国　　际　　电　　信　　联　　盟

ITU-R

国际电联无线电通信部门

ITU-R BT.2154-0建议书

（12/2022）

在各种类型显示设备上呈现的沉浸式视频高层系统架构

BT系列

广播业务（电视）

国际电信联盟

前言

无线电通信部门的职责是确保卫星业务等所有无线电通信业务合理、平等、有效、经济地使用无线电频谱，不受频率范围限制地开展研究并在此基础上通过建议书。

无线电通信部门的规则和政策职能由世界或区域无线电通信大会以及无线电通信全会在研究组的支持下履行。

知识产权政策（IPR）

ITU-R的IPR政策述于ITU-R第1号决议中所参引的《ITU-T/ITU-R/ISO/IEC的通用专利政策》。专利持有人用于提交专利声明和许可声明的表格可从http://www.itu.int/ITU-R/go/patents/zh获得，在此处也可获取《ITU-T/ITU-R/ISO/IEC的通用专利政策实施指南》和ITU-R专利信息数据库。

ITU-R 系列建议书

（也可在线查询 http：//www.itu.int/publ/R-REC/zh）

系列	标题
BO	卫星传送
BR	用于制作、存档和播出的录制；电视电影
BS	广播业务（声音）
BT	**广播业务（电视）**
F	固定业务
M	移动、无线电定位、业余和相关卫星业务
P	无线电波传播
RA	射电天文
RS	遥感系统
S	卫星固定业务
SA	空间应用和气象
SF	卫星固定业务和固定业务系统间的频率共用和协调
SM	频谱管理
SNG	卫星新闻采集
TF	时间信号和频率标准发射
V	词汇和相关问题

说明： 该ITU-R建议书的英文版本根据ITU-R第1号决议详述的程序予以批准。

电子出版
2023年，日内瓦

ITU-R BT.2154-0 建议书

在各种类型显示设备上呈现的沉浸式视频高层系统架构

（ITU-R第140-1/6和ITU-R第143-2/6号课题）

（2022年）

范围

本建议书为在各种类型显示设备上呈现沉浸式视频提供了一个高层系统架构。该架构由视频对象、场景描述、渲染器和播放器组成，并将其作为组件的最小集合。本建议书亦确定了需在渲染器和播放器之间传输的信息。

