新率还应该接着往上提高，比如目前 Oculus Rift CV1 和 HTC Vive 采用了 90Hz 刷新率，而 Sony Project Morpheus 采用的是 120Hz 刷新率。

一般说来，延迟产生的原因很多，在之前也提到过，首先从头部转动到传感器读到数据，然后数据需要经由单片机通过 USB 线传输到电脑，在硬件上传输完成后，就是软件算法处理过程，在将模拟信号转换成数字信号之后，数据中存在大量的噪声和漂移，于是需要复杂的数字信号处理方法将这些噪声和漂移过滤掉，接下来是运用 Time-warp 算法渲染场景，在渲染完后还需要做反畸变和反色散等处理，最后还需要的时间延迟就是传输图像到显示器的时间，不过现在的 OLED 技术已经将这个时间较少到了微秒级别。延迟中还涉及到 CPU 的性能、USB 丢包等问题。

目前看来，从 VR 硬件方面降低延迟是改善眩晕最好的方法，所有需要的处理时间不断压缩便能够降低画面的延迟。其次就像之前的 3D 眩晕一样，使用者经过一段时间的适应，才能使 VR 成为真正意义上的虚拟现实产物。

添加虚拟参考物

普杜大学计算机图形技术学院的研究人员于近年发现，只要在 VR 场景中添加一个虚拟的鼻子，就能在一定程度上解决头晕等问题。研究人员在各种虚拟场景中对 41 名参与者进行了测试，一部分人会有虚拟鼻子，一部分没有。结果发现，有鼻子的人都能保持更长时间的清醒。研究人员称，之所以能产生治晕的效果，可能是因人需要一个固定的视觉参照物，所以就算在 VR 中加入一个汽车仪表盘，也可能产生相同的效果。

电前庭刺激

目前国外正有一家名叫 vMocion 的公司，打算利用梅奥医学中心的航空航天医学和前庭研究实验室花费 10 多年时间研究的技术去解决这个问题。这项技术名为电前庭刺激 (GVS)，将电极放在策略性位置（每只耳朵后要放置两个电极，一个在前部，一个在颈背），追踪用户内耳的感知运动，并将视野范围的运动触发成 GVS 同步指令，刺激产生三维运动。如果行得通的话，它可以让用户完全沉浸在当前的环境中，真正感觉到自己驾驶的宇宙飞船在俯冲或转弯。

调节镜片之间的距离

在最原始的 HMD 中其实并没有考虑到这一点,考虑更多的只是怎样适配不同程度近视的用户(有些 HMD 甚至没有考虑这一点),三星首先采用了滑轮调节镜片的设计,可以自由调节两个镜片之间的距离。另外,也有方案显示可以通过蓝牙控制器等调节画面的中心点。从而保证画面中心、镜片中心、人眼中心三点一线。避免重影,避免晕眩。

光场摄影

一个光场快照可以在图片获取后对照片进行聚焦、曝光、甚至调整景深等操作。它不仅仅记录落在每个感光单元内所有光线的总和,光场相机还旨在测定每个进入光线的强度和方向。有了这些信息,就可以生成不只是一个,而是每一个在那一刻进入相机视野的可能的图像。例如,摄影师常常会调整相机的镜头,以便对面部进行聚焦,刻意模糊背景。也有人想要得到模糊的面部,背景要十分清晰。有了光场摄影,同一张照片,就可以获得任何效果。目前在这方面做得最好的 Magic Leap。Magic Leap 做的不是在显示屏上显示画面,而是直接把整个数字光场投射到使用者的视网膜上,从而可以让使用者可以根据人眼的聚焦习惯自由地选择聚焦的位置,以准确地虚实结合模拟人眼的视觉效果,而完全不会涉及到刷新率和分辨率等问题。

5.2.3 视角

传统设备中的视角处理方式包括镜头控制加角色控制,比如固定摄像机位置,游戏中的角色自由移动,或者摄像机随着游戏中角色的移动而移动,还有就是当观看全景视频时,镜头是跟着头部转动而移动的,也就是每次看到的都是整个球面中的一部分即视角区域,而且当在 VR 游戏中时,游戏设计师无法控制玩家的视角,任何类型的 VR 游戏都是主视角游戏,相机的行为控制是游戏是否会带来晕眩感的关键因素。

5.3 球面指标测度

目前,VR 市场上的产品质量参差不齐,需要有统一的标准来进行评价和判断。第 120 届 MPEG 会议上,ITU-T SG12 Q13/12 文档中概述了新的工作项目(虚拟现实的体验质量),包括应用场景和相关的 QoE 因素。其工作的目标是开发主观测试方法和质量模型,并且实际上已经有相关组织开始为 360 度 HMD 视频和 VR 游戏的主观评估提出一些新的建议。目前的基准提供了 VR 硬件和软件,VR 使用案例,影响因素分类,以及 VR QoE 指标。在 ITU-T SG12 全体会议上,华为公司提出了 QoE for VR(Quality of Experience for Virtual Reality)工作立项提案,该工作项目将输出关于 VR 业务质量领域的 Recommendations 文档,向业界提供关于 QoE 因素、QoE/QoS 方面需求、主观测试方法和客观质量评价模型等方面的建议。

