目 录

[1. 范围 2](#_Toc532406143)

[2. 规范性引用文件 2](#_Toc532406144)

[3. 缩略语、术语和定义 3](#_Toc532406145)

[3.1 缩略语 3](#_Toc532406146)

[3.2术语和定义 5](#_Toc532406147)

[4. VR视频服务用户体验指标体系 6](#_Toc532406148)

[4.1 性能指标 6](#_Toc532406149)

[4.2 指标框架 6](#_Toc532406150)

[4.2.1 VR业务QoE评价体系 6](#_Toc532406151)

[4.2.2 视频主观评价方法 7](#_Toc532406152)

[4.2.3 视频客观评价方法 9](#_Toc532406153)

[4.2.4 体验影响因素评测 10](#_Toc532406154)

[4.2.5 影响VR全景视频拼接质量的因素 11](#_Toc532406155)

[4.2.6 全景VR体验评估模型框架 11](#_Toc532406156)

[4.3参数采集及指标处理 12](#_Toc532406157)

[5. VR视频服务用户体验指标构成 12](#_Toc532406158)

[5.1 头戴显示器的发展状况 13](#_Toc532406159)

[5.1.1 沉浸式头戴显示器的发展状况 13](#_Toc532406160)

[5.1.2 透射式头戴显示器的发展状况 14](#_Toc532406161)

[5.1.3 头戴显示器未来的发展趋势 17](#_Toc532406162)

[5.2 视频源质量参数 18](#_Toc532406163)

[5.3 头戴显示器参数 19](#_Toc532406164)

[5.4 沉浸体验质量参数 20](#_Toc532406165)

[5.5 交互体验质量参数 21](#_Toc532406166)

[5.5.1 交互体验参数 21](#_Toc532406167)

[5.5.2 强交互体验业务指标 22](#_Toc532406168)

[5.6 网络KPI指标 24](#_Toc532406169)

[5.7 业务KQI指标 25](#_Toc532406170)

[5.8 Cloud VR业务时延要求 26](#_Toc532406171)

[6. VR视频服务用户体验评估算法 27](#_Toc532406172)

[6.1 总体模型综述 27](#_Toc532406173)

[6.2 模型三大模块综述 28](#_Toc532406174)

[6.2.1视频体验质量Q1 28](#_Toc532406175)

[6.2.2交互体验质量Q2 38](#_Toc532406176)

[6.2.3沉浸体验质量Q3 39](#_Toc532406177)

注：标黄色的标题暂缺内容。

# 1. 范围

本标准规定了应用于虚拟现实视频服务业务用户体验质量的评估场景和模型，分析了对VR视频业务用户体验质量的关键因素，同时定义用于评估VR视频服务用户体验质量的参数和技术方法。

本标准适用于对中国移动、中国联通和中国电信等电信运营商，广电运行商，VR硬件头显提供商，VR视频内容提供商，云VR服务提供商和其他相关厂商提供的VR视频用户体验质量进行综合评估，此外还适用于产业各方对影响VR视频服务质量的因素进行量化和分析。

该技术报告提供了与VR服务相关的QoE度量的研究。研究的重点是：

1. 研究可能影响VR服务体验的关键性能指标。

2. 确定VR QoE测量的设备参考模型。

3.识别国际标准中定义的与虚拟现实用户体验相关的现有QoE参数和度量。

4.分析现有关于VR视频体验因素的潜在改进，以便更好地适应VR服务。

# 2. 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅所注日期的版本适用于本文件。凡是不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

**[3GPP TR 26.929 V0.4.0]** VR用户体验的QoE参数和指标

**[3GPP TS 26.247], [3GPP TS 26.114]** VR QoE参数

**[3GPP TR 26.918 V15.2.0]** 3GPP上的VR媒体服务

**[3GPP TS 26.118 V1.0.0]—[3GPP TS 26.118 V15.0.0]** 流应用的3GPP虚拟现实配置文件。

**[ITU-T / SG 12 G.QoE-VR based on TD326-R1]** 1、识别VR业务及关键技术。2、识别影响VR QoE的关键质量因素。3、定义描述VR QoE质量的度量方法。4、研究VR质量的客观和主观评价方法论。5、输出保证良好VR QoE质量需求的建议。

**[ITU-T / SG 12 G.360-VR]** 研究具体如何评测360度全景视频的体验，包括主观测试方法，以及客观模型。

**[IEEE P 2048.1]** 该标准规定了VR与AR设备的种类和定义。

**[IEEE P 2048.2]** 该标准规定了沉浸式视频种类和质量指标。

**[IEEE P 2048.3]** 该标准规定了沉浸式视频文件格式和流媒体格式以及此类格式的功能性和交互性。

**[IEEE P 2048.4]** 该标准规定了在VR中个人身份认证的条件和方式。

**[IEEE P 2048.5]** 该标准详列了为工作站和内容消耗的环境提出的参考意见。

**[IEEE P 2048.6]** 该标准规定了在VR APP中使用沉浸式用户界面的条件和方式及沉浸式用户界面中的功能性和交互性。

**[IEEE P 2048.7]** 该标准规定了AR/MR APP的使用条件、系统、使用方式、检测和认证方式，以创建并使用其在真实世界中的虚拟对象映射。

**[IEEE P 2048.8]** 该标准规定了虚拟对象与现实世界之间的互操作性的使用的条件、系统、使用方式、检测和认证方式。

**[IEEE P 2048.9]** 该标准规定了沉浸式音频的分类和质量指标。

**[IEEE P 2048.10]** 该标准规定了沉浸式音频文件和流的格式，以及格式所支持的功能和交互。

**[IEEE P 2048.11]** 该标准定义了增强现实（AR）系统的总体框架，用于辅助车辆中的驾驶员和/或乘客。

**[IEEE P 2048.12]** 该标准定义了虚拟现实（VR），增强现实（AR）和混合现实（MR）的内容评级和描述符。

# 3. 缩略语、术语和定义

## 3.1 缩略语

以下缩略语适用于本文件。

|  |  |  |
| --- | --- | --- |
| 缩略语 | 英文全称 | 中文全称 |
| AR | Augmented Reality | 增强现实技术 |
| BAR | Block Area Ratio | 花屏面积占比 |
| CDN | Content Delivery Network | 内容分发网络 |
| CBR | constant bit-rate | 恒定比特率编码 |
| DPI | Deep Packet Inspection | 深度包检测 |
| DoF | Degree of Freedom | 自由度 |
| FCC | Fast Channel Change | 快速频道切换 |
| FoV | Field of View | 视场角/视野 |
| FR | Full Reference | 全参考 |
| FTP | File Transfer Protocol | 文件传输协议 |
| GOP | Group of Pictures | 图像组 |
| HG | Home Gateway | 家庭网关 |
| HTTP | Hyper Text Transfer Protocol | 超文本传输协议 |
| ICP | Internet Content Provider | 因特网内容提供商 |
| IGMP | Internet Group Management Protocol | 因特网组管理协议 |
| I/O | Input output system | 输入输出系统 |
| IP | Internet Protocol | 互联网协议 |
| ISP | [icosahedron projection](http://www.baidu.com/link?url=52ABlf7uqIzMPro_eBiWjtP-nq3gv24Khz3RlzLacM01gA1vM2YexRttUvfvd2cLk8lDKYFgEi6Cc7Qu-BE_7XNDGNtIHurG5tP8QCp48u3) | 二十面体投影 |
| HD | High Definition | 高清晰度 |
| HDR | High-Dynamic Range | 高动态范围 |
| HMD | Head Mount Display | 头戴式可视设备 |
| HTTP | Hypertext Transfer Protocol | 超文本传输协议 |
| HLS | HTTP Live Streaming | Apple的动态码率自适应技术 |
| ITU | International Telecommunication Union | 国际电信联盟 |
| KPI | Key Performance Indicator | 关键性能指标 |
| KQI | Key Quality Indicator | 关键质量指标 |
| MIMO | Multiple-Input Multiple-Output | 多入多出技术 |
| MOS | Mean Opinion Score | 主观体验评分 |
| MU-MIMO | Multi-User Multiple-Input Multiple-Output | 多用户多入多出技术 |
| MV | Motion Vector | 运动矢量 |
| NR | No Reference | 无参考 |
| OFDMA | Orthogonal Frequency Division Multiple Access | 正交频分多址 |
| OHP | octahedron projection | 八面体投影 |
| OQA | Objective Quality Assessment | 客观质量评估 |
| PPI | Pixels Per Inch | 像素密度-每寸 |
| PPD | Pixel Per Degree | 像素密度-每度 |
| PSNR | Peak Signal to Noise Ratio | 峰值信噪比 |
| QAM | Quadrature Amplitude Modulation | 正交振幅调制 |
| QP | Quantization Parameter | 量化参数 |
| QoE | Quality of Experience | 体验质量 |
| QoS | Quality of Service | 服务质量 |
| RGB | Red Green Blue | 红绿蓝颜色标准 |
| RR | Reduced Reference | 部分参考 |
| RTCP XR | Real-time Transport Control Protocol Extended Report | 实时传输控制协议扩展报告 |
| RTSP | Real Time Streaming Protocol | 实时流协议 |
| RTT | Round-Trip Time | 往返时延 |
| SMTP | Simple Message Transfer Protocol | 简单邮件传输协议 |
| sQuality | score of Quality | 质量评分 |
| sInteraction | score of Interaction | 互动体验评分 |
| sView | score of View | 播放体验评分 |
| sInteractionDeg | sInteraction Degradation | 互动体验损伤 |
| sViewDeg | sView Degradation | 播放体验损伤 |
| SQA | Subjective Quality Assessment | 主观质量评估 |
| SYN | Synchronous | 同步信号 |
| TCP | Transmission Control Protocol | 传输控制协议 |
| TR | Time Ratio | 花屏时长占比 |
| UDP | User Datagram Protocol | 用户数据报协议 |
| UHD | Ultra High Definition | 超高清晰度 |
| VBR | variable bit-rate | 可变比特率编码 |
| VC | Video Complexity | 视频内容复杂度 |
| vMOS | video Mean Opinion Score | 视频体验评分 |
| VR | Virtual Reality | 虚拟显示 |
| WS-PSNR | Weighted Spherical Peak Signal to Noise Ratio | 加权球形峰值信噪比 |

