

# МСЭ-R

Сектор радиосвязи МСЭ

## Рекомендация МСЭ-R ВТ.2154-0 (12/2022)

### Системная архитектура высокого уровня для иммерсивного видео с представлением на устройствах отображения различных типов

Серия ВТ  
Радиовещательная служба  
(телевизионная)



## Предисловие

Роль Сектора радиосвязи заключается в обеспечении рационального, справедливого, эффективного и экономичного использования радиочастотного спектра всеми службами радиосвязи, включая спутниковые службы, и проведении в неограниченном частотном диапазоне исследований, на основании которых принимаются Рекомендации.

Всемирные и региональные конференции радиосвязи и ассамблеи радиосвязи при поддержке исследовательских комиссий выполняют регламентарную и политическую функции Сектора радиосвязи.

### Политика в области прав интеллектуальной собственности (ПИС)

Политика МСЭ-R в области ПИС излагается в общей патентной политике МСЭ-T/МСЭ-R/ИСО/МЭК, упоминаемой в Резолюции МСЭ-R 1. Формы, которые владельцам патентов следует использовать для представления патентных заявлений и деклараций о лицензировании, представлены по адресу: <http://www.itu.int/ITU-R/go/patents/ru>, где также содержатся Руководящие принципы по выполнению общей патентной политики МСЭ-T/МСЭ-R/ИСО/МЭК и база данных патентной информации МСЭ-R.

### Серии Рекомендаций МСЭ-R

(Представлены также в онлайн-форме по адресу: <http://www.itu.int/publ/R-REC/ru>.)

Серия	Название
BO	Спутниковое радиовещание
BR	Запись для производства, архивирования и воспроизведения; пленки для телевидения
BS	Радиовещательная служба (звуковая)
<b>BT</b>	<b>Радиовещательная служба (телевизионная)</b>
F	Фиксированная служба
M	Подвижные службы, служба радиоопределения, любительская служба и относящиеся к ним спутниковые службы
P	Распространение радиоволн
RA	Радиоастрономия
RS	Системы дистанционного зондирования
S	Фиксированная спутниковая служба
SA	Космические применения и метеорология
SF	Совместное использование частот и координация между системами фиксированной спутниковой службы и фиксированной службы
SM	Управление использованием спектра
SNG	Спутниковый сбор новостей
TF	Передача сигналов времени и эталонных частот
V	Словарь и связанные с ним вопросы

*Примечание.* – Настоящая Рекомендация МСЭ-R утверждена на английском языке в соответствии с процедурой, изложенной в Резолюции МСЭ-R 1.

Электронная публикация  
Женева, 2023 г.

© ITU 2023

Все права сохранены. Ни одна из частей данной публикации не может быть воспроизведена с помощью каких бы то ни было средств без предварительного письменного разрешения МСЭ.

## РЕКОМЕНДАЦИЯ МСЭ-R ВТ.2154-0

**Системная архитектура высокого уровня для иммерсивного видео с представлением на устройствах отображения различных типов**

(Вопросы МСЭ-R 140-1/6 и МСЭ-R 143-2/6)

(2022)

**Сфера применения**

В настоящей Рекомендации описана системная архитектура высокого уровня для иммерсивного видео с представлением на устройствах отображения различных типов. Минимальный набор компонентов этой архитектуры включает в себя визуальные объекты, описание сцены, систему визуализации и средство воспроизведения. Кроме того, в Рекомендации определяется, какая информация должна передаваться между системой визуализации и средством воспроизведения.

**Ключевые слова**

Иммерсивное видео, 6DoF, описание сцены, объемное видео, адаптация устройств.

Ассамблея радиосвязи МСЭ,

*учитывая,*

- a) что иммерсивное видео, позволяющее конечным пользователям перемещаться в видеопространстве и смотреть видео в любом направлении с нефиксированных точек обзора, обеспечивает новый расширенный визуальный опыт;
- b) что иммерсивное видео представляет собой совокупность организованных определенным образом визуальных объектов, включая объемное видео, всенаправленное видео и двухмерное видео в трехмерном пространстве;
- c) что конечным пользователям доступны устройства отображения различных типов, такие как головные дисплеи, смартфоны и планшеты, которые требуют рассмотрения в контексте представления иммерсивного видео;
- d) что для распределения контента аудитории доступно все большее число интерактивных платформ доставки;
- e) что имеющие повышенную вычислительную мощность серверы в сетях, включая облачные и периферийные вычисления, будут эффективно использоваться для визуализации иммерсивного видео путем адаптации к устройствам отображения различных типов с разными вычислительными мощностями и возможностями отображения;
- f) что единая архитектура для представления иммерсивного видео на устройствах отображения различных типов будет способствовать развитию иммерсивных видеосистем и соответствующих приложений,