Facebook 也一直致力于推动 360 度技术的发展,而且也开拓创新了一系列的新概念,比如说有偏移立方体贴图(offset cubemaps),动态流式传输(dynamic streaming)和基于内容的流式传输(content-dependent streaming)。每一项新技术都对 360 度技术有重要的作用,同样重要的是每一项新技术与原先方法比较时的体验质量,即 QoE 的提高,最好能够提供一致的标准,并且能够对以后新出现的技术也作出衡量。

然而,目前 VR 行业还没有统一认可的 360 度内容评价标准,而传统的视频评价体系无法对 360 度视频的相关属性做出评价,如沉浸感和观影控制等。360 度视频每一帧都压缩了一个全方向的球形场景,如果要利用传统的评价体系,还需要将 360 球形帧还原成矩形帧,

这种翘曲操作会使得评价结果变得相当不可靠。而且，用户在 360 度视频播放期间可以控制任意时刻的观影方向，这也就意味着最终的视图质量取决于用户视场的帧区域，而不是整体的球形帧。

5.3.1 Facebook: SSIM360 和 360QVM

SSIM360

首先简单介绍一下传统非 360 度 SSIM。结构相似性指标 (Structural Similarity Index, SSIM) 是广泛应用于图像和视频编码的 QA 指标。评估过程为先输入两幅图像：参考图像（如原始内容）和变化图像（如编码内容），然后输出是 0 和 1 之间的分数。分数代表两幅图像的结构相似性，1 代表图像相同，因此质量保存完好，0 意味着两者完全不同，意味着严重的结构变形。该评估过程是逐帧完成的，而且仅在两个输入视频的长度和帧速率完全相同时才起作用。

跟大部分的图像处理算法类似，SSIM 不会一次性测量整幅图像。相反，它从两个输入图像中采样较小的区域（类似于图像压缩中的宏块）并比较样本。在最初的论文中，作者提出了一个用于这一采样的 11×11 高斯内核，而 ffmpeg 的 `vf_ssimg` 滤波器使用了一个统一的 8×8 方块。从每个样本中获取一个 SSIM 值。为了获取每幅图像的 SSIM，计算每个样本的 SSIM 平均数。为了获取整个视频的 SSIM，进一步平均计算所有帧的每幅图像 SSIM。

SSIM 中的采样区域是静态的（即相同的大小，相同的形状），并且在最终的平均计算中具有相同的权重。例如在上面的 ffmpeg 实现中，不管位置如何，图像中每一个 8×8 区域对最终的 SSIM 分数都具有同样的影响力。但 360 度媒介并非如此。前面提到的翘曲问题可能会使得同样的图像或视频得到不同的 SSIM。

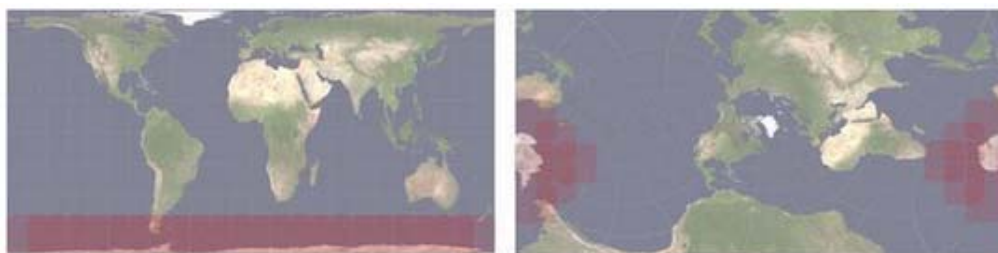


图 5.16 不同方向的球体展开

比较上面的两张世界地图。左边地图的编码器会降低画面下部的质量（南极洲）；而右边地图只是左边编码地图的重定向版本。当渲染至 3D 球体时，这两张地图应该看起来完全一样（除了方向），所以它们从 360 度内容 QA 中得到的分数也应该是相同。但是，如果我们使用 SSIM 作为我们的 QA 标准，则左边地图分数会较低。原因是由于等量矩形中靠近垂直中心的翘曲较少，所以右边地图“糟糕质量”部分（南极洲）所占的比例较小。在 SSIM 的 8×8 取样方块中（在两个地图上呈现为红色放宽），左边地图中南极洲占据的区域被采样 23 次，而右边地图只有 12 次，所以这导致了不同的平均 SSIM 分数。

一种防止翘曲影响的方法是，直接对 360 度球体中的渲染视图进行取样，而非采用等量矩形的渲染视图。我们从 360 度球体中的所有可能视角方向获取无限数量的方形快照。这种样本是平坦图像，同时进行了加权平均，所以翘曲问题将不复存在，而且 SSIM 将是一个有效的 QA。为了防止翘曲，在馈送至非 360 SSIM 之前先获取平面图像的快照 (V_1, \dots, V_n)。