## 3.2术语和定义

下列术语和定义适用于本文件。

3.2.1 用户体验质量 quality of User experience

用户感受到的服务过程中硬件操作、服务操作和内容展示的质量水平。

3.2.2 业务质量 quality of service

指网络利用各种技术，为指定的业务提供服务的能力。

3.2.3 用户体验关键参数 key parameter of User experience

对用户感受到的服务过程中硬件操作、服务操作和内容展示的质量水平有关键影响的因素，可以被表示为客观的体验质量参数。

3.2.4 业务性能参数 performance parameter of service

服务过程中可以衡量服务实现和内容展示质量水平的因素，可以被表示为客观的体验质量参数。

3.2.5 网络性能参数 performance parameter of network

对业务完成和业务质量程度有直接影响的网络关键性能指标。

3.2.6 2k

2K分辨率指屏幕横向像素达到2000以上，主流2K分辨率为2560x1440。我国数字影院放映机主要采用这种分辨率。许多高端手机屏幕也开始使用这种分辨率。其他的2048×1536（QXGA）2560×1600（WQXGA），2560×1440(Quad HD)也被注释成为不同的2K。

3.2.7 4K

4K分辨率是指水平方向每行像素值达到或者接近4096个，多数情况下特指4096\*2160分辨率。而根据使用范围的不同，4K分辨率也有各种各样的衍生分 辨率，例如Full Aperture 4K的4096\*3112、Academy 4K的3656\*2664以及UHDTV标准的3840\*2160等，都属于4K分辨率的范畴。

3.2.8 8K

8K分辨率是指水平方向每行像素值达到或者接近7680个，多数情况下特指7680 × 4320分辨率。

3.2.9 视频直播 Video streaming

根据用户发出播放请求，系统将实时的视频和音频流，通过传输网络传送到用户终端

3.2.10 视频点播 Video on Demand

根据用户发出点播请求，系统将存放在片源库中的节目信息检索出来，以视频和音频流文件，通过传输网络传送到用户终端。

3.2.11 VR弱交互业务

包含巨幕影院、360°全景视频、VR直播等，用户可以在一定程度上选择视点和位置，但用户与虚拟环境中的实体不发生实际的交互（如触摸）。

3.2.12 VR强交互业务

包含VR游戏、VR家庭健身、VR社交等，用户可通过交互设备与虚拟环境进行互动，虚拟空间图像需对交互行为作出实时响应而生成。

3.2.13视觉逼真度

视觉逼真度是用户感知到的虚拟现实系统的视觉输出（头戴显示设备上360度全景视频提供的虚拟场景）与真实世界视觉刺激的接近程度。

3.2.14听觉逼真度

与视觉逼真度类似，听觉逼真度代表了虚拟现实系统的听觉输出（虚拟现实系统中立体声或空间化声音所提供的虚拟场景）与真实世界听觉刺激的接近程度。

# 4. VR视频服务用户体验指标体系

## 4.1 性能指标

TBD

## 4.2 指标框架

视频服务质量受到终端、网络、平台、片源等多方面因素影响。具体的指标体系如下图所示：



图3 视频质量指标框架

数据源获取的数据经过处理可以得到KPI指标（网元相关指标）和KQI指标（业务相关指标）。

KPI指标一般指基于网元设备的关键指标，主要包括终端性能、网络传输质量、视频传输质量、音频体验质量、媒体的封装质量和视频的编码参数。

KQI指标从业务和应用层面出发反映业务使用体验，视频播放的流畅度，即视频在播放过程中是否有出现缓冲卡顿、花屏和马赛克的情况；视频源质量，包括视频源本身的编码格式、视频清晰度、媒体封装等；业务交互体验是视频在观看过程中业务交互的情况包括直播、点播业务的交互成功率以及交互时延等。KQI指标一般通过KPI指标处理得到。

QoE指标从用户层面出发，可以定义为终端用户对视频业务的总体主观感知。视频业务的感知体验就是反映用户观看视频业务体验的综合评分，用U\_MOS表示。定义U\_MOS=f(Qs, Qa, Qi,Qv)，即用户体验综合评分为（视频质量（Qs），音频质量Qa，交互质量（Qi），观看质量（Qv））的函数关系。

### 4.2.1 VR业务QoE评价体系

VR QoE须根据VR系统提供的沉浸水平以及用户与VR环境之间交互的质量进行评估，用户的沉浸水平可以通过频内容的质量和终端的展现来确定。

用户与虚拟环境之间的交互质量会对整体体验质量产生较大影响。沉浸感强的互动可以带来高品质的体验和强烈的存在感；另一方面沉浸感差的互动会导致不适。



图 4‑19 VR业务QoE评价体系

VR业务QoE评价体系，总体包含6个维度：视频质量、音频质量、网络质量、展示质量、互动质量、自由度。

**视频质量维度**：包含三维属性、每度像素数、码率、投影、视场角、压缩及帧率。

**音频质量维度**：包含三维属性、声道和采样率、压缩三个因素。

**网络质量维度**：网络是VR业务内容传输载体，其传输质量直接决定了VR用户的最终体验。网络质量影响因素较多，从最终用户体验角度考虑主要包含感知时延和卡顿次数。

**展示质量维度**：主要从播放质量、音视频同步、音视频空间对齐三个角度评估用户体验。

**互动质量维度**：VR的沉浸感一项重要的参考维度就是互动质量，互动质量需要从VR环境交互性、空间交互精确度两方面评估。

自由度：通常情况下是3DoF，6DoF等。针对不同阶段及不同的传感器配置，可能还包含3DoF+，Windowed 6DoF, Omnidirectional 6DoF等细分维度。

### 4.2.2 视频主观评价方法

视频的主观评价方法是基于用户的具有准确度高、真实性强的优点，而缺点是测试的时候需要耗费巨大的人力，物力和时间。同时在实际情况中会有各种的个人主观因素的影响，其评价方法实用性差。所以如果保证主观评价的准确性，做测试实验的时候需要足量人进行测试，测试条件要尽可能的符合标准。

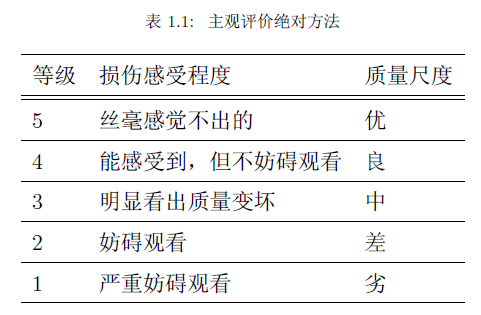
视频的主观评价方法主要根据ITU-R BT.500等标准中提到的方法，布置好相应的测试环境，让观测者按照自己的主观感受，根据规定好的评价标准对测试视频质量作出主观判断打分。根据测试者的打分进行处理，然后得到测试视频序列的主观分值。目前常用的主观评价方法主要包括：双刺激损伤尺度方法，双刺激连续质量尺度方法，双刺激连续评价方法，单刺激法，单刺激连续质量评价方法。上述的方法主观评价在组织方式和评价结果上都有优点和缺点。应用于不同的视频类型的主观评价的试验中，但都有成本昂贵、效率较低、过程不可重复，实验结果受观测者主观因素影响较大的缺点，使用范围比较局限。

#### 绝对测试方法

绝对方法是根据预先规定的评价尺度或自己的经验，对待测图像/视频给出等级评分，如双刺激损伤标度方法（Double Stimulus Impairment Scale，DSIS）和单刺激连续质量标度方法（Single Stimulus Continuous Quality Scale，SSCQE）等。

1. 双刺激损伤标度方法

DSIS方法成对呈现序列。 在每一对中呈现的第一个刺激始终是参考。 第二个刺激是被测系统受到损伤后的参考序列。在这种情况下，受试者被要求评估第二次刺激相对于参考的损害。 应该使用以下五级评级标准进行评级：



1. 单刺激连续质量标度方法（Single Stimulus Continuous Quality Scale，SSCQE）

一般选择绝对分类评级（ACR）方法， 受试者观察一个测试序列，测试序列一次一个地呈现，然后对该序列进

行评分。ACR方法使用以下五级评定等级：

|  |  |
| --- | --- |
| 5 | 极好 |
| 4 | 良好 |
| 3 | 一般 |
| 2 | 较差 |
| 1 | 很差 |

#### 相对测试方法

相对方法常用所谓的群优先尺度，即由观测者将一批图像/视频由好到坏进行分类，按照质量相对优劣给出相应的分数，如双刺激连续质量标度方法（Double Stimulus Continuous Quality Scale，DSCQS）

DSCQS方法使用连续质量标度来提供对视频序列内在质量的测量。每个观看者以由五个线性间隔的评级：

|  |  |
| --- | --- |
| 5 | 一群中最好的 |
| 4 | 好于该群平均水平 |
| 3 | 该群平均水平 |
| 2 | 差于该群平均水平 |
| 1 | 一群中最差的 |

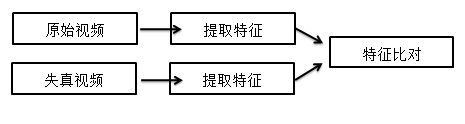
DSCQS方法可适用于VR经验不多的观看者。 为了保证分析结果的可靠性，可选择基于随机策略，让每个观测者从测试对里随机选取随机数量的视频对进行测试。

### 4.2.3 视频客观评价方法

客观视频质量评价方法则是通过构建数学模型，对视频质量进行评价。客观视频质量评价方法根据对原始视频的参照程度分为：全参考，半参考和无参考视频质量评价算法。与主观评价方法相比，客观评价方法更加方便，耗费时间少，应用性强。

#### 全参考质量评价

PSNR、SSIM、VQM等是2D视频质量评价方法，在3D及VR领域一般需要结合其他方法来进行质量评价。



* 基于视差的视频质量评价方法

结合视差图将PSNR、SSIM、VQM等2D评价方法对多视点视频进行质量评价，用来分析因视频质量失真而引起质量的下降情况。

* 基于视觉关注度的的视频质量评价方法

主要是利用人类视觉系统通常对感兴趣区的关注度比较高的原理。

* 基于BJND模型的的视频质量评价方法

当视频失真小于一定阈值时，人们就难于感知。对于立体视频，研究人类视觉系统能够感受到的最小失真变化量将有助于改善视频质量评价方法。

* 基于边缘差异的的视频质量评价方法

主要是利用图像边缘附近的变化最能引起人眼的视觉感知的原理。

#### 无参考质量评价

无参考质量评价也称之为自动化质量评价，即不存在参考视频。评测对象的好坏要完全根据视觉经验来判定，如图像纹理的丰富性、结构的保持性、立体感效果评价等。当需要对虚拟现实360度全景视频进行自动化评测时，需考虑一些客观模型的使用。这些客观模型通常通过采集一些关键信息，通过特定的算法模型来得出与主观评价相似的体验打分。客观模型的生成，通常称之为建模。建模是决定客观模型准确与否的关键。