*признавая*

- a) набор стандартов ИСО/МЭК 23090 – Информационная технология. Кодированное представление иммерсивных медиа;
- b) Рекомендацию МСЭ-R ВТ.2123 – Значения параметров видеосигнала для перспективных иммерсивных аудиовизуальных систем для производства международных программ и обмена ими в радиовещании;
- c) Отчет МСЭ-R ВТ.2420 – Сборник сценариев использования передовых иммерсивных сенсорных медиасистем,

рекомендует

проектировать иммерсивные видеосистемы, предназначенные для устройств отображения различных типов, в соответствии с системной архитектурой высокого уровня, описанной в Приложении 1.

## Приложение 1

### Системная архитектура высокого уровня для иммерсивного видео с представлением на устройствах отображения различных типов

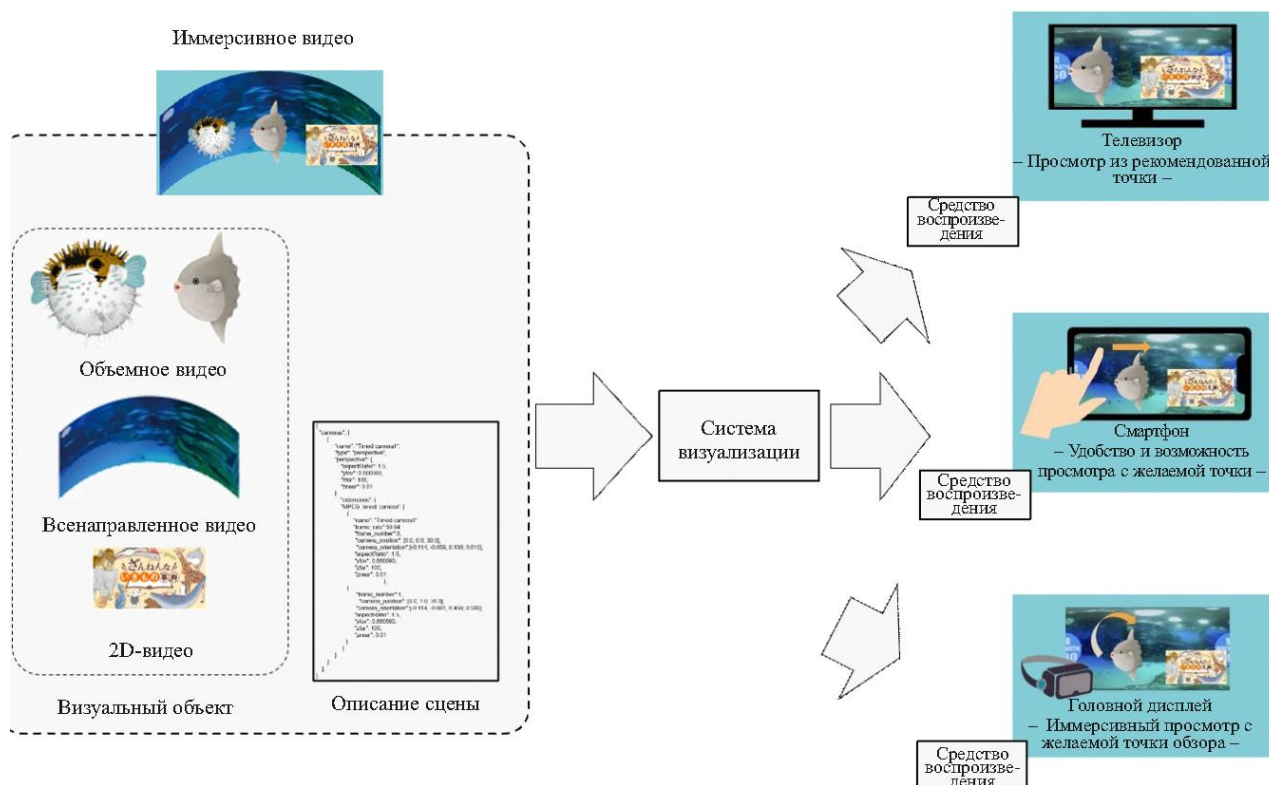
#### 1 Обзор

На рисунке 1 представлен обзор системной архитектуры для иммерсивного видео, начиная от составных элементов и заканчивая представлением на устройствах отображения различных типов.