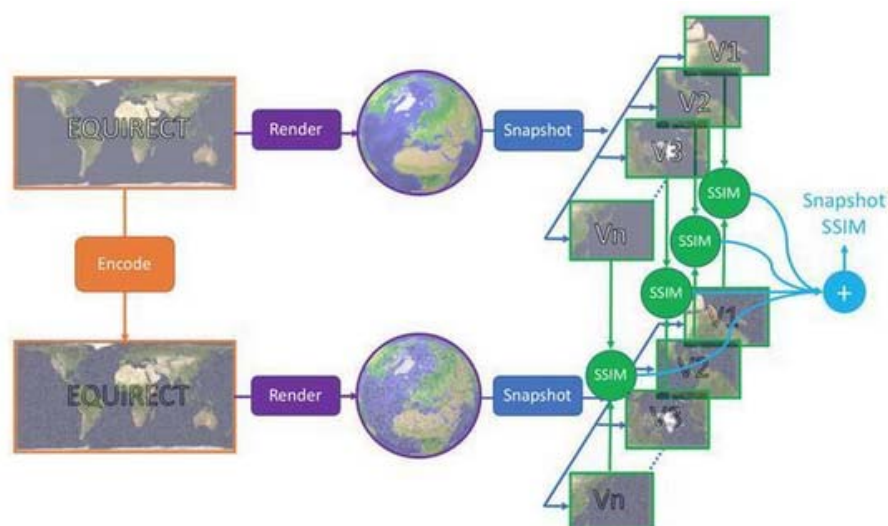


图 5.17 快照方式解决翘曲问题

最终的“Snapshot SSIM”分数将通过汇总所有的每快照 SSIM 分数进行计算。虽然可靠，但这种方法的计算复杂性使其成为不可行的 360 度图像 QA 解决方案。从给定视图中获取单个快照并不是一件简单的事情，更不用说获取无限数量的快照并且逐个运行 SSIM。但是，我们可以使用汇总的快照分数作为验证我们新 SSIM360 标准的基础事实。下面介绍 Facebook 提出的 SSIM360。

SSIM360 解决翘曲问题的方法是，在计算平均值时为每的样本 SSIM 得分加权。加权取决于采样区域在影像中的拉伸程度：拉伸越多，加权越小。这可以通过“样本覆盖的渲染球体比例”和“样本覆盖的帧比例”的比例进行计算。每个样本对最终分数的贡献不同，所以能有效地消除翘曲影响。

为了验证 SSIM360 的结果，将其与 Snapshot SSIM 进行比较。针对不同的纹理和动态特征的 360 视频，以及不同的编解码器，缩放和质量保存目标进行了验证实验。然后通过 1) SSIM, 2) SSIM360 和 3) Snapshot SSIM 评估这些降级测试用例。

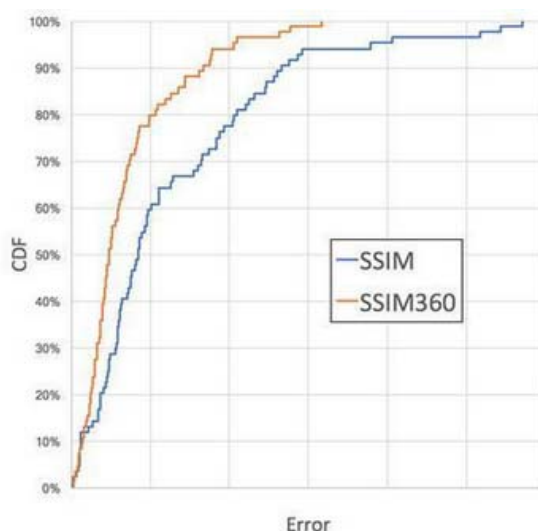


图 5.18 不同 SSIM 性能比较

上图中是 SSIM360 和 SSIM 的累积分布函数 (CDF)，可以看到 SSIM360 几乎能在所有的百分比中把错误减少约 50%。错误定义为 Snapshot SSIM 分数与每个 QA (SSIM360, SSIM) 输出分数之间的绝对差异。从计算角度而言，由于 SSIM360 只是通过可预先计算的权重图来替换每样本 SSIM 聚合中的统一加权，因此 SSIM360 与传统 SSIM 一样有效。

在 360 度视频中还涉及到视场问题，SSIM360 是在不确定用户注视点的情况下评估编码级别的质量。但在播放质量评估中，我们必须考虑到视图方向。在任何特定时间下，只有大约 15% 的 360 度场景会保留在用户视场中。这意味着通过 SSIM360 计算的整体球形帧质量将不再具有代表性。当在内容交付框架中采用基于视图的优化时，情况尤其如此。基于视图的优化技术（如偏移投影，基于显著性的编码，以及基于内容的流式传输）基本上将位分配（相当于大多数情况下的像素分配）偏向视频中感知更重要的区域。它们并没有提高整个画面的质量，而是优化用户最有可能注视的区域。而 SSIM360 由于侧重于整个画面无法捕捉到这种优化。

Facebook 通过两个映射解决了视场问题：一个是从视图方向到像素密度，另一个则是从像素密度到应用于 SSIM360 得分的比例系数。第一个映射考虑了编码中使用的投影，以及播放期间的视图方向和视场，并且询问每个时间点内有多少像素停留在视场之中。换句话说，FoV（Field of View）中有多少像素，通过已知的像素密度图，可以通过封闭型几何公式来有效地计算答案。

