# 云XR业务需求分析（定性）

为了给XR用户提供身临其境的用户体验，影响用户体验的关键因素值得被进一步研究（https://xinreality.com/wiki/Presence）。 呈现（presence）是一种在物理上和空间上位于环境中的感觉。呈现分为2种类型：认知呈现（Cognitive Presence）和感知呈现（Perceptive Presence）。

认知呈现是一个人思想的呈现。这可以通过观看引人入胜的电影或阅读引人入胜的书来实现。 认知呈现对于任何形式的沉浸式体验都很重要。

感知呈现是一个人感觉的呈现。要实现感知呈现，就必须欺骗人的感觉，视觉，声音，触觉和气味。为了创建可感知的呈现，XR设备必须能够欺骗用户的感觉，尤其是用户的视听系统。 XR设备通过对位置移动的跟踪来实现此目的。该跟踪系统的目标就是保持用户的临场感并避免被破坏。

当我们大脑被感官信息所欺骗而激活脑区，重新建立感官的适应，就可以实现这种呈现。例如用户会伸手去抓取并不存在的虚拟的苹果，或当他们在虚拟的屋顶上行走时，他们会感到紧张等等。

Brendan Iribe研究团队(<https://www.roadtovr.com/oculus-shares-5-key-ingredients-for-presence-in-virtual-reality/>)根据人眼视觉以及3GPP成员产品团队搜集的经验定义了视觉呈现的技术要求:

* + - 追踪（Tracking）
      * 6自由度追踪（6DoF）：能够追踪用户头部的旋转和平移运动
      * 360度追踪：无论用户面对哪个方向都能追踪用户的头部
      * 亚厘米精度：追踪不到一厘米的精度
      * 四分之一旋转追踪
      * 无抖动，设备上的画面保持平稳
      * 对室内游戏和体验，要有足够的追踪容量，即用户有足够大的空间移动但仍然能在仅有的2m的立方体里被捕捉。对于坐姿游戏和体验，则只需要较小的追踪容量就足够了。
      * 追踪需要频繁执行并处理XR观察者的最新姿势，最小的更新频率为大与或等于1000HZ，特别对于旋转跟踪则需要更高的频率。
    - 时延
      * 小于20ms的总体时间延迟，包括从你移动头部到你能看到设备移动的时间。
      * 最小化从姿势到渲染在到光子的时间。渲染内容要足够快，从渲染到光子的时间要小于50ms，从而避免错误的渲染内容。
      * 融合光学追踪和惯性测量单元数据
      * 最小化环路：追踪者-CPU-GPU-显示器-光学信号。
      * 最小化交互时延和依赖于应用的内容。
    - 持久度（persistence）
      * 低持久度：每2-3ms打开和关闭像素来避免拖尾或手势模糊。像素持久度是每一帧中显示器点亮的时间总和。每一帧持续时间越长，相对实际事物来说显示的精度就会越低。如果你转动大脑，在实际视野中大脑会看到不同的画面，但是通过VR大脑在这一帧只能接收到相同的画面。
      * 90HZ或者高于显示器的刷新频率消除画面闪烁。
    - 分辨率（Resolution）
      * 空间分辨率：无可见像素点，每英寸内的低分辨率和低像素会造成用户感知到像素并感觉像是透过屏幕在看外面。
      * 临时分辨率：尽管使用了异步时间扭曲，设备开发者仍然需要提供持续的90HZ或以上的帧频率才能传输一个传输一个舒适的VR而真正产生呈现的感觉。
    - 光学（optics）
      * 视野宽度（FOV）是指可观察世界的范围，典型场景下需要100-110度的FOV。
      * 舒适的眼镜盒，即人能通过眼镜舒适的观看画面的最小和最大眼镜距离。
      * 高质量的校准和纠错：纠正形变和色差用以符合眼镜的特性。

另一方面，感觉呈现（Sense of Presence）不仅仅对于VR体验很重要，对于沉浸式AR也很重要。为了达到增强现实的呈现效果，需要无缝集成虚拟内容和现实环境。例如在VR中，虚拟内容必须与用户的预期相匹配。对于真实的沉浸式AR特别是MR，则希望用户无法从真是对象中分辨出虚拟对象。

