

# 团 体 标 准

T/CARA 0006—2023

## 增强现实数据类型及格式技术要求

Technical requirements for data types and formats of augmented reality

2023-12-17 发布

2023-12-17 实施

中国增强现实核心技术产业联盟 发布



CARA



版权保护文件

版权所有归属于该标准的发布机构，除非有其他规定，否则未经许可，此发行物及其章节不得以其他形式或任何手段进行复制、再版或使用，包括电子版，影印件，或发布在互联网及内部网络等。使用许可可于发布机构获取。

## 目 次

前 言 .....	III
1 范围 .....	1
2 规范性引用文件 .....	1
3 术语和定义 .....	1
4 缩略语 .....	2
5 AR 应用概述 .....	2
5.1 概述 .....	2
5.2 应用能力说明 .....	2
5.3 数据参考组件 .....	3
6 AR 数据类型要求 .....	4
6.1 概述 .....	4
6.2 媒体类数据类型 .....	4
6.3 交互类数据类型 .....	4
7 AR 数据格式要求 .....	5
7.1 概述 .....	5
7.2 AR 媒体数据要求 .....	5
7.3 AR 交互数据要求 .....	7
7.4 其他要求 .....	7
参 考 文 献 .....	8

## 前 言

本文件按照GB/T 1.1-2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本文件由 中国移动通信有限公司研究院提出。

本文件由中国增强现实核心技术产业联盟归口。

本文件起草单位：中国移动通信有限公司研究院、浙江商汤科技开发有限公司、浙江大学、艾迪普科技股份有限公司、北京邮电大学、华东理工大学、青岛理工大学、厦门大学、山东经鼎智能科技有限公司、南昌虚拟现实研究院股份有限公司、中移（成都）信息通信科技有限公司、广州增强信息科技有限公司、宇宙猜想（北京）科技有限公司、海信视像科技股份有限公司。

本文件主要起草人：尹瑜坚、高飞、蒋慧、章国锋、盛崇山、姜翰青、王一鹏、史鹏程、李亚健、滕维宇、苏媛、乔秀全、蔡方、陈成军、郭诗辉、石祥仁、白晓明、孙其民、黄亚坤、李建阳、穆永宏、戴亨钺、叶云、郑贵桢。



# 增强现实数据类型及格式技术要求

## 1 范围

本文件规定了统一的增强现实（Augmented Reality，AR）表示中的框架结构、数据类型和数据格式，包括 3D 视频、模型数据、位姿数据、交互数据等内容。

本文件适用于跨平台AR数据的传输、存储以及解析应用。

## 2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；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。

GB/T 38247-2019 信息技术 增强现实 术语

YD/T 3078-2016 移动增强现实业务能力总体技术要求

T/CARA 0003-2022 增强现实内容格式技术规范

## 3 术语和定义

GB/T 38247-2019 和 YD/T 3078-2016 界定的以及下列术语和定义适合于本文件。

### 3.1

**增强现实 augmented reality**

采用以计算机为核心的现代高科技手段生成的附加信息对使用者感知到的真实世界进行增强的环境，生成的信息以视觉、听觉、味觉、嗅觉、触觉等生理感觉融合的方式叠加至真实场景中。

[来源：GB/T 38247-2019，定义 2.1.2]

### 3.2

**虚实融合 virtual-reality integration**

端或计算机通过数字技术产生的虚拟场景与真实场景进行实时的数字化混合，使虚拟物体和真实物体共同存在同一个场景中。

[来源：GB/T 38247-2019，定义 2.2.31]

### 3.3

**AR 内容 AR content**

所有可以被用于增强/提升用户对现实世界感知体验的多媒体对象，譬如文本、图片、视频、音频、三维模型等。通常，增强现实内容对应于特定增强现实目标。

### 3.4

**锚点 anchor point**

将虚拟物体放置在环境过程中所使用的基准点。

3.5

**AR 标识 AR marker**

在屏幕显示的一个数字对象，用于指示与增强现实目标所对应增强现实内容的可用性。

3.6

**AR 目标 AR target**

现实世界中真实存在的任何实体，譬如兴趣点、物品、人、车辆等。增强现实目标可以与一个或者多个增强现实内容相关联。

4 缩略语

下列缩略语适用于本文件。

3DoF: 三自由度 (Three Degrees of Freedom)

6DoF: 六自由度 (Six Degrees of Freedom)

AR: 增强现实 (Augmented reality)

FEC: 前向纠错 (Forward Error Correction)

FOV 视场 (Field of View)

VST: 视频透视式 (Video see-through)

OST: 光学透视式 (Optical see-through)

RTC: 实时通信 (Real time communication)

SLAM: 同步定位和地图构建 (Simultaneous location and mapping)