第二个映射通过引入反映像素密度变化的比例系数来调整 SSIM360 的质量分数。在非 360 度 QA 中，会使用类似的技术来评估缩小尺寸所导致的质量下降（即将图像“缩小”到更小的宽度/高度）。由于 SSIM 假定参考图像和变化图像具有相同的尺寸，因此缩小后的图像必须放大至原始尺寸，并重新采样以评估质量下降情况。不可以忽略这所需要的计算工作，特别是评估转码过程。转码过程会产生数十种具有不同维度的编码格式，而所有这些都需要在单独的 QA 过程中重新采样，如下图所示。

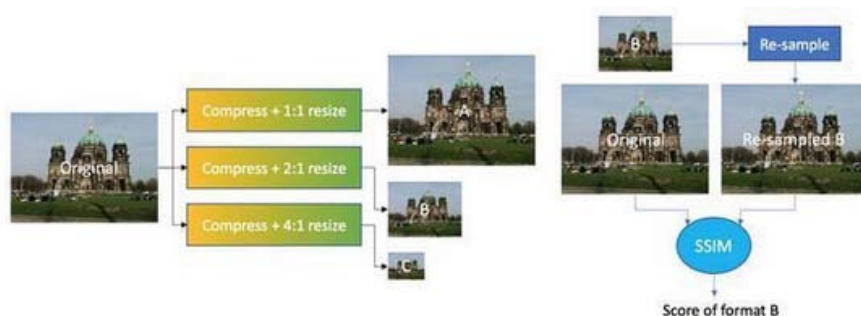


图 5.19 SSIM 过程需对不同大小的图像重新采样

要避免这种计算开销，我们可以将转码过程的 QA 解耦为两个步骤：压缩的 QA；调整大小的 QA。如下所示，SSIM 在压缩版本上执行，同时不调整大小。然后将惩罚因子应用于压缩版本来近似计算每个调整大小的格式的单独质量分数。调整大小的比例是决定惩罚因子的一个因素：调整大小的版本越小，损失的结构细节越多，所以惩罚因子越高。通过确定惩罚因子来取代重新取样更为有效，并且能够减少我们运行 SSIM 的次数。

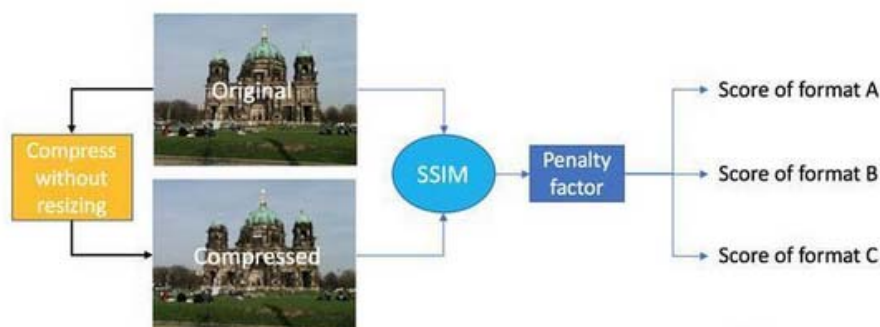


图 5.20 惩罚因子

在 360 度视频中，分辨率的变化不是来自调整大小，而是来自视场内像素密度的变化。随着用户注视点朝更高/更低像素密度的区域移动，观看内容（即视场内的内容）的分辨率将随之增加/减少。比例系数与惩罚因子不同，因为它可以减少或增加 SSIM360 得分。对于基于视图的投影，视场内的分辨率会发生变化，就好像视频分辨率通过“视场内的像素密度”与“整个帧的像素密度”之间的比率进行升级或降级一样。

360VQM

通过将以上的技巧应用到定制的 QA (quality assessment) 工作流程中，就可以解决 360 度视频会话中 QA 的翘曲和视场问题。SSIM360 取代 SSIM 来处理编码中的扭曲，而惩罚因子则由比例因子所取代（因视图改变引起的像素密度变而衍生出来）。这一 QA 工作流程的结果称为 360VQM：360 度视频质量指标。它可以准确高效地捕捉编码过程中的质量变化并在播放过程中将分辨率变化的影响体现在最终分数上。

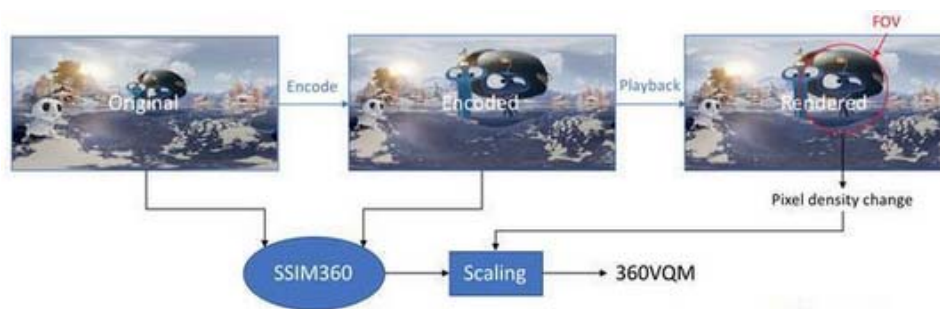


图 5.21 360VQM

5.3.2 VQEG/ITU 中的 3DTV

VQEG 是视频质量专家组 (Video Quality Experts Group) 的简称，该组织专门研究视频评价，其中的 Immersive Media Group 负责沉浸式媒体的质量评价，包括虚拟现实，增强现实，立体 3DTV，多视图等。