同样的对于VR和AR来说，特别是对于AR，不仅仅是要意识用户，还需要意识环境，其中包括：

* + - - 安全区域发现
    - - 动态障碍告警
    - - 几何和语义环境解析
    - - 环境照明
    - - 世界地图

对于AR，为了获得现实环境增强的视野，用户可以佩戴一个透视HMD而观察到由电脑对真实世界的视野进行叠加后产生的3D对象。这种透视HMD可以是光学HMD也可以是视频HMD。

除了前述的临场感和沉浸感之外，用户交互时延和内容老化对于沉浸式和非沉浸式交互体验至关重要。

用户交互延迟定义为在用户启动操作的那一刻到内容创建引擎根据此类操作开始进行处理的时间间隔。 比如在AR游戏中，用户交互延迟就是用户开始游戏互动的时刻到游戏引擎处理此类玩家响应的时间间隔。

内容老化定义为创建内容的时间与向用户呈现内容的时间之间的时间间隔。 比如在AR游戏中，这是游戏引擎创建视频帧到最终将帧呈现给玩家的时间间隔。

因此，双向交互延迟是内容老化和用户交互延迟的总和。 如果部分渲染是在AR服务器上完成的，并且服务器会生成帧缓冲区作为内容状态的渲染结果，则对于云游戏应用程序中基于栅格的分割渲染，以下过程会进一步增加这种延迟：

* + - 用户交互延迟

o在游戏客户端中捕获用户交互，

o将用户交互数据传递到游戏引擎，即服务器（又称网络延迟），

o游戏引擎/服务器对用户交互数据进行处理，

* + - 内容老化

o游戏引擎/服务器创建一个或几个视频缓冲区（例如，每只眼睛一个），

o将视频缓冲区编码为视频流帧，

o将视频帧传送到游戏客户端（又称网络延迟），

o游戏客户端对视频帧进行解码，

o向用户展示视频帧（又称帧速率延迟）。

以游戏为例，参考文献[1]和[2]提供了每种游戏类型的交互延迟容限阈值，如表4.2.2-1所示。 注意，该交互延迟是指如上所述的往返交互延迟。

|  |  |  |
| --- | --- | --- |
| 游戏类型 | 视角 | 延迟容限阈值 |
| 第一视角射击游戏(FPS) | 第一视角 | 100 ms |
| 角色扮演类游戏(RPG) | 第三视角 | 500 ms |
| 即时战略游戏(RTS) | 全向视角 | 1000 ms |

总体而言，不同的应用程序和用例需要不同的延迟要求。关于往返交互延迟，主要考虑了以下四个类别：

* + - 超低延迟应用程序：往返交互延迟阈值最多为50ms延迟。
    - 低延迟应用程序：往返交互延迟阈值最多为100ms延迟。
    - 中等延迟应用程序：往返交互延迟阈值最多为200ms延迟。
    - 非关键延迟应用程序：往返交互延迟阈值高于200ms延迟。

## 总结（定性）

# 云XR业务云网端需求与挑战（定量）

## 网络侧需求与挑战

<TBD>

## 云侧需求与挑战

<TBD>

## 端侧需求与挑战

### 强交互

### 弱交互

### AR业务

根据文献[3], 增强现实（AR）通过向用户提供额外的信息或人工生成的物品或内容并将它们实时叠加在当前现实世界的环境上。 此类信息或内容通常是基于视觉和（或）听觉的，用户对当前环境的观察可以是直接的，即没有中间的感测，处理和渲染；亦或是间接的，即用户对环境的感知是通过传感器传递的，并且是已经被增强或处理的。

文献[3]中列举了种类繁多的增强现实（AR）终端，有各种不同的尺寸规格。如Figure 3.3‑1所示。这些不同种类的终端在处理能力，通信类型以及功耗方面都有所不同的需求和挑战。