建模思路建议：

1. 建立视频质量和视频质量参数的模型
2. 建立视频逼真度和视频FOV，立体信息之间的模型
3. 建立音频质量和音频质量参数的模型
4. 建立音频逼真度和空间音频之间的模型
5. 建立视频质量，视频逼真度，音频质量，音频逼真度和整体沉浸感之间的模型
6. 建立交互质量和MTP参数之间的模型

建立整体沉浸感、交互质量和整体最终体验之间的模型。

### 4.2.4 体验影响因素评测

该方式可通过测量各种影响虚拟现实体验的因素，来侧面反映虚拟现实的用户体验如何。当进行该方式评测时，需尽可能考虑各个维度的影响因素。本文建议至少包含以下三个方面：

1. 多媒体质量

包括视频分辨率，码率，PSNR，音频采样率，视频卡顿率，初始缓冲时间等。

1. 交互性体验评测

主要为评测MTP时延，也就是移动头部到相应画面显示所需的平均时延。

1. 终端设备评测

包括MOV视场角, 传感器灵敏度，终端显示屏分辨率，刷新率等。

### 4.2.5 影响VR全景视频拼接质量的因素

基于全景图像、视频拼接的虚拟场景中需要解决以下问题：1、采用什么样的设备获取实景图像；2、如何精确计算出图像间的变换参数并进行拼接；3、如何得到相机或者摄像机的参数，并对图像进行几何修正；4、全景图像、视频的存储和传输；与之对应的技术包括图像的获取、图像的预处理、图像的配准、图像的融合、摄像机标定、颜色校正等技术。基于实景图像生成虚拟现实场景的过程中，由于图像配准算法和图像融合技术在不同的应用场景有一定的限制，所以图像配准方法准确度、图像融合误差等易导致视频帧拼接质量下降。

### 4.2.6 全景VR体验评估模型框架

图4-2 360度全全景视频临场感评估框架



沉浸感是关于虚拟现实系统提供的感知保真度的客观水平而临场感是关于用户对虚拟现实的主观心理反应。评估框架采用了包括三层的分层结构设计。如图5所示，这三层分别为沉浸因素层，中间质量层和临场感层。其中，沉浸因素层包含了针对不同感知线索如视频、音频、迟滞等的相关影响因素。这些因素被用来表征虚拟现实系统的沉浸感水平并进一步地与中间质量层中相应的质量建立关联。为了区分中间质量层中不同类型的质量，该层被划分为两个子层。上方的子层包括视觉逼真度、听觉逼真度和本体匹配度，下方的子层包括视频质量和音频质量。接下来，上方子层中的质量又进一步地联合产生临场感层中的临场感。下面将详细介绍评估框架中各质量的定义以及层与层之间的关系。

首先，定义中间质量层中五种不同类型的质量如下：

a) 视频质量——视频质量是用户对头戴设备上播放的360度全景视频的整体感知质量；

b) 音频质量——音频质量是用户对所听到声音的整体感知质量；

c) 视觉逼真度——视觉逼真度是用户感知到的虚拟现实系统的视觉输出（头戴显示设备上360度全景视频提供的虚拟场景）与真实世界视觉刺激的接近程度。

d) 听觉逼真度——与视觉逼真度类似，听觉逼真度代表了虚拟现实系统的听觉输出（虚拟现实系统中立体声或空间化声音所提供的虚拟场景）与真实世界听觉刺激的接近程度。

e) 本体匹配度——本体匹配度表征了头戴显示设备中画面和声音的变化与用户对头戴显示设备的操作（如头部活动）的匹配程度。

评估框架的输入均为沉浸系统技术层面可提取的参数。在定义了各种体验质量的基础上，视频比特率、视频分辨率、视频帧率、音频比特率、音频采样率被分别用来评估360度全景视频的视频和音频质量。这两个质量表征了用户对音视频序列的感知质量。考虑到头戴显示设备所能提供的特殊体验，评估框架中引入了两个附加的因素——视域和立体视觉——来进一步的评估虚拟场景与真实环境的视觉感知接近程度（即中间质量层中的视觉逼真度）。这里的视域是指头戴显示设备的水平视域，而立体视觉则是指360度全景视频提供的是二维视觉或三维视觉。这两个附加因素已经被证明是虚拟现实系统沉浸感水平的重要影响因素[2]。此外，框架使用空间化音频联合音频质量来评估虚拟场景与真实环境的听觉感知接近程度（即听觉逼真度）。此外，当用户在头戴显示设备上观看360度全景视频时，用户可以通过改变头部的方向来观看360度全景视频的不同区域。当头戴显示设备对头部活动信息的捕捉速度慢于头部的活动速度时，实际播放的画面将与期望播放的画面不匹配。这种不匹配将影响本体匹配度的水平，即虚拟场景的反馈与用户操作的匹配程度。这里将实际显示画面与期望播放画面间的时间间隔称为画面迟滞。类似的，这种迟滞同样可能出现在空间化音频中。将实际听到的声音方向与听到的虚拟场景中声音应该存在的方向间的时间间隔称为声音迟滞。这两个因素被用来表征虚拟现实系统的本体匹配度水平。

最后，将视觉逼真度、声音逼真度和本体匹配度进一步的与用户对虚拟现实系统的临场感建立关联。需要注意的是本章中所讨论的临场感为用户感觉到“在那儿”的感受，不涉及如社交和自感知等其他临场感的因素。

## 4.3参数采集及指标处理

TBD

# VR视频服务用户体验指标构成

用户的视频业务体验是多维度的，传统视频体验一般包括视频内容质量，网络传输质量，终端显示质量，以及用户与业务、终端与视频平台的交互质量体验等。相较传统视频，在VR视频服务中，沉浸感也成为衡量VR视频服务质量的重要参数。本章节将从多个维度阐述VR视频服务用户体验指标的构成。

## 5.1 头戴显示器的发展状况

随着计算机技术和移动显示的快速发展，头戴显示器近来发展迅速。它主要是由显示屏幕和显示目镜组成的，主要工作原理是由显示目镜把显示屏幕成像放大在较远的地方，它的一些具体设计参数，包括出瞳大小、出瞳距离、视场角、分辨率以及光学目镜的成像质量。大的出瞳直径可以让用户的眼球在更舒适的范围内进行运动，较大的出瞳距离可以让大家佩戴这个眼镜更便携更舒适，视场角、分辨率，以及光学目镜的成像效果决定了它呈现给大家的效果。所以近眼显示系统追求做大的出瞳直径、长的出瞳距离、大的视场角以及良好的成像效果。头戴显示器按照沉浸感划分为沉浸式头戴显示器和透射式头戴显示器。沉浸式头戴显示器又叫虚拟现实头戴显示器（VR HMD），是将周围环境完全隔离，使得人可以沉浸于虚拟环境中。透射式头戴显示器又叫增强现实头戴显示器（AR HMD），是在通过光学结构观察到真实环境的基础上，将虚拟物体叠加在真实环境上的显示设备。

### 5.1.1 沉浸式头戴显示器的发展状况

虚拟现实头戴显示器按照结构分为全沉浸式和半沉浸式结构。目前大多数产品采用的是全沉浸式结构、由单片式的镜片构成目镜系统，比如Oculus VR公司[1]主要的两款产品：Oculus Rift和Gear VR。Oculus Rift采用先进显示屏技术结合精准、低延时光点追踪系统，具有两个目镜，每个目镜的分辨率为 640×800，双眼的视觉合并之后拥有 1280×800 的分辨率，刷新频率为90Hz。并且具有陀螺仪控制的视角，使得沉浸感大幅提升，成品样机如图1所示。此外，还配有Oculus Touch，让双手也可以进行交互，如图2所示。Gear VR是配合三星[2]手机使用的一款虚拟现实头戴显示器，其可以兼容三星 GALAXY 多款旗舰智能手机，将手机卡入，虚拟现实体验即可呈现，如图3所示。



图1 Oculus Rift样机



图2 Oculus Touch



图3 Oculus Gear

HTC VIVE[3]也是选用全沉浸式结构的一款虚拟现实眼镜，采用了SteamVR™追踪技术，空间定位追踪（ROOM-SCALE）可达4×3米，当使用者坐在房间内，可以转瞬移动到无尽的虚拟世界，或是规划专属的VR游玩区域，在其中自由走动。眼镜配置前置摄像头、双眼舒压设计，并且配有手柄给使用者以触觉反馈。显示屏采用了一块OLED 屏幕，单眼有效分辨率为 1200 x 1080 ，双眼合并分辨率为 2160 x 1200，刷新频率为90Hz。2K分辨率大大降低了画面的颗粒感，用户几乎感觉不到纱门效应。成品样机如图4所示。