关键词

沉浸式视频、6自由度、场景描述、立体视频、设备自适应

国际电联无线电通信全会，

考虑到

a) 沉浸式视频使终端用户能够在视频空间中移动，并从自由视点全方位观看视频，提供了一种新的增强视觉体验；

b) 沉浸式视频的表示方式为在三维空间中排列诸如立体视频、全向视频和二维视频的视频对象；

c) 终端用户可以使用各种类型显示设备，例如头戴式显示器、智能手机和平板电脑，因此需要考虑使用这些设备来呈现沉浸式视频；

d) 可利用日益增多的交互式传送平台向观众分发内容；

e) 包括具有提高的计算能力的云和边缘的网络服务器将被有效地用于渲染适应于具有不同计算和显示能力的不同类型显示设备的沉浸式视频；

f) 用于在各种类型显示设备上呈现沉浸式视频的通用架构将促进沉浸式视频系统和应用的发展，

认识到

a) ISO/IEC 23090标准集 – 信息技术 – 沉浸式媒体的编码表示；

b) ITU-R BT.2123建议书 – 用于广播节目制作和国际节目交换的高级沉浸式视听系统的视频参数值；

c) ITU-R BT.2420报告 – 高级沉浸式感官媒体系统使用场景集合，

建议

应根据附件中所述的高层系统架构来设计面向各种类型显示设备的沉浸式视频系统。

附件

在各种类型显示设备上呈现的沉浸式视频高层系统架构

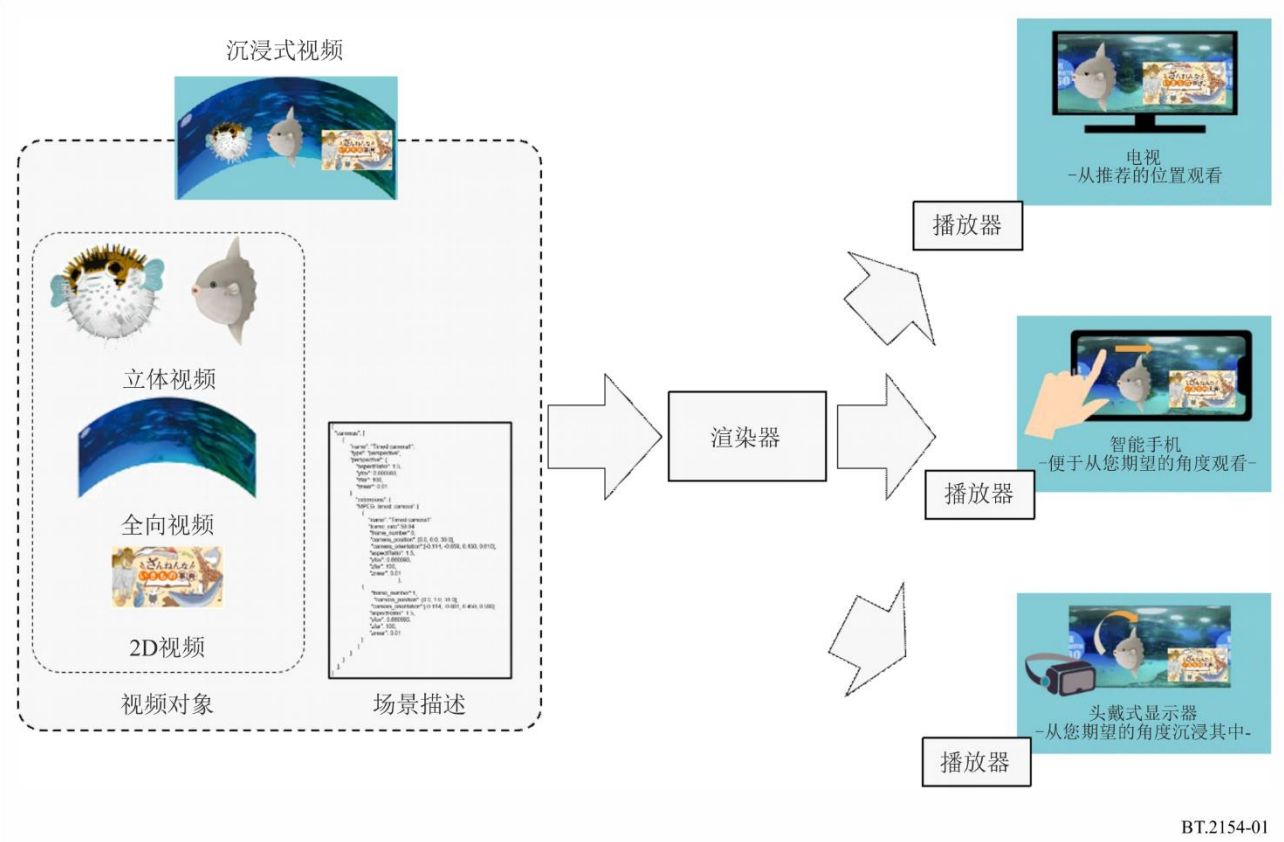
1 概述

图1为沉浸式视频从合成到在各种类型显示设备上呈现的系统架构概览图。

沉浸式视频由场景描述和从场景描述引用的多个视频对象组成，其中包括可以表示对象的三维形状和纹理的立体视频、围绕对象的全向视频和二维（2D）视频，以便用户可以从任何位置、任何方向观看视频。全向视频和2D视频可具有其深度信息。场景描述提供了沉浸式视频的时间序列三维表示，例如每个对象的位置、方向和大小及其在三维空间中的空间和时间排列。

图1

沉浸式视频从合成到呈现的概览图



渲染器借助由场景描述指示的各种视频对象构建三维空间，并产生要从用户的位置和方向观看的视频。根据用户的观看位置和方向，每个显示设备的播放器以最适合该设备的方式呈现经渲染的视频。