5 AR 应用概述

5.1 概述

AR 应用程序可以处理具有不同特征的各种媒体类型，实现 AR 应用虚实融合的应用功能。本节介绍了 AR 应用通用功能能力，以及应用功能所依赖的媒体和数据，并形成数据参考组件，作为数据类型和格式要求的研究基础。

5.2 应用能力说明

5.2.1 AR 内容获取

支持 AR 应用服务器推送 AR 内容、AR 标识、锚点等数据至客户端，实现 AR 内容在指定位置加载。

5.2.2 AR 内容交互

支持通过相机采集位姿数据，包含相机 FOV 以及表情等交互数据，完成数据类型的采集与生成，用于虚实渲染。其中强计算能力终端可生成空间信息数据，计算当前 FOV 中的环境信息，环境点云信息，来确定环境中的可安置锚点的平面。

支持在虚拟融合环境中建立场景三维坐标系，用于描述相机、AR 内容、用户环境统一尺度和



位置关系。

### 5.2.3 AR 内容传输

在视频和实时通信业务中，根据 AR 内容存储和处理节点，支持 AR 应用相关媒体和数据的分发传输，包括媒体压缩、数据组打包、数据同步等功能。

支持通过网络接口实现 AR 数据共享和同步。

### 5.2.4 AR 内容处理显示

支持点云等模型的虚实渲染，进行光照估计、尺度估计、深度估计、虚实遮挡、运动跟踪等。

支持现实交互渲染，可结合传感器采集的交互数据，在内容中实时更新现实设备及识别目标的空间变换信息，这个变换需要反映出现实中真实的空间关系和操作指令。

支持数据组帧渲染能力，包括交互数据、环境数据、媒体数据合并渲染。

支持通过内容分发从网络获取资源，并在 AR 应用中加载并正确显示资源。

### 5.2.5 AR 内容存储管理

支持指定 AR 内容的类型，包括模型数据、环境数据、识别图等资源；

支持增加、删除、修改、查询 AR 内容资源。

## 5.3 数据参考组件

为支持上述 AR 应用能力，本文将 AR 数据分为如图 1 所示三个部分，包括媒体类数据、交互类数据以及应用控制数据。其中媒体类数据、交互类数据用于实现 AR 内容获取、内容交互、内容传输、内容处理现实和存储管理等，应用控制数据属于应用通用要求，用于应用会话建立和资源调用等控制面功能，因此，本文将主要对 AR 应用的媒体类数据、交互类数据进行统一要求。



图1 AR应用数据参考组件

媒体类数据用于呈现或显示 AR 多媒体内容，可以进行渲染并叠加到用户真实的场景中，包括文本、图片、音频、2D/3D 视频，3D 模型、点云、场景描述及动画等类型，提供沉浸式多元的增强现实应用体验。

交互类数据用于应用感知交互环境并反馈用户操作，用于环境感知包括空间建模、锚点、统一坐标系以及 AR 内容和用户的位姿等，用于用户操作的交互信息包括用户 FoV、面部表情、动作姿态以及设备操作等信息。

## 6 AR 数据类型要求

### 6.1 概述

针对上述应用功能和参考组件,本章制定相应的数据类型说明,满足 AR 沉浸式多媒体数据在个人或者行业场景中的共享和使用,包括适用于移动终端 VST 和 AR 头显终端 OST 中的 AR 应用,满足业务中 AR 应用的内容及数据交换。

### 6.2 媒体类数据类型

媒体类数据需支持包括文本、图片、2D 视频、3D 模型、3D 视频、音频、场景及动画多媒体数据类型,该类数据需满足 AR 内容获取、传输、处理显示和存储管理的应用能力要求。媒体类数据类型具体应满足以下要求:

1) 支持 2D 视频数据,应支持携带彩色(例如 RGB)信息和透明信息(alpha\_channel)的 2D 视频。

2) 可支持 3D 视频数据处理,应支持携带彩色信息和深度信息(例如 RGBD)的视频处理,包括体积视频等。

3) 支持点云建模数据,点云数据应至少包括一个三位坐标系统中一组向量的集合,可以附带 RGB 信息,即每个坐标点的颜色信息。若点云建模数据作为 SLAM 过程的一部分,需包含视觉特征、关键帧和空间贴图用于映射真实世界。

4) 支持 3D 模型数据,对象描述为顶点、面和其他元素的列表,以及相关的属性。需至少包括几何数据和拓扑信息两个方面:①几何数据:用于确定 3D 模型在空间中的位置与大小,包括顶点、边、面,顶点、边、面之间可以互相导出。②拓扑信息:用于表示 3D 模型的构造方法,描述该 3D 模型各个组成元素的数目及相互关系,如相邻关系,链接关系,包含关系等。