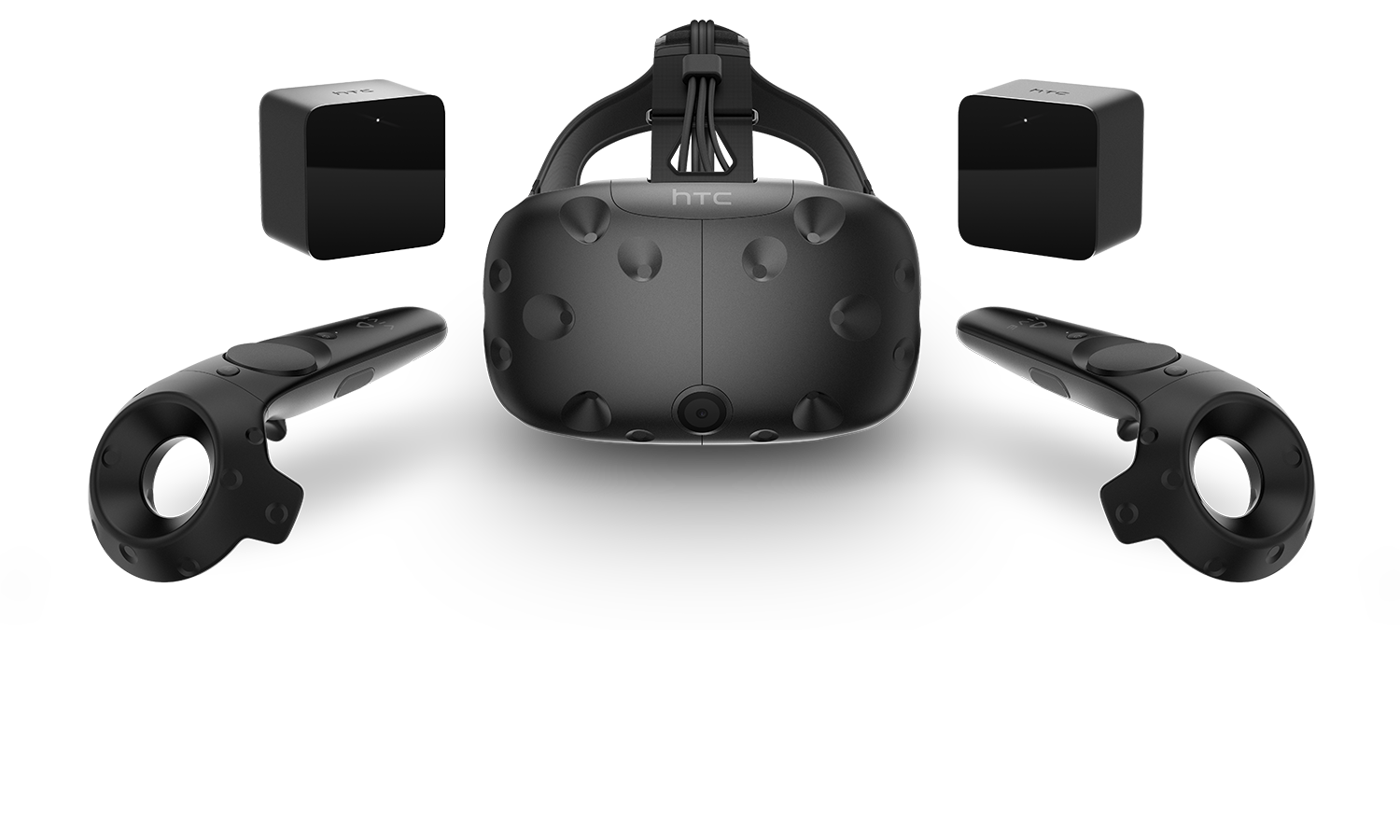


图4 HTC VIVE

Sony公司[4]的PlayStation®VR采用OLED显示屏实现全沉浸式显示结构，显示尺寸为5.7寸，并且结合了3D音频技术，沉浸感很高。成品样机如图6所示。



图6 Sony PlayStation®VR

### 5.1.2 透射式头戴显示器的发展状况

增强现实头戴显示器由于其光学系统需要光学穿透显示, 结构就显得复杂一些，所以它的结构形式就较多一些。最早受到大家关注的AR眼镜是谷歌眼镜[5]，其结构是一个潜望式的光学结构。图像从图像源出来经过三次的反射，然后进入到人眼。同时人眼透过半透半反的玻璃观看真实世界，虽然它比较小巧，但是它的视场角比较小，而且由于只有一个表面有光焦度，所以其解析度和视场角都略显不足。Google Glass主要结构包括，在眼镜前方悬置的一台摄像头和一个位于镜框右侧的宽条状的电脑处理器装置，配备的摄像头像素为 500 万，可拍摄 720p 视频。镜片上配备了一个头戴式微型显示屏，它可以将数据投射到用户右眼上方的小屏幕上。显示效果如同 2.4 米外的 25 英寸高清屏幕。使用该设备可以将智能手机的信息投射到用户眼前，通过该设备也可直接进行通信，如图7所示。



图13 Google Glass显示功能示意图

第二种结构是利用自由空间耦合面结构，目前结合一些手机的屏幕可以实现一个较大的视场角，较低成本的近眼呈现。Meta公司[6]的Meta 2就是采用的这种结构，结合一些手机的屏幕可以实现90°的大视场角，成本较低，不需要手柄，将设备连接到笔记本电脑上就可以与三维场景进行交互，但缺点是头戴显示器重量较大，成品样机如图14所示。



图14 Meta 2 成品样机示意图

为实现更大视场角的显示，自由曲面棱镜的结构被用于头戴显示器的设计中。自由曲面棱镜是有3个自由曲面的表面，自由曲面就是一个各个方向都不对称的一个结构，它引入了很多自由度，可以实现更大视场角更便携的一种近眼的呈现。耐得佳[7]的NED+ GLASS X2和悉见[8]X1S就是采用自由曲面棱镜的光学结构，自由曲面就是一个各个方向都不对称的一个结构，它引入了很多自由度，可以实现更大视场角更便携的一种近眼的呈现。在此基础上，北京理工大学提出了一种新型的自由曲面的拼接方案[9]，通过光学整体注塑成型，还有自由曲面多参数的特性，解决了拼缝明显、系统笨重、很难进行穿透式显示问题，获得了一个较好的拼接效果,如图17所示。同时探讨了自由曲面拼接的不同方案，为了获得更大的分辨率，同时保证一个较大的视场角。

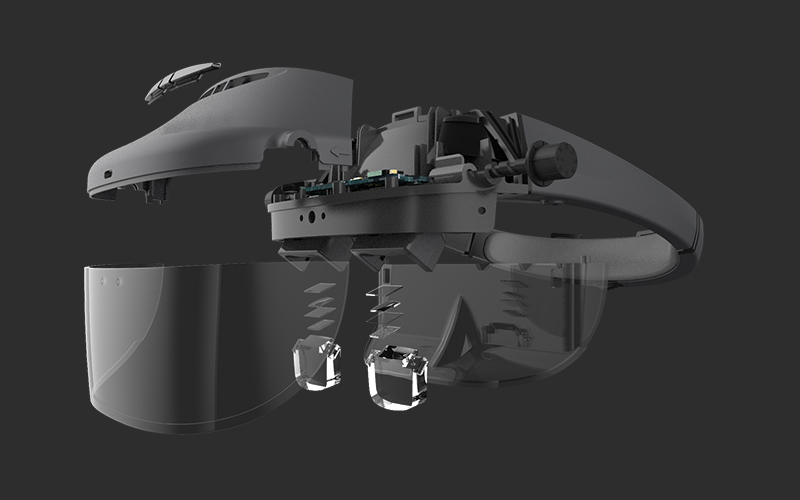


图15 耐得佳NED+ GLASS X2 增强现实眼镜拆解图



图16 悉见X1S



图17 拼接式近眼显示效果

自由曲面棱镜的结构虽然小巧，而且进行了一些优化，但是用户可能对这个小巧程度还不够满足，目前自由曲面的厚度大概也会在10mm左右。所以为了进一步让系统更加小巧更加薄，光波导也被引入了近眼显示的设计当中。光波导技术是通过光线在内部进行全反射，来增加光学长度，从而来减小光学厚度。这种光波导的近眼显示结构，光需要进入到光波导内，然后由光线在光波导内部的传输，最后从光波导耦出进入人眼。根据耦出端的不同，结构又分为四种。爱普生公司[10]的EPSON BT-300采用了曲面的反射进行耦出，使用双目透视型（See-through） Si-OLED光学引擎，对比度为100,000:1，分辨率为1280\*720，长达6小时的续航时间，而重量只有69g（BT-300）、119g（BT-350）。Optinvent ORA公司[11]，采用了自主提出的棱镜表面进行了耦出，可以使得出瞳更大；Lumus提出了多反射镜的方案，并对于产业化和杂散辐射进行了一些优化，获得良好的效果；衍射光学表面也被很多产品引入到近眼显示的设计当中，比如Hololens和Sony的AR产品。微软[12]HoloLens采用的是全息光学元件和内置传感器，可以实现头部移动时仍可以看到立体显示场景，并且有手势和语音的交互，如图18所示。

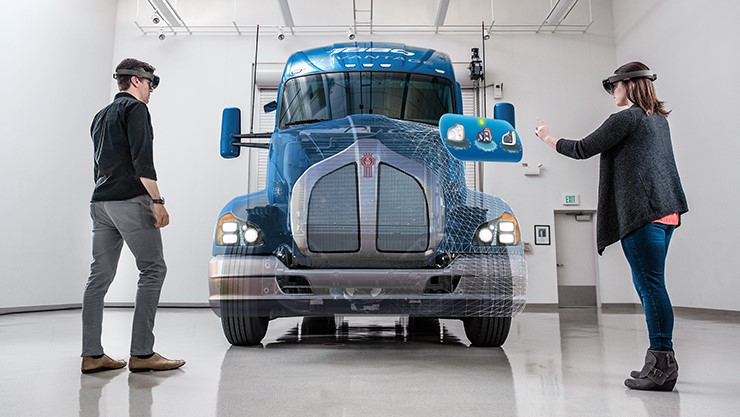


图18 HoloLens 功能示意图

### 5.1.3 头戴显示器未来的发展趋势

现在正在研究的头戴式显示器正在向小型化、大视场的高分辨率化以及真实化发展。

在小型化方面，北京理工大学在自由曲面研发基础上结合光波导的技术[13]，希望通过光波导技术可以使得镜片的厚度进一步降低，通过自由曲面棱镜增加它的优化参数，使得光学性能更好，从而达到最优的效果。搭建了基于这个原理的样机如图19所示，并对系统的杂散辐射进行了一部分优化，最后获得了一个比较良好的成像效果。另外，在研究过程中发现，近视或者是远视的用户在使用近眼显示设备的时候，需要同时佩戴自己的矫正眼镜，这样的话近眼显示设备就比较厚，而且佩戴起来比较复杂，基于这种想法，提出了一个结合了自由曲面的光学目镜以及矫正的近眼显示设备，使两镜片一体化，设备更小巧更方便。

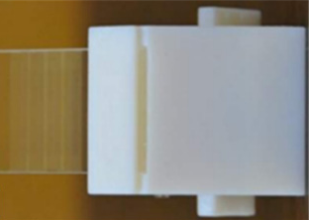


图19 基于光波导技术的原理样机

同时，Innovega公司提出了一个更好的想法，它想把近眼显示和隐形眼镜结合在一起，通过隐形眼镜增加一个微透镜，将显示屏上一个物体呈现到较远的一个位置，同时除了这个透镜之外的地方进行一些镀膜的处理，使屏幕上的一些显示并不能进入人眼当中，而且这块外界的物体通过其他部分呈现到人眼。这样系统更加小型化，但是目前它的量产性评估和生理性评估正在进行。

由于现在一些VR和AR设备的屏幕分辨率是一定的，随着视场角的增加，角分辨率也会降低，这个关系在学术上称为分辨率/视场角不变量。为了解决这种不变量，除了增加屏幕的分辨率，学者们还想到了一些其他方案，比如光学拼接的方案[14]，通过多通道光学系统，通过通道拼接来实现高分辨率，在保证了每个通道内的分辨率的基础上形成更大的视场角。