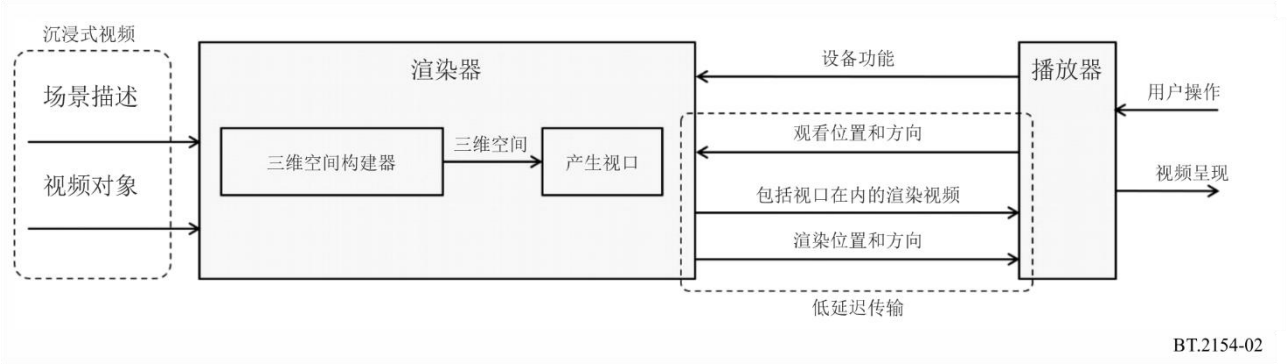
2 高层系统架构

2.1 定义

定义图2所示高层系统架构的目的是在各种类型显示设备上呈现沉浸式视频。

图2

用于沉浸式视频的高层系统架构



2.2 沉浸式视频

沉浸式视频代表一个时间序列的三维空间，由视频对象和场景描述组成。

视频对象包括表示对象的三维形状和纹理的立体视频、围绕对象的全向视频以及二维（2D）矩形视频。全向视频和2D视频可与深度信息相关联。

场景描述参考多个视频对象，并识别每个对象的位置、方向和大小及其在三维空间中的空间和时间排列，并以此来定义时间序列三维空间。

场景描述亦可包括有关内容创建者推荐的用户观看位置和观看方向的信息，即推荐的视口，以便在显示设备上呈现。

2.3 渲染器和播放器

渲染器借助各种视频对象和场景描述来构建三维空间，并产生视频，以便从用户的位置和方向观看。

根据用户的观看位置和方向，播放器以最适合设备的方式呈现来自渲染视频的视频。

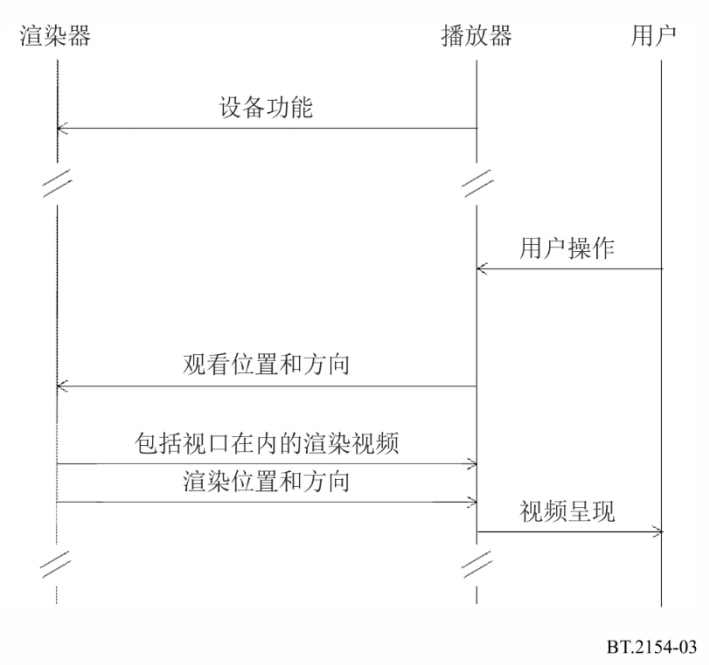
渲染器和播放器的功能应彼此分离，这样播放器便无需处理三维空间中的所有信息，而只需处理需要在设备上呈现的部分。这种功能上的分离减少了要由播放器处理的视频对象的数据量和播放器的处理负荷，从而实现了处理播放器的轻量化。即使将来要引入其他类型的视频对象，亦只需更新渲染器来为其提供支持，而无需更新播放器。