5) 支持场景描述,用于描述 3D 场景的组成,通常使用树或图结构来参考和定位场景中的不同 2D 和 3D 内容。

6) 支持常规 2D 音频数据,可支持多声道和空间音频、3D 音频等音频类型。

7) 支持动画组件:为 AR 3D 模型添加组件用于增强显示效果,支持材质组件切换模型贴图效果、碰撞器组件增加碰撞检测效果、刚体组件增加重力等物理属性、声音组件用于播放和检测场景中的声音以及依附于模型的脚本组件。

### 6.3 交互类数据类型

交互类数据需支持用于空间感知的空间建模、锚点、统一坐标系以及 AR 内容和用户的位姿等,用于用户操作的交互信息包括用户 FoV、面部表情、动作姿态以及设备操作等信息,该类数据主要满足应用能力要求中的内容交互和内容处理,具体应满足以下要求

1) 支持点云建模数据、锚点、统一坐标系以及 AR 内容和用户位姿在同一个坐标系统中的同一表征,包括绝对位置和相对位置的坐标值和方向向量。

支持点云建模数据、锚点、统一坐标系以及 AR 内容和用户位姿实时同步更新和更新后内容渲染。

2) 支持 AR 锚点,用户空间中用于锚定 AR 内容(2D 或 3D)的点,该元数据支持文本、图片、视频等 AR 内容的精准覆盖或呈现。

3) 支持用户位姿表示用户的位置和方向。

4) 支持视野(FOV)数据,用于表示任何给定时刻可观察到的世界的范围,包括横向 FOV 和纵向 FOV。同时支持视口数据,为用户视野在目标显示器上的投影。

5) 支持终端 Action 数据,包含动作姿态、面部表情、设备操作和时间戳信息的实时持续读取,

用于触发所构建的虚拟环境的模型或场景发生动作或状态的改变。

7 AR 数据格式要求

7.1 概述

本章根据不同的AR视频和实时通信业务的AR内容处理节点，规范支持AR应用相关媒体和数据的格式要求，以满足上述AR应用能力中的内容获取、内容交互、分发传输、处理显示和存储管理，例如处理显示要求中支持空间定位，满足基于2D、3D视频数据类型的空间锚点和平面检测，标记虚拟对象位置，能够根据环境进行实时调整，能够在真实场景中检测到水平平面和垂直平面的位置、边缘以及大小，使虚拟对象紧贴垂直平面或水平平面，避免对象被放置到空中或穿到环境物体内，实现虚拟2D特效、3D模型等与之动态匹配，再将叠加的结果按要求格式输出；内容交互中支持识别增强，满足文本、音频、视频数据类型的识别，如特征点识别、语音识别、手势识别、语义识别等，同时可结合交互数据（FOV、位姿、表情等）满足点云、模型的虚实渲染，进行光照估计、尺度估计、深度估计、虚实遮挡、运动跟踪等；存储管理要求中支持满足上述格式要求的媒体类数据：点云建模数据、3D模型数据、识别图等存储、删除、修改、查询。

7.2 AR 媒体数据要求

AR应用服务器和客户端支持上述媒体类数据交换。其中

- 1) 文本应至少支持txt格式，可选择支持csv；
- 2) 图片应至少支持jpg，png格式，可选择支持tga，gif等；
- 3) 音频封装应至少支持mp3和wav，可选择支持wmv，m4a等；
- 4) 视频封装应至少支持mp4，可选择支mov，avi，swf等。2D、3D视频源参数至少满足以下要求：

表 1 视频源参数要求

内容分辨率	1080P
内容帧率	30fps
色深	8bit
采样/像素比特深度	4:2:0/12bit

表 2 全景视频参数要求

内容分辨率	4K
内容帧率	30fps
色深	8bit
采样/像素比特深度	4:2:0/12bit

表 3 体积（3D）视频参数要求

形式	点云、3D mesh
内容帧率	25fps
色深	8bit
采样/像素比特深度	4:2:0/16bit

5) 点云建模数据应至少支持\*.las和\*.pcd 点云数据格式，可选择支持\*.ply, \*.off或其他通用的点云数据格式。

6) 3D模型模型类型应使用polygon建模，场景描述应支持glTF2.0, JSON forma模型数据格式和国产MSD模型数据格式，用于描述模型、场景、模组、材质、物理属性以及其他可扩展信息，可选择obj, fbx, glb, USDZ（Universal Scene Description）。

其中glTF由json和外部数据构成，构成如下：一个json格式的文件（.gltf）包含完整的场景描述：阶段层级结构、材质、相机以及mesh、animation等的描述信息。二进制文件(.bin)，存储geometry、animation数据以及其他基于buffer的数据纹理文件(.jpg/.png)；MSD以Json格式进行编写，在格式中使用简单的键值描述了对引擎、场景树、物件结构、属性参数、纹理材质、动画效果、逻辑编排、以及文件IO等信息进行处理和描述。