另外，在学术界和产业界一个比较重要的研究方面是构建真实感的近眼呈现。人在观察真实世界当中，人观察物体时，单目会调节聚焦到物体平面上的，同时通过双目视差感，它的辐辏也是感觉形成到这个平面上。但是对于目前商用化的近眼显示，在空间当中，只有一个呈现平面，使得单眼聚焦在这个平面上；但是为了产生3D感，通过视差的方式，辐辏则不在这个平面上，这种聚焦和辐辏的不一致会导致佩戴VR眼镜和AR眼镜感觉不舒适，无法长期佩戴。针对这个问题，研究者提出了很多解决和缓解的方案。亚利桑那大学华宏教授团队在近眼显示设备中插入了液体透镜结构[15]，通过改变液体透镜的电压，从而改变液体透镜的焦距，从而使得成像平面在空间中形成多个，通过这种方案来缓解辐辏和聚焦的不一致。北京理工大学使用两片胶合自由曲面目镜[16]，通过胶合位置使用半透半反膜使得用户可以同时看到两个屏幕，控制结构位置使得呈现在不同的深度上，来实现多深度重建方案并且获得较好的成像效果。斯坦福大学Wetzstein团队提出了多层显示屏的近眼显示方案[17]，通过计算机一些优化方案充分利用多层显示屏上的像素，来构建进入人眼的稠密的光场，但是他们算法较为复杂，暂时无法进行实时渲染。英伟达公司Lanman博士使用集成成像的方法[18]，构建稠密的光场进入到人眼当中，而且它的光场每一条光线，都严格地跟像素一一对应，并且他们使用了GPU技术实现了实时的光场渲染。北京理工大学通过使用微透镜阵列[19]形成光场，同时利用光学目镜的放大作用，实现一个大视场角可穿透式光场呈现，我们完成了基于该原理的样机，并通过将相机在出瞳的位置模拟人眼进行了拍摄，获得前后聚焦的景深效果。

参考文献：

[1] Oculus VR公司官方网站：https://www.oculus.com/.

[2] 三星公司官方网站：https://www.samsung.com/cn/.

[3] HTC公司官方网站：https://www.htc.com/cn/.

[4] Sony公司官方网站：http://www.sony.com.cn/.

[5] 初见公司官方网站：http://tech2ipo.com/.

[6] Meta公司官方网站：http://www.metavr.com/.

[7] 耐得佳公司官方网站：http://www.nedplusar.com/.

[8] 悉见公司官方网站：http://xiijan.com/.

[9] Song W , Cheng D , Deng Z , et al. Design and assessment of a wide FOV and high-resolution optical tiled head-mounted display[J]. Applied Optics, 2015, 54(28):E15.

[10] 爱普生公司官方网站：http://www.epson.com.cn/.

[11] Optinvent ORA公司官方网站：http://www.optinvent.com/.

[12] 微软HoloLens官方网站：

https://www.microsoft.com/zh-cn/hololens/.

[13] Dewen Cheng, Yongtian Wang, Chen Xu, Weitao Song and Guofan Jin. Design of an ultra-thin near-eye display with geometrical waveguide and freeform optics [J]. Optics Express, 22(7), 20705-20719, 2014.

[14] W. Song, D. Cheng, Z. Deng, Y. Liu, & Y. Wang, “Design and assessment of a wide-FOV and high-resolution optical tiled head-mounted display”, Appl. Opt. 54(28), E15-E22 (2015)

[15] S. Liu, H. Hua, and D. Cheng, “A novel prototype for an optical see-through head-mounted display with addressable focus cues,” IEEE Trans. Vis. Comput. Graph. 16(3), 381–393 (2010).

[16] D. Cheng, Q. Wang, Y. Wang, G. Jin, "Lightweight spatial-multiplexed dual focal-plane head-mounted display using two freeform prisms." Chin. Opt. Lett. 11(3), 031201 (2013).

[17] F. C. Huang, K. Chen, and G. Wetzstein, “The light field stereoscope: immersive computer graphics via factored near-eye light field displays with focus cues,” ACM Trans. Graph. 34(4), 60 (2010).

[18] D. Lanman and D. Luebke, “Near-eye light field displays,” ACM Trans. Graph. 32(6), 1–10 (2013).

[19] W. Song, Y. Wang, D. Cheng, and Y. Liu, “Light field head-mounted display with correct focus cue using micro structure array,” Chin. Opt. Lett. 12(6), 060010 (2014).

## 5.2 视频源质量参数

视频源质量参数主要是对视频源的分辨率、帧率等参数进行评估。需要说明的是，因超高清及VR视频在传输过程中可能产生画质损耗，且在终端播放过程中可能受限于终端设备的解码及显示性能，可能存在视频播放质量较源质量下降的情形。

表格 5‑1 VR视频内容分辨率及帧率评分参考值

|  |  |  |  |
| --- | --- | --- | --- |
| **分辨率** | **参考分值（VR）** | **帧率** | **参考分值（VR）** |
| 6480P (16K) | 100 (11520 x 6480) | 120 | 100 |
| 4320P (8K) | 90 (8K x 4K) | 90 | 90 |
| 2160P (4K) | 80 (4K x 2K) | 60 | 75 |
| 1440P (2K) | 60 | 30 | 50 |

分辨率和帧率是用以衡量视频源质量的两项重要参数。表格5-1中体现了传统评估方式下VR视频分辨率及帧率的评分参考值。一般认为评分100代表极致体验；80以上代表可达到用户喜爱的优质体验，其中80-90代表用户体验的舒适区，超过90则代表体验优秀，服务可能给用户带来意外的惊喜；60-80代表可被用户接受的体验；60以下代表体验无法被用户接受。

实际上，关于视频分辨率，在VR环境中，更适合采用每度像素数量（pix/deg）的概念来取代用于图像的每像素数量或用于显示的每英寸像素数量，以体现如何将视场内的部分精确地呈现给用户的眼睛。一般说来，20像素/度可确保内容的分辨率始终高于显示分辨率。

VR视频的高感知质量需要支持球面度空间图像的高视频分辨率（全球面的立体角角度为4π）。若发送的视频源分辨率为4K（3840 x 2160），头显平均提取率仅约为4K图像的12-14％，需将其升高至1080分辨率后进行显示，因此观众感知质量较低。当支持整个4π球面度空间的图像视频分辨率需达到2.828（即约等于3）倍于4K分辨率时（即11520 x 6480），才可为观众带来好的体验。

除此之外，在创建VR内容时，通常需要在一组摄像机设备上进行采集，然后再进行拼接处理。拼接步骤是影响质量的非常重要的步骤之一。拼接包括将来自不同摄像机的视图合并到一个视图中。在两个相邻视图的边界可能发生的任何故障，都很容易被人眼看到，并大幅降低VR内容的感知质量。这个步骤通常由软件自动执行，并且有多种算法，处理成本和质量结果不同，重要的是在接收方不会引入额外的拼接错误。

## 5.3 头戴显示器参数

5.1章节中描述了视频源质量的几项特征参数。与此同时，VR视频服务的体验与终端设备的解码及显示性能密切相关，本章节将做相关描述。

上文中提到每度视频像素的数量是高质量VR的重要指标，20像素/度可确保内容的分辨率始终高于显示分辨率，给用户带来高质量的VR感知。但尽管如此，与传统电视机相比，这个分辨率仍然只有1080p感知的一半，4K的四分之一。如表格5-2所示，目前尚未有头显能够显示40pix/deg，Oculus Rift的最高分辨率为18pix/deg。

关于帧率，良好的VR业务体验要求PC头显帧率不低于90fps，移动设备帧率不低于60fps。除了以60fps运行的Samsung Gear VR之外，现在大多数头显都能提供90fps，部分头显可以提供120fps的最大帧率。此外，90fps帧率需提供足够低的延迟，以防止由于不正确的MTP延迟造成的不适感。关于MTP延迟将在后文中做阐述。

表格 5‑2 主流VR头显参数参考值

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **设备** | **分辨率** | **最大帧率** | **视场角（水平&垂直）** | **每度像素** |
| 三星Gear VR | 2560 x 1440 | 60fps | 最大96° | 水平13  垂直15 |
| Oculus Rift | 2160 x 1200 | 90fps | 水平67°-96°  垂直66°-107° | 水平11-16  垂直11-18 |
| HTC Vive | 2160 x 1200 | 90fps | 水平76°-100°  垂直77°-113° | 水平11-14  垂直10.5-15.5 |
| PlayStation®VR | 1920 x 1080 | 120fps | 最大100° | 水平9.5  垂直11 |
| 电视（全高清） | 1920 x 1080 | 30fps | 平均：水平45°  垂直25° | 水平43  垂直43 |
| 电视（4K） | 3840 x 2160 | 120fps | 平均：水平45°  垂直25° | 水平86  垂直86 |

头显设备镜片与视频源之间的图像也是影响体验的因素之一。正常物体通过透镜观看后会呈现卷曲效果，亦可称为枕形畸变。基于头显透镜的畸变系数可以通过数学模型比对真实图像与失真图像得到。

## 5.4 沉浸体验质量参数

章节5.1和5.2重点围绕视频分辨率和帧率对VR视频服务体验的影响做了阐述。但人眼对感知深度信息非常敏感，包括双目视差、运动视差、遮挡和聚散度等几种方式，VR要实现让人眼无法分辨虚拟和现实的目标，必须首先满足人眼在自然状态下观看外部世界的条件。

当人在自然状态下观察目标时，眼睛在不断变焦，并且能够通过前庭视反射消除画面切换时的抖动和模糊。这是因为在人眼的周围有无数的光线，眼睛通过不断地改变焦距，采集来自不同方向和位置的光线，聚焦在不同的平面。加上头部的运动，人可以在很大的范围内不断地聚焦在所关心的物体上。因此为了观察世界，人需要时刻感知到周围来自不同位置和方向的光线，这些光线的集合便是光场。