2.4 需在渲染器和播放器之间传输的信息

图3显示了需在渲染器和播放器之间传输的信息流。

图3

需在渲染器和播放器之间传输的信息



1 在开始渲染过程之前，渲染器基于场景描述和视频对象构建三维空间，且播放器向渲染器通知包括显示分辨率、视野和帧速率在内的设备功能。

2 当用户开始观看内容时，播放器根据用户的操作向渲染器通知用户的观看位置和方向。在用户观看视频期间，其观看位置和方向可能发生改变。

3 根据所通知的用户的观看位置和方向，渲染器从三维空间产生包括要呈现的视口在内的视频。渲染器可在比视口更宽的区域范围内产生视频，以支持观看位置和方向的快速移动。此外，渲染器亦可基于场景描述中的推荐视口信息（如包括在内的话）来产生推荐视口视频。

4 通过指示产生视频时所使用的三维空间中的渲染位置和方向，将经渲染的视频传输给播放器。低延迟传输将用于传输视频及有关渲染位置和方向的信息。

5 播放器根据用户的观看位置和方向呈现全部或部分传输的视频。

附件的
后附资料
（参考资料）

高层系统架构的实施示例

1 概述

本后附资料给出了一个示例系统，其中实施了附件中定义的、用于在各种类型显示设备上呈现的沉浸式视频的高层系统架构。

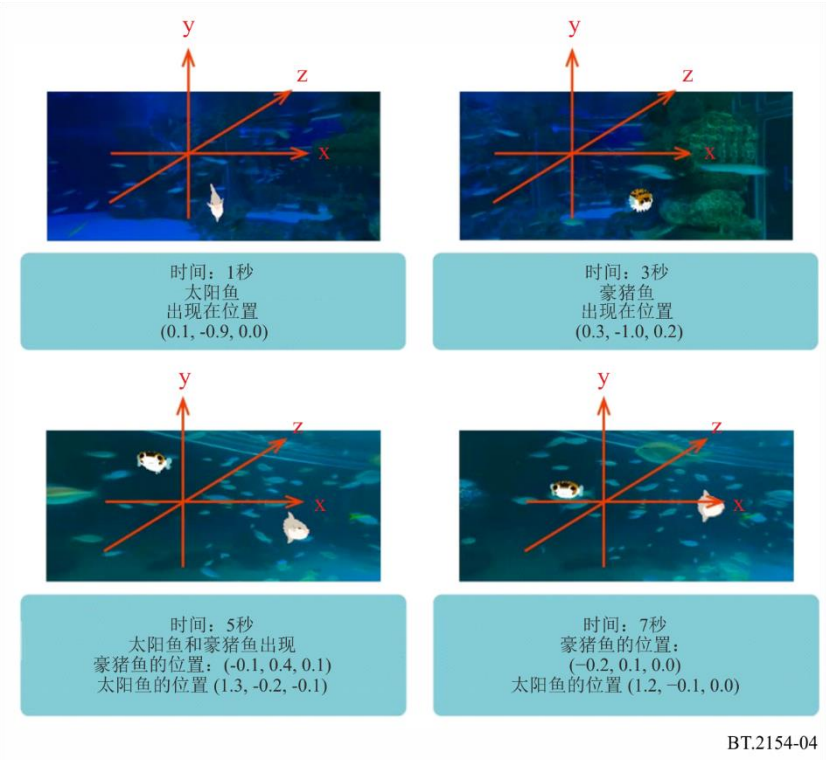
2 沉浸式视频

2.1 场景描述

场景描述的概念如图4所示。如图所示，当时刻为1秒时，一个太阳鱼形状的对象出现在三维空间中的位置（0.1, −0.9, 0.0）。2秒后，当时刻为3秒时，一条豪猪鱼出现在位置（0.3, −1.0, 0.2）。在这一时刻，太阳鱼形状的对象消失了。场景描述便以这种方式来指定每个时刻三维空间中对象的位置、方向和大小。