目前来说，该组织正在制作一组独特的视频序列来进行实验 (GroTruQoE 数据集)，第一步是使用配对比较方法进行大规模实验，因为序列比较简单，受试者可以轻松地提供对总体偏好的判断。相应的测试计划是“为 3D 视频质量评估中的评估方法建立 3D 体验质量的基础真相” (GroTruQoE3D1)，第二步将使用配对比较测试的结果作为 groundtruth 数据库，以研究哪种更具时效性的主观测试方法可用于预测配对比较测试的结果。

5.3.3 Visbit 360 基准视频

VR 行业有很多关于 4K 与 8K，传统渲染与注视点渲染等话题的比较讨论。开发者可能想知道以 8K 质量制作的内容是否会对 360 度 VR 视频用户体验产生任何影响。VR360 度视频的最终清晰度受到源视频质量、压缩、渲染管道、屏幕分辨率、屏幕像素结构、VR 镜头质量，甚至是眼睛锐度等多种因素的影响。更复杂的是，这一切无法简单地叠加。这涉及大量的重新采样，混叠和插值。最后，即使没有上述的复杂问题，当每个人都用他们自己的视频进行比较时，仍然很难轻松分辨出差异，因为其质量和分辨率都有所不同。

为此，Visbit 推出了基准视频 (VB2018VR)，可以在各种播放器中轻松进行测试，并且判断其质量差异。下面是 Visbit Benchmark 360 VR Video (VB2018VR) 的截图，它可以用来测试、评估和选择最佳的播放解决方案，也可以用来检查后期制作流程，判断管道中是否存在任何质量丢失。

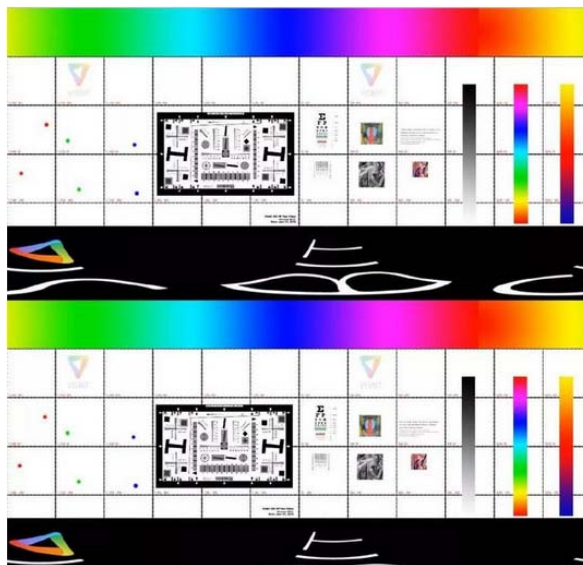


图 5.22 VB2018VR V0.5 的截图 (7680 x 7680 TB, YUV444p)

这个基准视频长约 10 秒，采用 7680×7680 分辨率（立体上下），色域为 YUV444p，速率是 30 FPS，采用 H.264 编码（如需 HEVC 版本，可联系 Visbit）。视频包括三个可视区域，专门用于测试细节和可读性、动态范围和帧速率。

测试的时候，只需要将 8K 基准视频上传至任意平台或播放器，即可进行比较：(1)肉眼比较；(2)截图比较。其中有关于颜色矫正的注意事项：(1)为补偿 VR 透镜的颜色失真，有的播放器（如 Visbit 8K VR Player）会校正颜色。色彩校正可以改善 VR 中的观影体验。但在屏幕截图中，你可能会看到锐利边缘出现色溢，从而令细节变得模糊。(2)如有需要，你可以获取使用非颜色校正版本以进行基准测试。(3)否则，你应该仅比较视场中心，因为这里的颜色校正程度最小。

如果你瞄准的是最高清晰度，以下便是截至 2018 年 6 月的映维网推荐的测试硬件：Pico Neo, HTC Vive Focus 或者搭载 Galaxy S9 的 Gear VR。在观影基准视频时，以下是可能用到的各种质量测试。

视场测试

对于这个基准视频，你可以轻松看到每个头显的视场。需要注意的是，VB2018VR 是以度来表示视场角。打开头显并播放视频后，你就可以读取水平和垂直视场。例如，Gear VR + S9 上的视场大约是 70 度。

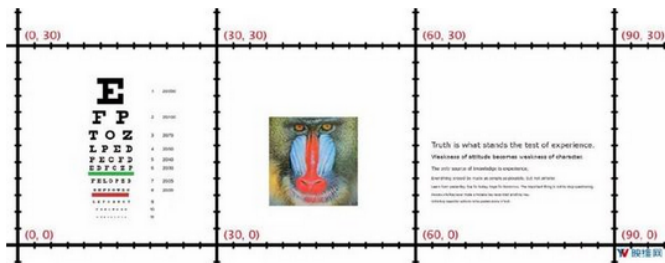


图 5.23 VB2018VR 中的视场角

细节与可读性测试：面部与文本

面部和文本可能是最重要的两种对象。通常而言，人类对面部和文本更加敏感。下面将纳入著名的 Lena, Barbara 和 Baboon 图，以及包含不同字体大小的文本，如下图所示。