目前市面上很多VR/AR公司只能人工合成360°全景图片或视频，其内容首先缺乏双目立体视觉效果。即便采用立体相机拍摄了三维空间信息，获得了双目立体视差，也仍然缺乏运动视差，不能形成走入走出的沉浸感。通过在VR内容中提供交互式视差，用户可以像在真实环境中那样在前景中查看背后的物体，而当视差不存在时，相同的图像似乎会粘在背景上。前景中的物体提高了沉浸感，同时前景中的物体应足够远以防不适感，并且具有背景的交互视差应存在于这些物体上。但研究表明，如果它们距离使用者过近（距离小于3米），将会导致观看不适感。作为虚拟现实领域最具有变革性的新技术，光场能同时提供双目视差、运动视差、选择性聚焦三方面的视觉信息，从传统的2D平面图像提升到4D光场图像，大大提升沉浸感和逼真度，模糊虚拟和现实的边界。

除视频质量外，关于VR的音频质量，经常使用由不同类型的音频编码提供的浸入度。通俗来说，VR音频需要能带来临场感的声音，即还原用户在现场实际能听到的声音。前后声音是否能很好地区分、混响大小模拟是不是足够好、戴耳机和没戴耳机的效果是否相似等等，都是决定一段音频是否完善的要素。VR音频需至少做到跟随头部运动，以便静止的声源在虚拟环境中静止不动，而不与用户的头部绑定。

与此同时，用户沉浸式交互强烈依赖外部设备的动作跟踪。头部动作跟踪是当下VR技术的一种重要手段，主要依靠内/外置陀螺仪传感器来实现。其中，三轴陀螺仪（Gyroscope）可以在xyz三轴上检测角速度的变化；六轴陀螺仪（Accelerometer + Gyroscope）可以在xyz三轴上检测加速度与角速度的变化；九轴陀螺仪（Accelerometer + Magnetic + Gyroscope）在xyz三轴上检测加速度、地磁场、角速度三者的数据变化。六轴与九轴陀螺仪均需要融合算法。

## 5.5 交互体验质量参数

### 5.5.1 交互体验参数

传统视频业务中的交互体验可以按照业务类型做区分，其中直播业务关注的是频道切换时间，点播VoD业务关注的是点播的加载时间。表格5-3中提供了直播频道切换质量的参考评分和点播加载质量的参考评分。

表格 5‑3 交互质量参考评分

|  |  |  |  |
| --- | --- | --- | --- |
| **直播切换时长** | **参考分值** | **点播等待时长** | **参考分值** |
| <100ms | 100 | <0.5s | 100 |
| 500ms | 80 | 1s | 80 |
| 1000ms | 60 | 5s | 60 |

除以上视频业务常规交互指标外，Motion-to-Photon即MTP延迟是影响VR沉浸感和交互感的决定性因素。MTP指从用户运动开始到相应画面显示到屏幕上所花的时间，目前定义的最大值是20毫秒。Motion-to-Audio是指从用户运动开始到听到对应声音所花的时间。与视频类似， Motion-to-Audio，即从用户运动开始到听到对应声音所花的时间应能立即相应任何用户交互。Motion-to-Audio延迟应与Motion-to-Photon延迟保持同步，最大值20毫秒。

### 5.5.2 强交互体验业务指标

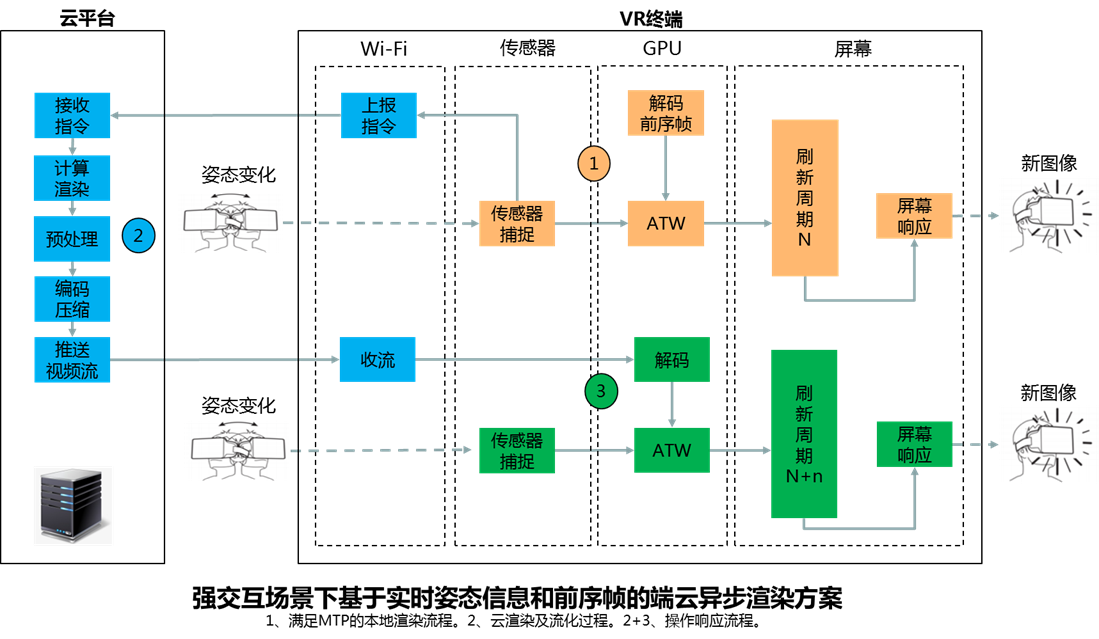


图5-4 强交互体验

下表是强交互体验的业务指标。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **规格分类** | **分类** | **规格项** | **参考** | **定义&拆分** |
| 强交互业务体验指标 | 视频质量 | 分辨率（FOV） | 3K FOV 2880x1600 |  |
| 帧率 | 72FPS |  |
| 视场角 | 120 | 人眼可观察部分的边缘与观察点连线的的夹角。 |
| 色深(bit) | 8bit |  |
| 码率 | 30~50Mbps |  |
| 编码方式 | H.264 |  |
| 立体视觉 | 3D |  |
| 音频质量 | 音频码率 | 128kbps |  |
| 音频编码 | AAC |  |
| 源声道数量 | 2声道 | 声道数也就是录制时音源数量或回放时相应的扬声器数量 |
| 音频采样率 | 48KHz | 音频采样率是指录设备在一秒钟内对声信号的次数 |
| 终端质量 | 终端分辨率 | 1440x1600x2 | 屏幕支持的像素数 |
| 终端视场角 | 101、110 |  |
| 终端刷新率 | 90Hz | 刷新率是显示器每秒从图形处理单元 GPU获取新图像的次数 |
| 终端声道数 | 双声道 | 音频需要被重新计算并更新方位来还原用户真实空间感 |
| 音视频同步时差 | 同步 | 音视频时间差 |
| 反畸变 | 枕型反畸变 | VR内容通过透镜观看时，可以带来更大的视场角，但也会让图像产生畸变 |
| 解码能力 | H.264 |  |
| 解码时延 | 9~15ms |  |
| 异步扭曲时延 | <5.5ms |  |
| 屏幕响应时延 | 5~6ms | 和屏幕类型相关 TFT-LCD，AMOLED |
| 定位质量 （NOLO） | 交互自由度-头盔 | 6DOF |  |
| 交互自由度-手柄 | 6DOF |  |
| 定位信息采集精度-头盔 | ±2mm |  |
| 定位信息采集精度-手柄 | ±2mm |  |
| 姿态信息采集精度-头盔 | 1.00E-04 |  |
| 姿态信息采集精度-手柄 | 1.00E-04 |  |
| 位置信息采集频率-头盔 | 120Hz |  |
| 位置信息采集频率-手柄 | 120Hz |  |
| 姿态信息采集频率-头盔 | 1000Hz |  |
| 姿态信息采集频率-手柄 | 1000Hz |  |
| 信号传输延迟 (2.4G WiFi) | 2~3ms |  |
| 信号处理延迟 | 2~3ms |  |
| 网络质量 | 网络丢包率 | <1E-5 |  |
| 网络传输时延 | <20ms |  |
| 观看质量 | 卡顿次数（TCP/UDP） | 0 | 当相邻有效帧（对应不同的观看视角）解码时间间隔较大时会出现卡顿 |
| 卡顿时长（TCP/UDP） | 0 |  |
| 花屏次数（UDP） | 0 | 统计每秒受损的帧数，考虑关键帧的影响，考虑统计难度折算成 视频丢失率 |
| 花屏时长（UDP） | 0 | 统计每秒受损的帧数x帧间隔 |
| 交互质量 | 程序加载时长 | 30s |  |
| 头动感知延时（MTP） | 20ms | 终端（头部姿态感应 1~2ms +信号处理 2~3ms+（帧刷新周期1000/90ms）∪（异步扭曲5~6ms）+ 屏幕响应 5~6ms）<20ms 异步扭曲在帧刷新周期内完成。 |
| 云渲染及流化时延 | 70ms | 云端处理 30ms+网络传输 20ms； |
| 操作响应延时 | 100ms | 终端上行：手柄定位 1~2ms +指令传输 2~3ms+驱动程序处理 1~2ms； 云端处理：30ms； 终端下行：【（解码时延 9~15ms+异步扭曲 5~6ms）∪（帧刷新周期同步 1000/90）】+ 屏幕响应 5~6ms； 网络传输：20ms； |
| 黑边越限次数（ATW） | 0 | ATW视角与有效帧视角之间的差值超过一定阈值的次数 |
| 最大黑边角度（ATW） | 0 | ATW视角与有效帧视角之间的差值的最大值 |

## 5.6 网络KPI指标

网络的传统KPI 指标，例如覆盖、干扰、在线用户数、小区负荷等不能直接衡量视频性能。一般说来，视频传输质量的网络因素主要通过延时、丢包、抖动来体现。

采用UDP协议传输时，最主要的影响因素是丢包。根据解码器对丢包的处理机制不同，现象可能是冻结，也可能是花屏。采用冻结机制时，解码器遇到丢包，则冻结住画面，等待下一个完整的I帧后再继续，用户侧体现为画面冻结；采用忽略机制时，则忽略丢包继续解码，用户侧体现为花屏。

采用TCP协议传输时，由于采用可靠传输，在带宽足够且稳定的条件下，少量丢包可以通过TCP重传得到补偿。当分片下载不及时赶不上播放速度时，缓冲区会耗尽（缓存下溢），从而造成播放缓冲现象。

用户对于冻结/花屏和缓存下溢的体验，既体现在次数，即频繁程度，也体现在持续时长，即严重程度。

表格5-4 用户对于冻结/花屏体验的参考评分

|  |  |  |
| --- | --- | --- |
| **冻结/花屏次数/每小时** | **冻结/花屏时长/每小时** | **参考评分** |
| 0 | 0 | 100 |
| 1 | <0.5s | 80 |
| 2 | <2s | 60 |