图4

利用场景描述在三维空间中排列时序对象



在本例中，采用了GL传输格式（glTF2）的扩展格式（相关规范见 https://github.com/KhronosGroup/glTF/tree/master/specification/2.0）来进行场景描述。图5为场景描述的一个示例。

图5

场景描述示例


```
[
    "frame_number": 618,
    "rotation_object": [0.03668982873033452, 0.7522537201043805, 0.017108748113350298, -0.6576286853497625],
    "scale_object": [0.03900000000000042, 0.03900000000000042, 0.03900000000000042],
    "translation_object": [-83.94561853512538, -15.251572393537403, 13.225600523272751],
    "visible": 1
],
[
    "frame_number": 619,
    "rotation_object": [0.02024137343578336, 0.23900985486236237, 0.03505908720575184, -0.970172889996628],
    "scale_object": [0.03900000000000042, 0.03900000000000042, 0.03900000000000042],
    "translation_object": [-148.076839849297, -12.958146408306028, -38.03696833117341],
    "visible": 1
],
[
    "frame_number": 620,
    "rotation_object": [0.03316152485827769, 0.6266292729985842, 0.023219949684294753, -0.7782653155749667],
    "scale_object": [0.03900000000000042, 0.03900000000000042, 0.03900000000000042],
    "translation_object": [-101.243284426844, -9.882305069562054, -50.61199066105607],
    "visible": 1
],
```