- 引擎描述：引擎信息包含了引擎加载素材时所需要的基础配置，其中包括引擎制式、刷新频率等信息。
- 场景树：描述了需要渲染呈现的三维空间场景。在场景树中，可以包含天空盒、灯光、摄像机等基础显示配置，也可以包含三维网格、二维平面、文字、线条等一系列内容。
- 物件结构：为了降低场景中的数据存储、并提升加载网格加载速度，MSD支持存储离散后的结构数据。对于一些通用格式内容，也可在存储时仅通过文件间的引用关系进行保存。
- 属性参数：在MSD中记录着所有渲染所需的空间属性、物件属性、特技属性等参数。通过属性参数可以控制三维空间的搭建，以及效果呈现。
- 纹理材质：用于描述各个加载的网格物件在渲染时所需要的纹理属性。描述信息需要区分Phong材质以及PBR材质。对于Phong可支持环境光、漫反射、高光及自发光效果参数的添加，而对于PBR则支持基础色、金属度、粗糙度、法线、凹凸、置换等参数的填写。对于纹理，可以通过外部贴图实现，也支持记录对外部文件的引用关系。
- 动画效果：动画方面包含了所有在动画轨道中使用的素材变化曲线参数。通过关键帧、参数值以及贝塞尔曲线参数予以描述。同时动画中还包含了对于循环方式及循环区间的表述。
- 逻辑编排：节点编辑无代码技术，通过将多个逻辑模块进行互联运算，从而形成一整套逻辑处理流程。通过MSD格式可以记录所有需要执行处理的关联关系，并记录所有的注释及展示信息。
- 文件IO：MSD文件可以描述一个完整的工程场景，也可以描述一个模型素材，或是描述一段逻辑编排。在进行使用时可以将器整合成一个更大的场景或时工程使用。因此在进行编排时，可以把每个MSD文件当作一个处理单元，将一个MSD与其他MSD文件的关联详细进行逻辑连接。因此需要支持记录每个MSD向外部的引出内容。这些引出内容可作为输入参数，也可作为输出结果。

6) 动画组件：材质组件支持的模型贴图为 tga、png 格式，脚本组件支持 json 数据格式。

### 7.3 AR 交互数据要求

#### 1.交互输入应满足如下技术要求:

- 1) 应支持输入数据类型, 如: 文本、音频、视频、触摸/传感数据等, 输入格式支持上述7.2节要求。对于输入为其他感数据时, 应支持将其转换为指令数据。
- 2) FoV和视口数据可根据相机参数, 如焦距、主点、校准参数和相机的姿态, 获取三维场景中的点与捕获图像中的像素之间的相关性。
- 3) 骨骼和表情数据可使用关键点阵列表示。
- 4) 运动捕获Action数据采用bvh format, 频率至少为1khz.
- 5) 用户位姿数据需包含用于方向的四元数和用于位置的三维矢量组成。时间戳由64位单调递增的基于纳秒的整数表示。
- 6) 注视点的三维矢量用注视方向的四元数表示。
- 7) 获取投影到3D场景相关联的参数, 包括透视、正交和全向等投影类型。

2.交互输出应支持以多种形式进行渲染结果的输出, 如: 音频、图片、视频流、3D 模型等; 输出格式满足上述 7.2 节要求。

### 7.4 其他要求

- 1) 根据 AR 媒体处理节点, 支持 AR 应用相关音视频媒体数据通过网络以实时通信 (RTC) 的方式传输给终端, 利用 FEC 丢包纠错方式减少重传, 并根据网络状态动态调整 FEC 达到减少带宽占用的效果。
- 2) 支持多元数据同步传输, 包括终端音视频数据同步和交互类数据的同步。
- 3) 交互类数据满足 1khz 以上实时刷新。
- 4) 为实现渲染过程的可视化和交互性, 可采用渐进式渲染显示的方式, 在渲染过程中逐步显示图像的细节, 并逐步改善图像的质量。用户可以通过设置渐进式渲染的参数, 并在渲染过程中实时观察和调整渲染结果, 可以随时暂停和恢复渲染任务, 支持的渲染状态为.vrprog 文件, 包括渐进渲染的信息。

## 参 考 文 献

- [1] GB/T 36341.1-2018 信息技术 形状建模信息表示 第 1 部分：体系结构和组件
- [2] GB/T 36341.2-2018 信息技术 形状建模信息表示 第 2 部分：特征约束
- [3] GB/T 36341.3-2018 信息技术 形状建模信息表示 第 3 部分：流式传输
- [4] GB/T 36341.4-2018 信息技术 形状建模信息表示 第 4 部分：存储格式

CARA