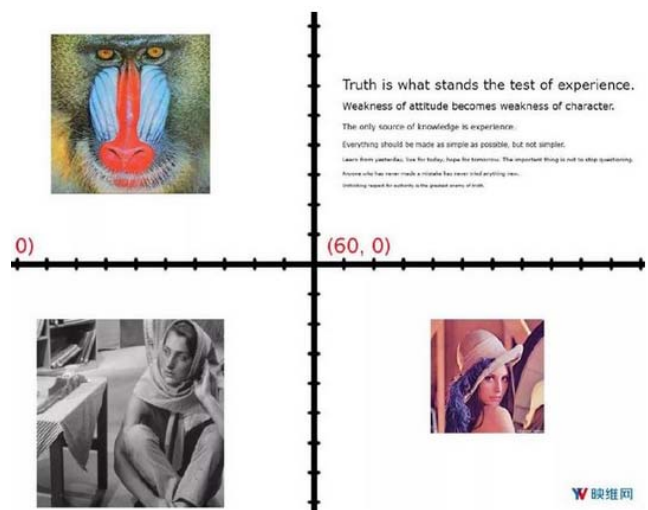


图 5.24 面部与文本

要留意的内容：能快速阅读哪一行文字？是否能清楚地看到面部的细节，如眼睛，嘴唇和鼻子？是否能看到女人衣服或动物皮毛上的纹理？

细节与可读性测试：视力表

VB2018VR 纳入了两张视力表。你最低能看到哪一行呢？要注意的是，左边的视力表本身有点模糊，VB2018VR 将在之后进行优化。

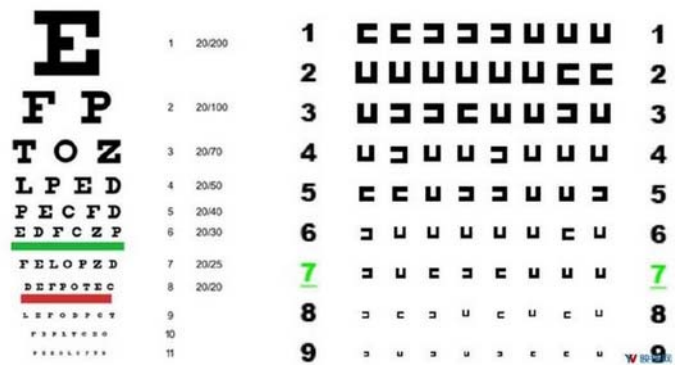


图 5.25 分辨率测试标版

下图则包含了典型的 ISO-12233 分辨率图表。图标的锯齿十分明显。你可以容忍多少锯齿呢？这个问题有点主观。回答这个问题前你应该回到面部与文本的测试，然后再来看看是否能接受其锯齿程度。