BT.2154-05

2.2 视频对象

对于立体视频的视频对象，所使用的点云流是通过以ISO/IEC 23090-5“基于可视立体视频的编码和基于视频的点云压缩”压缩点云格式的立体视频而获得的。

对于全向视频，所使用的视频是通过以ISO/IEC 23090-2“全向媒体格式（OMAF）”存储的等矩形投影（ERP）转换的360度视频而获得的。

此外，使用了二维矩形视频来进行叠加呈现。

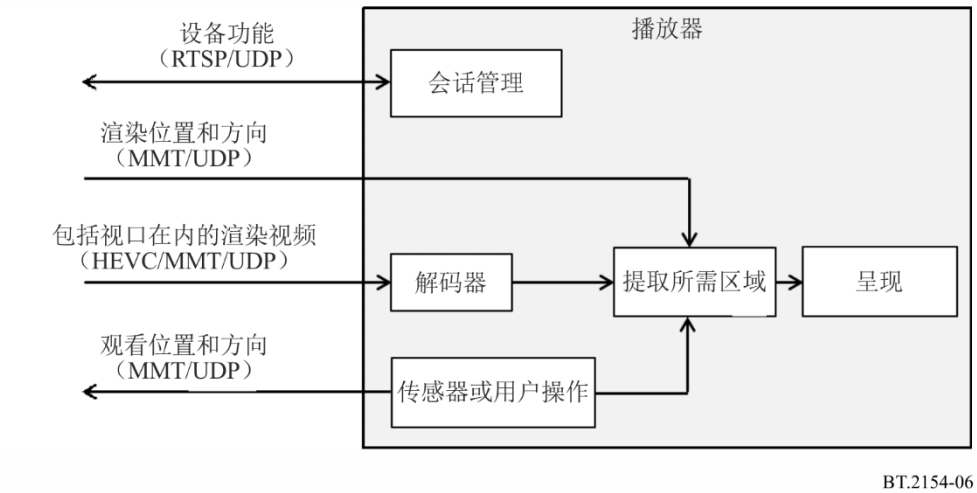
3 渲染器和播放器的实施

3.1 播放器的实施

已分别为头戴式显示器、智能手机/平板电脑和传统电视机开发了播放器。传统电视机的播放器不允许用户改变观看位置和方向。有关此类设备功能模块的描述见图6和图7。

图6

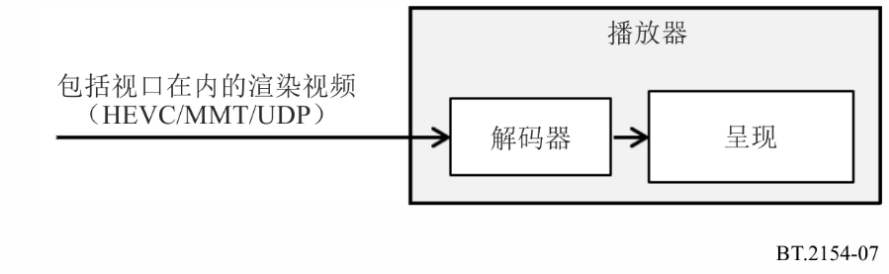
头戴式显示器和智能手机/平板电脑播放器的功能模块



BT.2154-06

图7

无用户操作设备（假设为电视机）的播放器的功能模块



播放器使用实时流协议（RTSP，IETF RFC 7826）设置方法来建立与服务器的会话，并将其功能通知服务器，其中包括其显示器分辨率、帧速率、视野以及用于压缩包括视口在内的视频的可用编码方法。

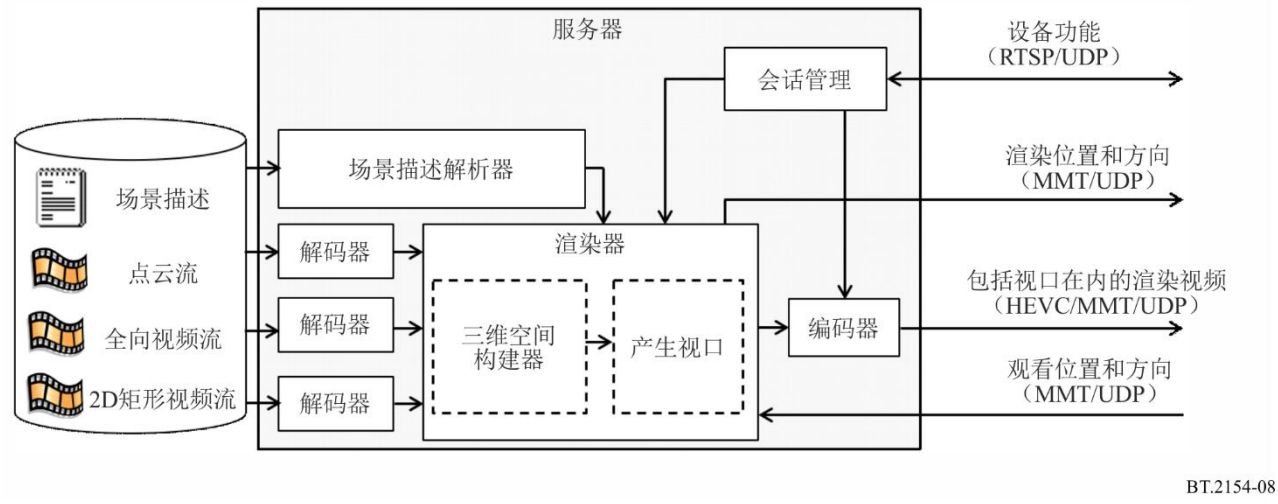
用户的观看位置和方向以MPEG媒体传输（MMT，ISO/IEC 23008-1）消息的格式通知给服务器。在头戴式显示器的情况下，根据用户自身的移动来确定观看位置和方向；在智能手机/平板电脑的情况下，则根据用户的屏幕操作来确定观看位置和方向。

3.2 渲染器的实施

具有渲染器功能的服务器与播放器是分别开发的。根据设备类型，需要不同类型的播放器，但是，不管播放器的类型如何，服务器都是通用的。图8展示了具有渲染器功能的服务器的功能模块。

图8

具有渲染器功能的服务器的功能模块



服务器解析场景描述，实时解码所需的视频对象，并根据场景描述将其排列在三维空间中。然后，渲染器从三维空间产生视频作为视口，该视口具有根据播放器通知的观看位置和方向而确定的显示分辨率。系统亦根据包括在设备的场景描述中的推荐视口信息产生另一视口，其中不执行设备功能通知，且观看位置/方向不变。