大多数的增强现实终端是一种光学头戴式的显示器（HMD）。它会在佩戴者眼前配置显示屏。 显示屏可以在佩戴者的眼前显示传输的数据流，图像和其他虚拟的数字信息。 一些HMD终端在用户的两只眼睛前面分别配置显示屏，还有一些HMD终端仅在用户的一只眼睛前面配置显示屏。在这些终端中，显示屏通常是透明的，虚拟的数字信息会叠加到真实的物品上实现增强现实。 AR终端允许用户在观看眼前显示屏上加载的数据流和图像叠加层的同时还可以看到周围真实世界的环境。佩戴者还可以通过诸如语音命令，手势以及控制器之类的输入方法与投影的数字内容进行交互。

HMD终端一般包括如下组件：

* + - 光学系统：包括显示屏和镜头
    - 跟踪传感器以及其他额外的传感器
    - 摄像头
    - 与AR相关的处理单元（概括为AR引擎），包括GPU，CPU，ASIC（例如专用的多媒体编解码器）等。
    - 通信功能，例如支持5G连接功能的Uu调制解调器

根据文献[3], AR终端必须具备跟踪以及即时定位与地图构建（SLAM）的能力，这也对AR终端提出了新的挑战，包括：

* + - 内向外追踪技术：一种常用的位置跟踪方法，专门用于跟踪头戴式显示器（HMD）和运动控制器配件的位置，根据摄像头或其他传感器的位置来确定物体在空间中的位置是否位于被跟踪的设备（例如HMD）上。
    - 世界追踪技术：一种创建AR体验的方法， AR终端的摄像头用来捕获设备方向和位置信息，检测现实世界的物理表面包括已知的图像或对象。用户可以体验由此构建的来自周围现实世界的虚拟内容。
    - 即时定位与地图构建（SLAM）：是在构建或更新未知环境地图的同时跟踪未知环境中用户位置。

对于SLAM所涉及的空间映射能力，包括创建周围环境的地图，以及定位处理（即在该空间内确定用户和物体的位置）对AR终端尤为重要。 结合多个传感器的输入信息可以获得更好的定位精度，这些传感器包括单眼/立体/深度摄像头，无线电信标，GPS，惯性传感器等。下面列出了涉及到的一些方法：

* + - 空间锚点用于在AR体验中建立3D对象的位置信息，从而使得该位置信息与用户的个人视角无关。 空间锚点在有限的空间内（例如Microsoft®Mixed Reality Toolkit的空间半径大约为3m）是准确的。 多个空间锚点可用于更大的空间。
    - 视觉定位，例如vSLAM，视觉定位系统（VPS）等，可以使用来自例如移动摄像头的视觉数据以及其他传感器数据进行定位。

空间映射和定位处理可以在AR终端上完成。 但是，网络可以以不同方式支持相关操作：

* + - 云服务可用于存储，检索和更新空间数据。 对于较大的公共场所，可以使用众包（crowdsourcing）来保持数据更新并提供给所有人。
    - 空间计算服务器，从多个数据源收集数据并对其进行处理以创建空间地图，包括但不限于来自于AR设备视觉和惯性数据流。 该服务可以将该信息提供给其他用户，并且还可以根据从其他用户设备接收的数据来帮助他们进行定位处理。

室内和室外的空间映射和定位处理会有不同的要求和限制。 服务提供商在扫描室内空间并存储空间特征时，尤其是在与全球定位相关联时，需要探索隐私问题。

文献[3]同样也提供了关于AR终端的概述。

智能手机（定义为AR5G-P1类型）通常可以用于AR。 通常，内置的AR引擎可用于支持来自传感器的数据处理，视野渲染以及SLAM处理。 通常5G调制解调器，多媒体处理单元以及 AR处理单元都集成在这类AR设备中。 对AR终端而言，功耗指标一般非常重要。但由于这类终端的电池容量相对较大，功耗指标并不是至关重要的。