Иммерсивное видео включает в себя описание сцены и множество визуальных объектов, создаваемых на основе информации из описания сцены, и состоит из объемного видео, с помощью которого могут передаваться трехмерная форма и текстура объектов, всенаправленное видео, окружающее объекты, и двухмерное видео для обеспечения пользователям возможности просматривать видеоизображение с любых точек и в любых направлениях. Всенаправленное и двухмерное видео могут иметь собственную информацию о глубине. Описания сцен задают параметры трехмерного представления иммерсивного видео в конкретной временной последовательности, включая позицию, ориентацию и размер каждого объекта, а также их пространственно-временное расположение в трехмерном пространстве.

РИСУНОК 1

Обзор иммерсивного видео: составные элементы и представление



Система визуализации конструирует трехмерное пространство из различных визуальных объектов, указанных в описании сцены, создавая видео для просмотра из точки и направления просмотра пользователя. Средство воспроизведения каждого устройства отображения представляет созданное системой визуализации видео наиболее оптимальным для конкретного устройства образом исходя из точки и направления просмотра пользователя.

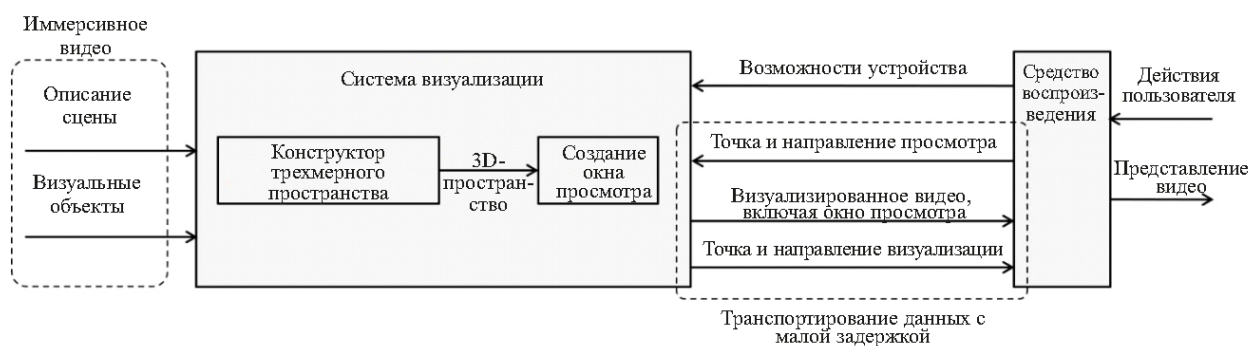
## 2 Системная архитектура высокого уровня

### 2.1 Определение

На рисунке 2 показана системная архитектура высокого уровня, определяемая для представления иммерсивного видео на устройствах отображения различных типов.

РИСУНОК 2

Системная архитектура высокого уровня для иммерсивного видео



ВТ.2154-02

### 2.2 Иммерсивное видео

Иммерсивное видео представляет собой организованное во временной последовательности трехмерное пространство и состоит визуальных объектов и описания сцены.

Визуальные объекты подразделяются на объемное видео, передающее трехмерность форм и текстуру объектов, всенаправленное видео, окружающее объекты, и двухмерное (2D) прямоугольное видео. Для всенаправленного видео и 2D-видео может быть предусмотрена информация о глубине.

Описание сцены определяет трехмерное пространство во временной последовательности, предоставляя информацию о множестве визуальных объектов с указанием позиции, ориентации и размера каждого объекта, а также их пространственно-временного расположения в трехмерном пространстве.

В описании сцены может также содержаться информация о рекомендованных создателем контента точке и направлении просмотра пользователя, то есть о рекомендованном окне просмотра, для представления видео на устройстве отображения.

### 2.3 Система визуализации и средство воспроизведения

Система визуализации конструирует трехмерное пространство из различных визуальных объектов и на основе описания сцены. Кроме того, она создает видео для просмотра из точки и направления обзора пользователя.

Средство воспроизведения представляет созданное системой визуализации видео наиболее оптимальным для конкретного устройства образом исходя из точки и направления просмотра пользователя.

Функции системы визуализации и средства воспроизведения должны быть разделены, с тем чтобы средству воспроизведения требовалось обрабатывать не всю информацию о трехмерном пространстве, а только ту ее часть, которая должна быть представлена на устройстве. Такое разделение функций уменьшает объем связанных с визуальными объектами данных, которые приходится обрабатывать средству воспроизведения, и снижает нагрузку на него, благодаря чему могут применяться средства воспроизведения для более легких процессов обработки данных. Даже если в будущем будут внедрены дополнительные типы визуальных объектов, для их поддержки необходимо будет обновить только систему визуализации, а средство воспроизведения обновлять не потребуется.