渲染器生成的包括视口在内的视频采用高效视频编码（HEVC，ISO/IEC 23008-2 | ITU-T H.265建议书）作为二维视频，并以MMT格式传输给播放器。同时，用于产生视口的渲染参数以MMT消息格式传输给播放器。

4 在三种不同类型显示设备上的呈现

4.1 头戴式显示器

如图9所示，使用头戴式显示器令用户得以从期望的位置在期望的方向上观看视频，同时以高沉浸感自由地四处移动。这使得用户不仅能从前面，亦能从后面和侧面观看对象。

图9

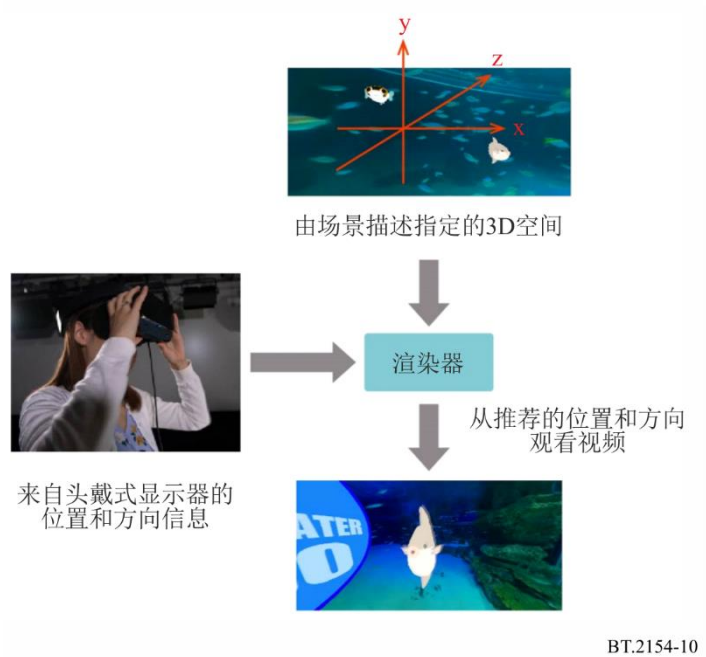
使用头戴式显示器观看

BT.2154-09

在此系统中，渲染器根据由头戴式显示器的传感器检测到的用户观看位置和方向来产生视频，播放器则在头戴式显示器上呈现产生的视频。用于在头戴式显示器上呈现视频的机制见图10。

图10

在头戴式显示器上呈现视频的机制



BT.2154-10

4.2 智能手机

观看位置和方向可通过在智能手机屏幕上进行操作来改变，这令用户得以从其期望的位置及期望的方向上观看视频（见图11）。

图11

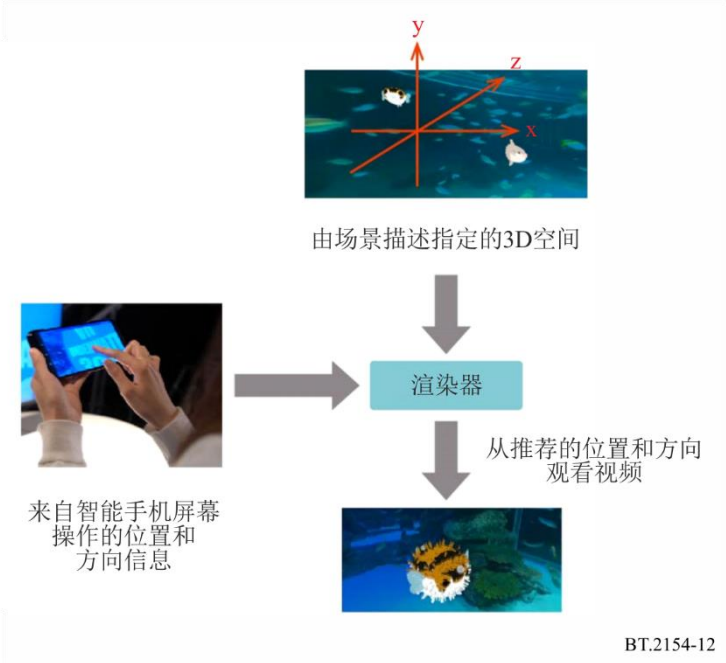
在智能手机上观看

BT.2154-11

与在头戴式显示器上呈现的情况一样，渲染器根据场景描述产生要在智能手机上呈现的视频。在智能手机上，播放器呈现的是在用户屏幕操作所指定的位置和方向上所观看的视频（见图12）。