对于AR眼镜类的终端，各种约束条件则直接影响终端的设计。终端产品的设计既要满足外观时尚的要求，同时也要考虑终端的重量和功耗。一般这类终端为了避免电池消耗过快以及满足必要的散热要求，一定程度上降低了对处理能力的要求。现阶段可穿戴AR终端中运用了两种不同的透视方案: 光学透视和视频透视。AR的应用场景通常是基于眼镜的，也可以称为光学透视。 同时，使用具有视频透视功能的HMD也可以实现AR业务的体验。例如，在[4]中提供了各种不同类型AR终端的比较。 一般来讲，AR终端分为以下类型：

* + - AR5G-A1线源连接AR穿戴式眼镜：这种类型终端已于2019年面世。终端包括最少数量的传感器，可用于AR定位的摄像头以及显示屏。电源，AR处理以及连接能力都是由终端外部提供。
    - AR5G-A2 无线连接AR穿戴式眼镜：这种类型终端尚未面世。这类终端包括最少数量的传感器，可用于AR定位的摄像头，电源以及用于提供连接能力的无线调制解调器。 AR处理，AR定位以及网络连接都是通过与其无线相连的外部设备（例如智能手机）来提供。
    - AR5G-A3 智能视频透视式AR HMD：这种类型终端是可以认为是能够应用AR应用的入门级终端。此类终端类似于VR的终端，可以通过摄像头捕获现实世界的画面经过转换后实时地投射到显示屏上以实现增强现实。这样看上去就像直接看到了外部影像，好像是你可以看穿物体。
    - AR5G-A4 独立可穿戴式AR眼镜：这种类型终端尚未面世。对于此类设备，预计5G调制解调器，电源以及所有多媒体处理单元以及AR处理单元都将集成在单个设备中。
    - AR5G-A5 智能无线穿戴式AR眼镜：这种类型终端尚未面世。除了AR5G-A2设备之外，此类设备还将包括至少一定数量的AR 处理和多媒体处理功能，例如编/解码器和AR处理单元。

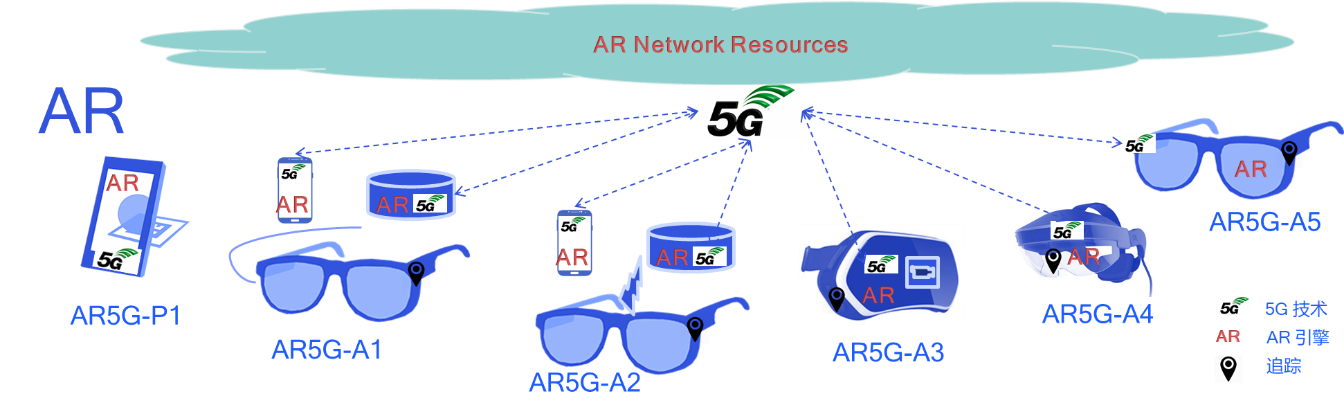


Figure 3.3‑1AR 终端规格

Table 3.3‑1中总结了不同AR终端类型的区别，包括部署了5G Uu调制解调器的外部设备与AR设备（如果适用）之间的连接方式，5G Uu调制解调器的部署方案，支持场景识别和视野渲染的AR引擎的部署方案，供电方式以及最大功率典型值。需要说明的是，这里是假设所有的传感器都部署在终端上。