表格5-5 用户对于缓存下溢体验的参考评分

|  |  |  |
| --- | --- | --- |
| **缓存下溢次数/每小时** | **缓存下溢时长/每小时** | **参考评分** |
| 0 | 0 | 100 |
| 1 | <5s | 80 |
| 3 | <10s | 60 |

在网络侧评估缓存下溢评分时，考虑的主要因素是带宽，TCP重传率和时延。其中足够且恒定的带宽对保证播放质量的作用最大，其次是TCP重传的影响。

## 5.7 业务KQI指标

用户的视频业务体验是多维度的，包括视频内容质量，网络传输质量，以及用户与业务、终端与视频平台的交互质量体验等。对用户的视频业务体验进行评估的体系一般需要实现三个目标：能对视频用户体验进行客观的测量，能指示影响视频用户体验的原因，并能为视频网络规划提供参考。与此同时，评估体系需要能够覆盖视频业务的端到端过程，能够从内容源质量、网络传输质量、业务平台、终端四个方面进行评估。因此，VR用户服务体验的质量评估建议由一个体现视频体验质量的综合指标VR-QoE，一组支撑VR-QoE的规范化KQI指标，以及如章节5.1-5.5中与KQI相关联的一系列KPI构成。

如图5-1所示，基于构成VR视频服务用户体验的五项KPI指标：内容质量、传输质量、头显质量、交互质量和沉浸质量，可以形成视频体验、交互体验、沉浸体验三项核心KQI，并以此衡量最终的服务体验质量。

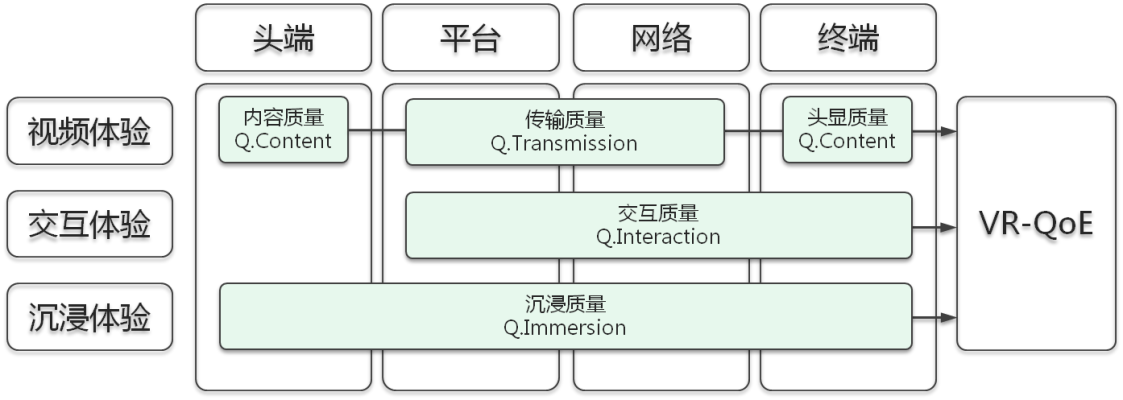


图5‑1 VR-QoE的主要因素

## 5.8 Cloud VR业务时延要求

对于Could VR视频业务，参考4K视频的分析，网络时延在20~40ms即能满足要求，不同的时延只是对画面加载时间产生影响。但是对于Could VR强交互业务，为了保证用户的交互感和愉悦感，对时延有如下三个要求：

1.MTP时延要求≤20毫秒（端云异步渲染方案下MTP时间不再依赖网络和云渲染，可由终端保证）。

2.云渲染及流化时延要求为30~70毫秒。

2.1在起步阶段，云渲染及流化时延建议≤70毫秒，画面黑边和质量的劣化在可接受范围。

2.2 在舒适体验阶段，云渲染及流化时延建议≤50毫秒，此时基本消除黑边现象。

2.3 理想体验阶段，云渲染及流化时延建议≤30毫秒，此时位置移动时画面的扭曲将让人无法察觉。

3.操作响应时延要求≤100毫秒。 操作响应时延由云渲染及流化时延，加上终端的二次渲染、异步扭曲和刷新呈现构成。操作响应时延减去云渲染及流化时延的要求70ms后，剩余30ms，大于终端对二次渲染、异步扭曲和刷新呈现的MTP时延要求。 因此只要保证云渲染及流化时延要求，就能同时满足操作响应时延要求。

# VR视频服务用户体验评估算法

用户的视频业务体验是多维度的。比如传统视频体验一般包括视频内容质量、网络传输质量、终端显示质量，以及用户与业务、终端与视频平台的交互质量体验等。相较传统视频，在VR视频服务中，沉浸感也成为衡量VR视频服务质量的重要参数。本章节将综合多维指标，提出面向VR视频服务用户体验的评估算法。

影响VR服务体验的因素可分成客观和主观两方面。客观因素包括：视频源指标，例如分辨率和帧率；网络指标，例如花屏、卡顿等；交互时延，例如MTP等；以及终端设备限制等客观因素。主观因素主要包括用户的身心状态、视觉系统的生理构造、对业务质量的期望值、业务体验经历以及在使用业务过程中的情绪等难以量化的因素。

现有的VR服务体验评价方法多以主观质量评价为主，通常通过一组用户群体直接对业务的QoE进行评价。主观评价方法至今仍普遍应用于音视频业务的QoE评价上，比如ITU-T E.800、ITU-T P.810、P.833、P.910和ITU-R BT.500中对此类主观评价都有相关定义。

针对视频源质量、网络质量等则更适合采用客观质量评价方法，即利用评估模型来自动计算业务质量。本章节中所探讨的VR服务体验综合算法属于客观质量评价方法，旨在通过量化的指标和算法最大程度的还原用户对于VR服务体验的满意程度。而完整的VR服务体验需考虑将主观与客观质量评价相结合。

## 6.1 总体模型综述

VR用户服务体验质量评估算法建议由体现视频体验质量的综合指标VR-QoE来体现。VR-QoE反映终端用户对VR视频服务整体的主观接受程度，需要通过量化的方法来呈现终端用户的体验和感受，并反映当前VR服务质量与用户期望之间的差距。

VR-QoE的量化方法建议采用定量的方式，即由具体的业务类型及其评价参数决定。如章节5.6中所述，VR-QoE可以拆解为一组支撑VR-QoE的规范化KQI指标Qx，以及与Qx相关联的一系列KPI构成。综合性的VR-QoE指标反映用户对VR视频业务的整体接受度，通过对不同Qx进行组合计算得到，Qx可规范化为比如0~100的分值，部分Qx可通过探针等设备采集上报的性能指标计算得到。

表格6‑1 VR用户服务体验质量指标:

|  |  |  |
| --- | --- | --- |
| **指标** | **指标描述** | **指标特征** |
| VR-QoE | 用户VR服务体验质量综合指标，根据Q.x计算 | 反映VR视频用户的整体体验质量 |
| KQI | 支撑VR-QOE的规范化KQI指标，根据KPI计算 | 用于计算VR-QOE |
| KPI | 与KQI相关的一系列KPI，采集自网络、设备等 | 用户评估KQI指标并对相应问题进行分析 |

图5-1中体现了影响VR服务体验的KQI和KPI之间的映射关系。其中，视频体验建议由内容质量、传输质量和头显质量构成；交互体验重点由交互质量体现；沉浸体验由沉浸质量体现。

由各项KQI构成的VR服务质量体验计算模型建议为：

其中，代表视频体验的系数，代表交互体验的系数，代表沉浸体验的系数，。

## 6.2 模型三大模块综述

如章节6.1所述，VR服务质量体验计算模型由视频体验、交互体验、沉浸体验三大模块组成。

### 6.2.1视频体验质量Q1

视频体验质量Q1即QVideo由内容质量、传输质量和头显设备质量三项KPI共同构成。在三项KPI对视频体验质量影响程度相同的情况下，Q1建议为：

其中，内容质量、头显设备质量和传输质量可通过专业分析设备和探针上报获得。同时，因用户观看到的VR视频质量同时受到源视频质量参数、压缩效果、渲染渠道、终端质量参数等多方面因素综合影响，参数之间难以简单叠加，所以建议以用户观看到的视频质量QWatching来体现QContent和QHMD的综合影响效果：

关于QWatching的测量，建议可以采用Visibit的基准视频作为参照，用以判断用户观看到的视频是否存在质量丢失，可测试的参数包括视场、帧率、画面细节展示程度、动态范围等。

Mode0：

Mode0是基于码流层的轻量级模型，主要衡量视频显示质量。显式包含*BitRate, ScreenSize ,PPI*变量，其余变量在不同的应用场景下（分辨率、编码器类型）为定值。

Model1：

Mode1是基于码流层的复杂模型，需要从编码数据包及比特流中采集视频帧关键编码信息（详见5.3.5），同时衡量视频播放质量和视频压缩对于视频源质量的损伤情况。

Mode2：

Mode2是基于图像层模型，通过对视频画面关键质量指标的衡量，刻画视频不同纬度的质量。Mode2（详见5.3.6）根据图像的模糊度，块效应，对比度，噪点度，色彩丰富度和曝光度等指标评价视频源质量。

#### 6.2.1.1 Paramedic Model (mode 0)

注：Mode0 参数统一由字母{ci}表示。

Qs为视频体验质量得分，主要衡量视频显示质量，其取值范围 [1，5]，公式为：



其中，Qs’ 表示显示质量，即在固定 PPI 和屏幕尺寸条件下（如720P视频在42’电视播放）能够取得的最高质量得分； 表示相应最低分值，本文中取 1.0。



在不同尺寸的屏幕显示视频时，需要用有效显示*PPI*及屏幕尺寸（*ScreenSize*）来评估在特定设备上播放视频时用户的体验质量。有效显示*PPI*由原视频的分辨率，屏幕分辨率以及屏幕尺寸计算得到。



以上的系数均为大数据统计所得。根据不同的编码方式有不同的对应系数取值。

#### 6.2.1.2 Bitstream Model (Mode 1)

注：Mode 1 参数由字母{ci }表示。

Qs 为视频体验质量得分，综合考虑视频显示质量和视频编码质量，其取值范围 [1，5]，视频部分的质量公式为：



Qs’ 表示显示质量，在不同尺寸的屏幕显示视频时，需要用有效显示*PPI*及屏幕尺寸（*ScreenSize*）来评估在特定设备上播放视频时用户的体验质量，公式引用（0-0-2）；