## 2.4 Информация, передаваемая между системой визуализации и средством воспроизведения

На рисунке 3 показано движение информации, передаваемой между системой визуализации и средством воспроизведения.

РИСУНОК 3

### Информация, передаваемая между системой визуализации и средством воспроизведения



ВТ.2154-03

- 1 Прежде чем приступить к процессу визуализации, система визуализации конструирует трехмерное пространство на основе описания сцены и визуальных объектов, а средство воспроизведения сообщает системе визуализации информацию о возможностях устройства, включая разрешение отображения, поле просмотра и частоту кадров.
- 2 Когда пользователь начинает просмотр контента, средство воспроизведения сообщает системе визуализации информацию о точке и направлении просмотра пользователя, которые в процессе просмотра видео могут меняться в зависимости от действий пользователя.
- 3 С учетом переданной информации о точке и направлении просмотра пользователя система визуализации создает из трехмерного пространства видео для представления, включая окно просмотра. Для реагирования на быструю смену точки и направления просмотра система визуализации может создавать видео для более широкого круга областей, чем те, что охвачены окном просмотра. Кроме того, система визуализации может создавать видео с рекомендованным окном просмотра в случае, если информация о рекомендованном окне просмотра включена в описание сцены.

- 4 Созданное системой визуализации видео передается на средство воспроизведения с указанием точки и направления визуализации в трехмерном пространстве, которые были использованы при создании видео. Передача видео и информации о точке и направлении визуализации требует транспортирования данных с низкой задержкой.
- 5 Средство воспроизведения представляет все переданное видео целиком либо его часть, в зависимости от точки и направления просмотра пользователя.

**Прилагаемый документ  
к Приложению  
(информационный)**

**Пример реализации системной архитектуры высокого уровня**

**1 Обзор**

В настоящем Прилагаемом документе приводится пример системы, в которой реализована определенная в Приложении 1 системная архитектура высокого уровня для иммерсивного видео с представлением на устройствах отображения различных типов.

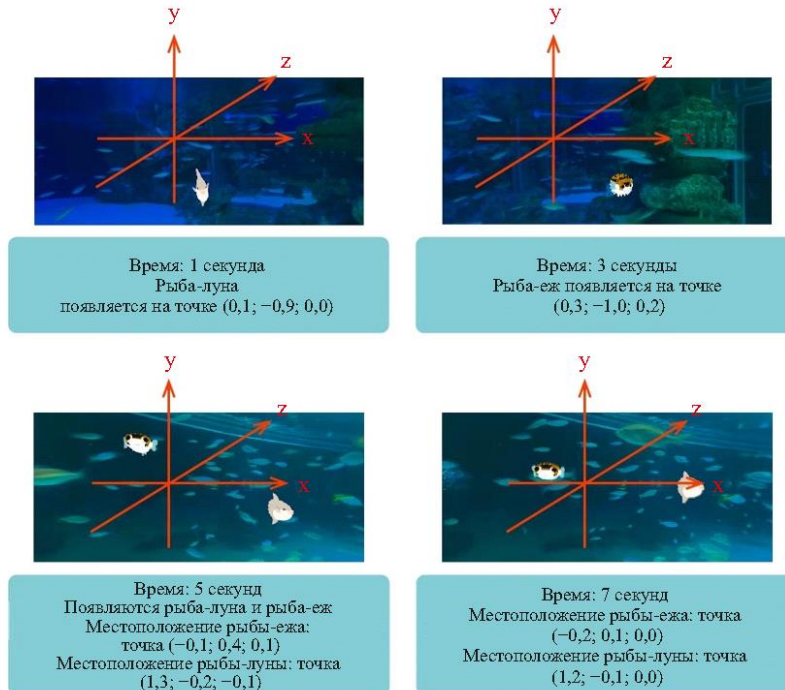
**2 Иммерсивное видео**

**2.1 Описание сцены**

Концепция описания сцены проиллюстрирована на рисунке 4. Как показано на рисунке, в момент времени 1 секунда в трехмерном пространстве появляется объект "рыба-луна" на точке  $(0,1; -0,9; 0,0)$ . Спустя 2 секунды, на 3-й секунде, появляется рыба-еж на точке  $(0,3; -1,0; 0,2)$ . К этому моменту времени объект "рыба-луна" уже исчез. Подобным образом в описании сцены указываются позиция, ориентация и размер объектов в трехмерном пространстве на каждый момент времени.