对于AR引擎的部署方案有以下定义：

* + - 外部设备：即该类设备仅支持显示，并接收可以直接显示的完全渲染的视野数据。设备上没有任何场景识别（如果适用）。
    - 分拆渲染：即外部设备根据传感器信息对视野进行预渲染，而AR终端则根据最新传感器信息进行最终渲染。如前所述，存在不同程度的功能分割。同样，场景识别也可以进行拆分运算。
    - AR终端：即该类终端将完全渲染AR终端中的视野，传感器信息仅在本地处理。场景识别也完全有AR终端完成（如果适用）。

Table 3.3‑1 AR 终端类型

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| AR Type  Number | AR 终端类型 | 非独立终端连接外部设备方式 | 5G Uu 调制解调器部署方案 | AR 引擎部署方案 | 供电方式 | 最大功率典型值 |
| AR5G-P1 | 智能手机 | n/a | AR 终端 | AR 终端或分拆渲染 | 内部电池 | 3-5 W |
| AR5G-A1 | 线源连接AR穿戴式眼镜 | USB-C | 外部设备 | 外部设备 | 外部电池 | 1-3 W |
| AR5G-A2 | 无线连接AR穿戴式眼镜 | 802.11ad/y, 5G sidelink. etc. | 外部设备 | 外部设备 | 内部电池 | 0.5 – 2 W |
| AR5G-A3 | 智能AR视频透视 HMD | n/a | AR 终端 | AR 终端或分拆渲染 | 内部电池 | 3-7 W |
| AR5G-A4 | 独立可穿戴式AR眼镜 | n/a | AR 终端 | AR 终端或分拆渲染 | 内部电池 | 2 - 4 W |
| AR5G-A5 | 智能无线穿戴式AR眼镜 | 802.11ad/y, 5G sidelink. etc. | 外部设备 | AR 终端或分拆渲染 | 内部电池 | 0.5 – 2 W |

在设计多媒体处理，AR引擎和5G连接能力时，重要的是要了解AR设备中集成的各个不同组件的功耗值。建议考虑以下内容：

* + - 追踪与感知
    - 3DoF跟踪功耗较低（例如低于1瓦）
    - 6DoF跟踪由于涉及例如捕获摄像头信息，捕获用于眼睛和手部跟踪的LED等，耗电量较大
    - 显示
    - 显示屏的功耗至关重要。不同的品牌功耗差别较大。
    - 显示屏功耗可一般在0.3W至1W的范围内
    - 渲染（GPU）
    - GPU的功耗取决于帧频，分辨率以及显示技术
    - 根据使用情况，功耗可能从几毫瓦到几瓦范围内
    - 计算和多媒体处理（CPU）
    - 与GPU类似
    - 如果涉及编码，则功耗通常会较高。
    - 连接性
    - 5G等无线连接功能模块的功耗取决于多种因素，包括比特率，与无线接入网的距离，无线信道条件以及工作频率等。
    - 根据使用情况，功耗可能从几毫瓦到几瓦。

每个组件都可以进一步改进以解决能耗问题。在开发AR设备的技术规范时，必须考虑每个组件的功耗，这一点非常重要。

# 合作与展望

[1] "Cloud Gaming: Architecture and Performance", Ryan Shea and Jiangchuan Liu, Simon Fraser University; Edith C.-H. Ngai, Uppsala University; Yong Cui, Tsinghua University; IEEE Network-July/August 2013.

[2] M. Claypool and K. Claypool. Latency and player actions in online games. Communications of the ACM, 49(11):40–45, 2006.

[3] TR 26.928 “Technical Specification Group Services and System aspects:eXtended Reality (XR) in 5G”

[4] Long Qian, Alexander Barthel, Alex Johnson, Greg Osgood, Peter Kazanzides, Nassir Navab, and Bernhard Fuerst, "Comparison of optical see-through head-mounted displays for surgical interventions with object-anchored 2D-display", Int J Comput Assist Radiol Surg. 2017 Jun; 12(6): 901–910, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5891507/.