图12

在智能手机上呈现视频的机制



4.3 电视机

尽管用户无法像在头戴式显示器和智能手机上那样改变电视机的位置和方向，但仍可从内容创建者推荐的观看位置和方向上轻松欣赏视频（见图13）。

图13

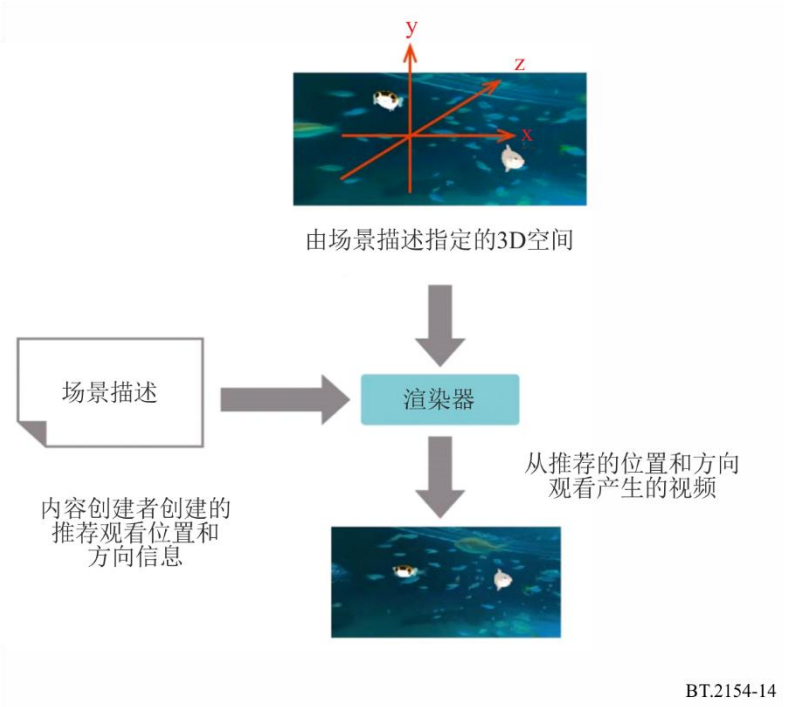
在电视机上观看

BT.2154-13

即使在这种情况下，渲染器亦基于由场景描述指定的三维空间信息来产生视频。不过，由于没有用户操作，有关观看位置和方向的信息由场景描述作为推荐的视口信息给出。根据该信息，渲染器产生要呈现的视频，如图14所示。

图14

在无用户操作显示器（电视机）上呈现视频的机制



BT.2154-14

5 参考文献

以下网址提供了在三种设备上呈现视频方面的内容：
https://www.nhk.or.jp/strl/english/open2021/tenji/3/index.html

在上述实施过程中使用的规范如下：

ITU-T H.265建议书 | ISO/IEC 23008-2 (2020)：信息技术 – 异构环境中的高效编码和媒体传输 – 第2部分：高效视频编码

ISO/IEC 23008-1:2017：信息技术 – 异构环境中的高效编码和媒体传输 – 第1部分：MPEG媒体传输

ISO/IEC 23090-2:2021：信息技术 – 沉浸式媒体的编码表示 – 第2部分：全向媒体格式

ISO/IEC 23090-5:2021：信息技术 – 沉浸式媒体的编码表示 – 第5部分：基于可视立体视频的编码（V3C）和基于视频的点云压缩（V-PCC）

IETF RFC 7826 (2016：实时流协议版本2.0

glTF 2.0 Khronos组，GL传输格式（glTF）2.0规范，可从以下网址获得 https://github.com/KhronosGroup/glTF/tree/master/specification/2.0/