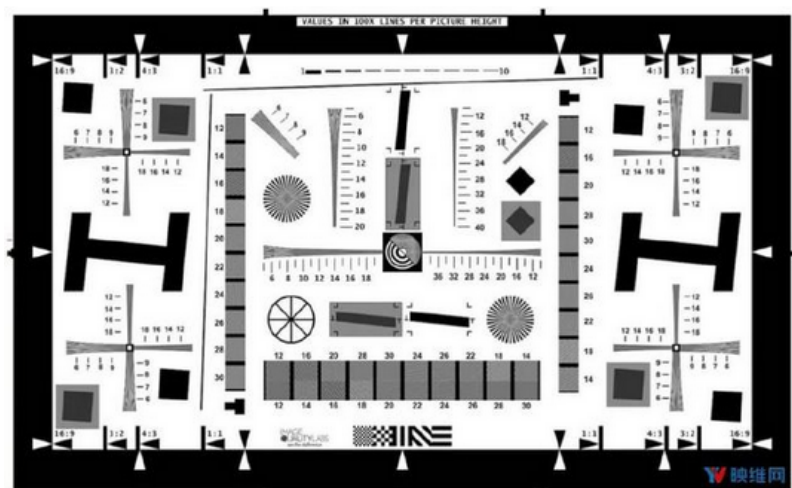


图 5.26 ISO 分辨率测试标版

动态范围测试

如果播放器没有很好地处理色域或动态范围，你将看到渐变颜色出现带状效果。

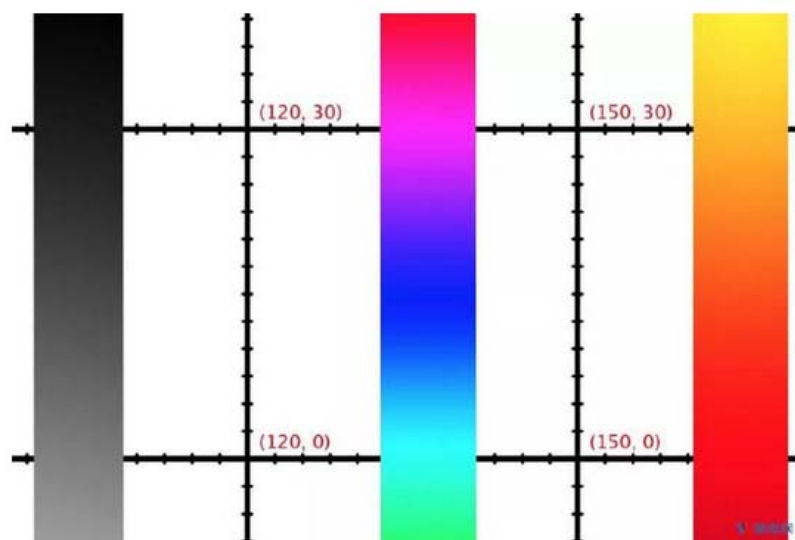


图 5.27 用于动态范围测试的色彩渐变

帧率测试

VB2018VR 同时包括以不同速度和 3D 路径移动的多个点。另外，视频最上方和最下方正在快速旋转。你可以依次判断运动的流畅度。

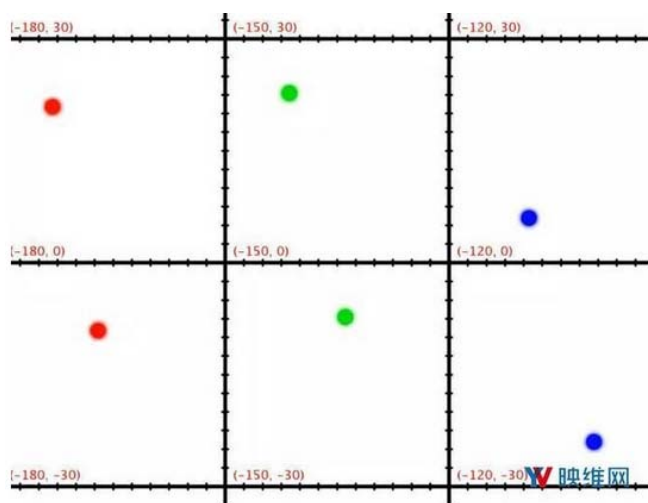


图 5.28 帧率测试的移动点

文件对比测试

VB2018VR 提供了一系列的测试视频，内容相同，只是分辨率不同。你可以自由下载并用任何视频播放器进行测试，然后分享比较结果：

1. visbit-8k-stereo: YUV444P, 7680 x 7680 @ 30fps
2. visbit-6k-stereo: YUV444P, 5760 x 5760 @ 30fps
3. visbit-6k-stereo-yuv420: YUV420P, 5760 x 5760 @ 30fps
4. visbit-4k-stereo: YUV444P, 3840 x 3840 @ 30fps
5. visbit-4k-stereo-yuv420: YUV420P, 3840 x 3840 @ 30fps
6. visbit-4k-squeeze-stereo: YUV444P, 3840 x 1920 @ 30fps
7. visbit-4k-squeeze-stereo-yuv420: YUV420P, 3840 x 1920 @ 30fps

5.3.4 游戏体验质量建模

在消费级沉浸式媒体服务中，游戏是占很大比例的一项内容。相比于 2D 屏幕，在虚拟全景世界中进行游戏体验是极为真实刺激的。从目前火热的游戏直播的角度而言，Twitch TV 是其中相当成功的一个平台，经统计，2015 年 Twitch TV 每月的访问者有近 1 亿。游戏直播是指游戏主播向平台传输游戏画面，并通过直播平台向广大观众展示自己的游戏过程。与这一概念类似，用户终端无需任何高端处理器和显卡，只需基本视频解压能力就可以直接进行游戏的方式被称为“云游戏”，而在沉浸式媒体领域，“云游戏”的最大瓶颈就在于延迟。

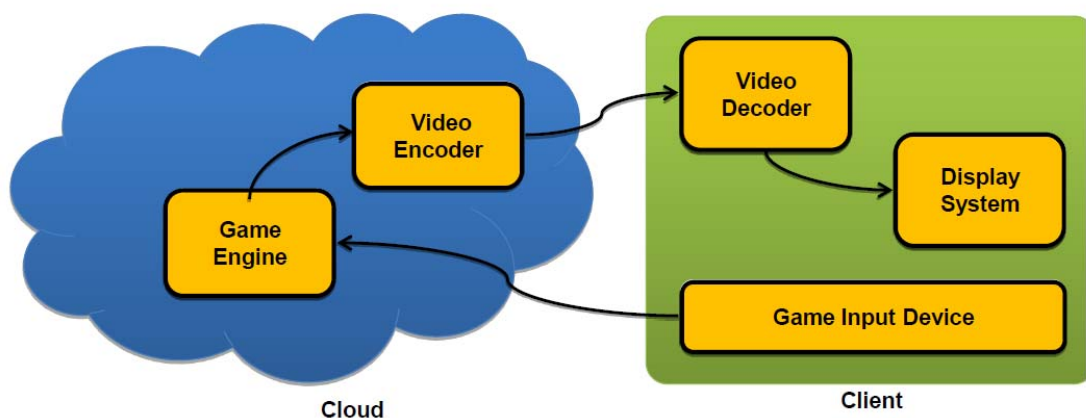


图 5.29 “云游戏”