Mode 1 视频编码质量*Qcod* 由比特流中采集视频帧关键编码信息计算获得，衡量视频压缩对与视频源质量的损伤情况，可以表示为如下函数映射关系：



视频编码质量综合衡量了编码过程中关键帧率因子关键帧率因子*kfr*、量化因子*QPimp*和编码复杂度因子*cpximp*：



*kfr*为关键帧率因子，通过I帧之间平均距离*d*，和视频帧率*fr*来衡量I帧平均间隔对视频质量的影响：



量化因子主要衡量视频量化过程中造成的质量损伤，其计算过程为：



编码复杂度通过当前帧采用skip帧间预测模式下块的跳过比例*SkipRatio*来衡量。*SkipRatio*是基于帧内编码单元的帧级统计量，可以兼容多种编码标准。



上式中，*brv*视频码率，*AvgByteI*是I帧的平均大小(字节为单位)。

以上的系数均为大数据统计所得。根据不同的编码方式有不同的对应系数取值。

#### 6.2.1.3 Hybrid Model (Mode 2)

1）块效应Blockings：块效应是指在编码过程中造成的块边界不连续的情况，块效应严重时，视频在将出现明显块状缺陷，影响视觉效果，降低观赏感受。

原理公式





具体计算过程：

将一帧视频分为LxL块, wn，其中n为块下标 ，

|  |  |  |
| --- | --- | --- |
| 序号 | 公式 | 复杂度 |
| (1) | 计算每一分块最后一列与倒数第二列像素灰度值差值绝对之和verticalInnerSum | O(*M\*N*) |
| (2) | 计算每一分块最后一列与下一分块第一列像素灰度值差值绝对之和verticalOuterSum | O(*M\*N*) |
| (3) | 计算每一分块最后一行与倒数第二行像素灰度值差值绝对之和  horizontalInnerSum | O(*M\*N*) |
| (4) | 计算每一分块最后一行与下一分块第一行像素灰度值差值绝对之和horizontalOuterSum | O(*M\*N*) |
| (5) |  |  |
| (6) |  |  |

其中，

*M*：第*n*帧横向像素个数

*N*：第*n*帧纵向像素个数

：视频序列的帧数

2）模糊度Blurriness：模糊度是指对图像模糊程度的衡量，模糊是指基于图像像素灰度的梯度幅度变化，该变化可以表征图像边缘信息，当梯度幅度过小时，该条边缘将不够清晰明显，影响视觉效果，降低观赏感受。

原理公式 



具体计算过程：

视频序列复杂度：O(*f\*M\*N*)

将一帧视频分为LxL块, wn，其中n为块下标 ，

|  |  |  |
| --- | --- | --- |
| 序号 | 公式 | 复杂度 |
| (1) | 对每一块wn，估算标准差 |  |
| (2) | 计算视频帧每一像素的水平方向及垂直方向梯度如下，形成水平方向，垂直方向梯度矩阵 | O(*M\*N*) |
| (3) |  | O(1) |
| (4) | 通过如下方式判断像素点是否属于边缘：  ,则标记当前*f*(i,j),为边缘 | O(*M\*N*) |
| (5) | 通过如下方式计算边缘的宽度： | O(*M\*N*) |
| (6) |  | O(1) |

其中，*f*：第*i*帧像素点(*m,n*)的灰度值

*M*：第*n*帧横向像素个数

*N*：第*n*帧纵向像素个数

*edgewidth：*边缘宽度

*edgewidth：*边缘数量

：视频序列的帧数

3）对比度Contrast：合理的对比度可以显示生动、丰富的色彩，展现更多的细节、更好的清晰度以及灰度层次。然而对比度过高则视频将产生失真感，对比度过低画面则将表现为灰蒙蒙。不合理的对比度将影响人眼的主观感受

原理公式  O(*M\*N*)

具体计算过程：

每一帧复杂度：O(*M\*N*)

视频序列复杂度：O(*f\*M\*N*)

其中，

*M*：第*n*帧横向像素个数

*N*：第*n*帧纵向像素个数

：第*n*帧的图像对比度

4）噪点度 Noise：噪点度定义对像素色度值的浮动的衡量，该种浮动对图片整体质量无正影响且无固定规律。在被压缩视频中通常存在多种噪点类型。其中最常见的是量化噪声及蚊式噪声。其中量化噪声主要由于对像素值的量化引入，其分布具有小范围随机特性，且在整体图像上分布不统一。蚊式噪声主要表现为物体边缘的噪声以及运动纹理干扰。

原理公式：3\*3大小的窗口卷积 及

具体计算过程：

每一帧复杂度：O(*M\*N*)

视频序列复杂度：O(*f\*M\*N*)

将一帧视频分为LxL块, wn，其中n为块下标 ，

|  |  |  |
| --- | --- | --- |
| 序号 | 公式 | 复杂度 |
|  | 对每一块wn，估算标准差 |  |
|  | 计算各块标准差统计直方图 |  |
|  | Find |  |
|  |  | O(1) |
|  |  | O(1) |
|  | Find the blocks  whose |  |
|  |  | O(M\*N) |
|  |  | O(M\*N) |

其中，\*：与每帧图像做卷积的算子

{wn}：提取出的符合要求的图像块集合

：第*n*帧像素点(*x,y*)与算子做卷积后的计算值

：第*n*帧图像的噪点度

，与MOS的分段对应关系如下表：

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Qs\_i  Metric | 1 | 2 | 3 | 4 | 4.5 | 备注 |
| Blockiness | [0, 0.3) | [0.3, 0.6) | [0.6, 0.8) | [0.8, 0.9) | [0.9, 1.01) | min=0，max=inf，数值越大，块效应越小 |
| Blurriness | [50, 70) | [30, 50) | [10, 30) | [5, 10) | [0, 5) | min=0,max=70，数值越大,模糊度越大 |
| contrast | [0, 15)  &  [120, 100) | [15, 30)  &  [80, 100) | [30, 40)  &  [60, 80) | [40, 45)  &  [55, 60) | [45, 55) | min = 0，max= 120，数值越大，对比度越高 |
| noise | [20, 30) | [10, 20) | [5, 10) | [3.5, 5) | [0, 3.5) | min = 0，max= 30，数值越大，噪点度越高 |

5）色彩丰富度Color：色彩越丰富，能够提供的量化的图像信息就更多。我们使用了一种基于颜色统计直方图的色彩丰富度指标。

原理公式：+

具体计算过程：

每一帧复杂度：O(*M\*N*)

视频序列复杂度：O(*f\*M\*N*)

将彩色图像的RGB值首先被转换为改进的色调（Improved Hue）、亮度（Luminance）和饱和度（Saturation）等信息，并计算色彩丰富性评分C。

|  |  |  |
| --- | --- | --- |
| 序号 | 公式 | 复杂度 |
| (1) | 计算亮度，并划分等级，用做上标记 | O(1) |
| (2) | 计算饱和度 | O(1) |
| (3) | 计算色调 | O(1) |
| (4) | 对每一个亮度等级，应用 | O(M\*N) |
| (5) | 计算图片的色调均值（Hue Mean） | O(1) |
| (6) | 计算每个亮度等级中的均值长度，归类像素点到子集中 | O(M\*N) |

其中，

*M*：第*n*帧横向像素个数

*N*：第*n*帧纵向像素个数

：视频序列的帧数

M为有效的色调均值的数目

每个子集中的有效色调均值的个数

为x点处的色调

为x点处的亮度

为x点处的饱和度

6）曝光度Exposure：由前端相机曝光控制和色调映射算法共同决定，后者是HDR视频的热门研究领域。曝光度指标需要量化感知帧像素级的曝光质量。方案使用了一种基于亮度直方图范围检验的评价方法。

原理公式：

具体计算过程：

每一帧复杂度：O(*M\*N*)

视频序列复杂度：O(*f\*M\*N*)

将一帧视频分为LxL块, wn，其中n为块下标 ，

|  |  |  |
| --- | --- | --- |
| 序号 | 公式 | 复杂度 |
| （1） | 对每一块wn 计算平均亮度 |  |
| （2） | 计算单帧曝光率 | O(1) |
| （3） | 计算曝光度 | O(M\*N) |

其中，

*M*：第*n*帧横向像素个数

*N*：第*n*帧纵向像素个数

：视频序列的帧数

*La*：亮度明的边界。

*Lb*：亮度暗的边界

两套参数分别处理过度曝光和欠曝光

### 6.2.2交互体验质量Q2

针对VR的交互体验质量Q2，MTP是最为重要的衡量指标之一。MTP 延时大于20 ms便会导致用户体验到较为明显的眩晕。在理想状态下，总延迟应降至10ms 甚至更低。建议以20ms为用户可以接受的最低限度为衡量标准，将MTP时延进行量化并体现为分值。

除此之外，针对直播类业务，建议以频道切换时延作为交互体验的衡量指标；针对点播类业务，建议以起播时延作为交互体验的衡量指标。衡量标准可参考表格5.3中的体验评价参考分值。

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 变量 | 物理意义 | 计算所需变量 |
| 2.a | *QZapping* | 直播时频道切换质量 | 2.c |
| 2.b | *QLoading* | 点播初始加载质量 | 2.d |
| 2.c |  | 直播时频道切换时延 | 5.1.9 |
| 2.d |  | 点播时初始缓冲时延 | 5.1.11 |
| 2.e | *t* | 当前已播放时长 | 系统自动记录 |
| 2.f | *T* | 最大遗忘时长 | 系统参数 |

交互体验质量首先考虑频道切换和初始播放时延因素对体验的影响，其它因素可在后续进一步补充。

#### 6.2.2.1直播、

该部分衡量直播中基于频道切换事件的交互体验质量，可以定义为关键性能指标（频道切换时延）的函数映射关系如下。



基于频道切换时延的直播交互体验质量最终得分为：



= 

会话评分是在实时评分的基础上，根据遗忘曲线进行衰减：



#### 6.2.2.2、

该部分衡量点播中基于初始加载事件的交互体验质量，可以定义为关键性能指标（初始加载时延）的函数映射关系如下。



基于初始加载时延的点播交互体验质量最终得分为：



=

会话评分是在实时评分的基础上，根据遗忘曲线进行衰减：



### 6.2.3沉浸体验质量Q3

关于沉浸体验质量Q3，除一些客观指标如所使用的自由度（3DoF、3DoF+、6DoF等）、音视频频编解码标准等易于衡量外，仍有大量体验因子目前较难通过客观并量化的计算方式来呈现。主观测量仍是VR沉浸体验的经典测量方式。