随着类似上述的各类游戏方式的诞生，游戏体验质量的评估也有待完善。目前，ITU-T 的 Q.13/SG 12、Q7/SG 12 项目都涉及到游戏应用的 QoE。VQEG-IMG 工作组于 2018 年根据多编码方式、多分辨率、多比特率版本的 16 个游戏视频对游戏 QoE 完成了一项研究。该研究利用多种特征如平均差 (MAD, Mean Average Difference)、空间信息 (SI, Spatial Information)、时间信息 (Temporal Information) 进行游戏体验质量的建模，其大致建模框架如下图所示：

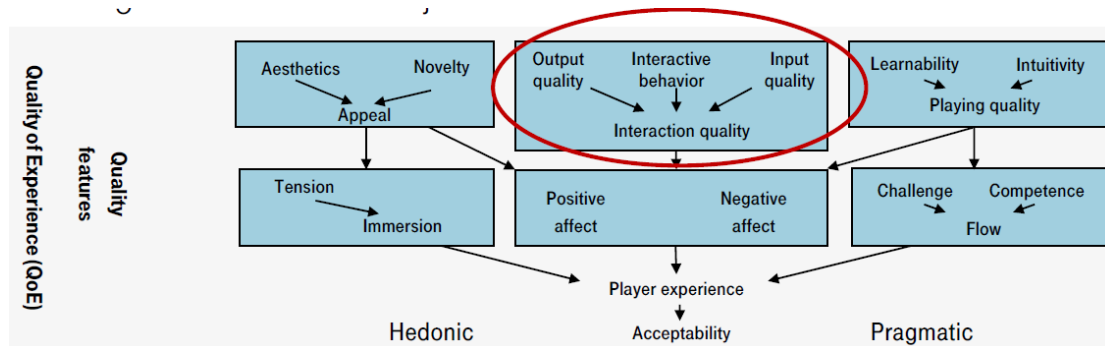


图 5.30 游戏体验质量建模框架

多重因素建模的 QoE 分数与主观测试的 MOS 进行拟合，最终得到以下等式：

$$MOS = 1.102 + 0.59 * PosAffect + 0.24 * Reactiveness + 0.25 * VideoQuality \quad (5.1)$$

其中， $Reactiveness = \exp(0.84 + 4.43 / \text{帧率})$ 。此模型是一种无参考质量评价算法，同时未涉及比特率模型，游戏中的正面影响和输出流是两种主要的分数贡献因素。

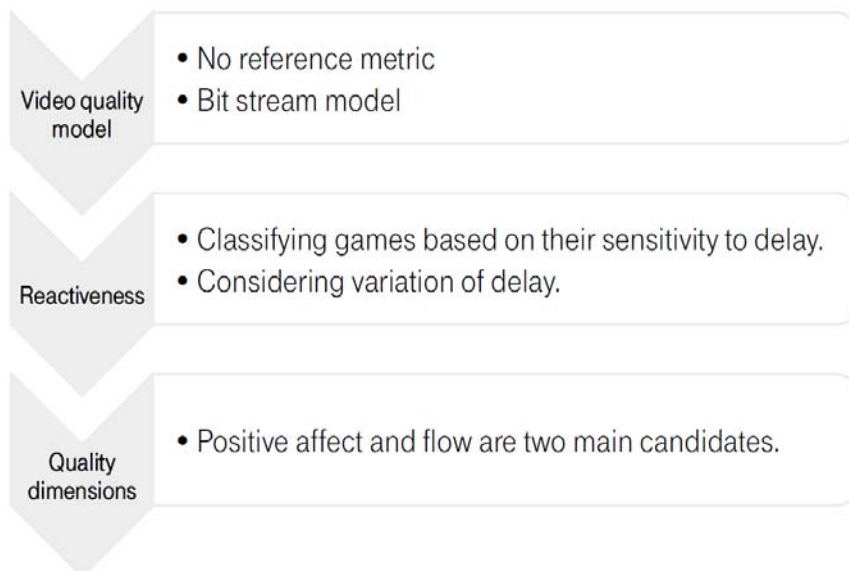


图 5.31 游戏 QoE 建模特点

本章参考资料：

- [1]<https://mp.weixin.qq.com/s/JFrZzdxGjUZNkdfDfQ0-Pw>
- [2]<http://www.52vr.com/article-2115-1.html>
- [3]<https://blog.csdn.net/tcpipstack/article/details/52024537>
- [4]<http://baijiahao.baidu.com/s?id=1594786358644717130&wfr=spider&for=pc>
- [5]<https://www.its.bldrdoc.gov/vqeg/projects/immersive-media-group.aspx>

- [6]<https://www.its.bldrdoc.gov/vqeg/meetings/madrid-spain-march-19-23.aspx>
- [7]Brunnström K, Sjöström M, Imran M, et al. Quality of Experience for a Virtual Reality simulator[C]// Human Vision and Electronic Imaging. 2018.
- [8]Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4):600-612.