РИСУНОК 4

Расположение объектов в трехмерном пространстве во временной последовательности на основе описания сцены



ВТ.2154-04

В данном примере для описания сцены используется расширенный формат GL Transmission Format (glTF2), спецификации которого представлены по адресу: <https://github.com/KhronosGroup/glTF/tree/master/specification/2.0>. На рисунке 5 приведен пример описания сцены.

РИСУНОК 5

Пример описания сцены

```
[
  {
    "frame_number": 618,
    "rotation_object": [0.03668982873033452, 0.7522537201043805, 0.017108748113350298, -0.6576286853497625],
    "scale_object": [0.039000000000000042, 0.039000000000000042, 0.039000000000000042],
    "translation_object": [-83.94561853512538, -15.251572393537403, 13.22560052327275],
    "visible": 1
  },
  {
    "frame_number": 619,
    "rotation_object": [0.02024137343578336, 0.23900985486236237, 0.03505908720575184, -0.970172889996628],
    "scale_object": [0.039000000000000042, 0.039000000000000042, 0.039000000000000042],
    "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": [0.039000000000000042, 0.039000000000000042, 0.039000000000000042],
    "translation_object": [-101.243284426844, -9.882305069562054, -50.61199066105607],
    "visible": 1
  }
],
```

ВТ.2154-05

## 2.2 Визуальные объекты

Для объектов объемного видео используются потоки облаков точек, получаемые путем сжатия объемного видео в формате облаков точек в соответствии со стандартом ИСО/МЭК 23090-5 "Визуальное объемное кодирование на основе видео и сжатие облака точек на основе видео".



В качестве всенаправленного видео используется видео, получаемое путем преобразования видео формата 360° на основе проекции равных прямоугольников (ERP), в формате, предусмотренном стандартом ИСО/МЭК 23090-2 "Формат всенаправленных медиа (OMAF)".

Кроме того, используется двухмерное прямоугольное видео для представления наложения изображений.

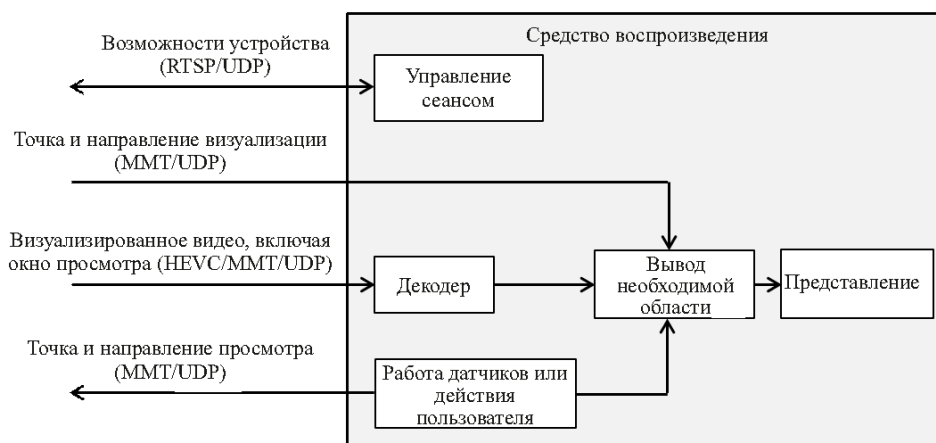
### 3 Реализация системы визуализации и средства воспроизведения

#### 3.1 Реализация средства воспроизведения

Были разработаны средства воспроизведения для головных дисплеев, смартфонов/планшетов и обычных телевизоров. Средство воспроизведения для обычного телевизора не позволяет пользователю менять точку и направление просмотра. Функциональные блоки соответствующих устройств показаны на рисунках 6 и 7.

РИСУНОК 6

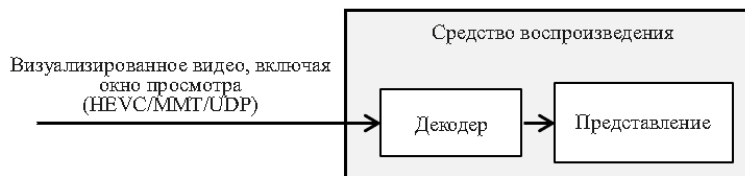
Функциональные блоки средства воспроизведения для головного дисплея и смартфона/планшета



ВТ.2154-06

РИСУНОК 7

Функциональные блоки средства воспроизведения для устройств, не предполагающих действий со стороны пользователя, на примере телевизора



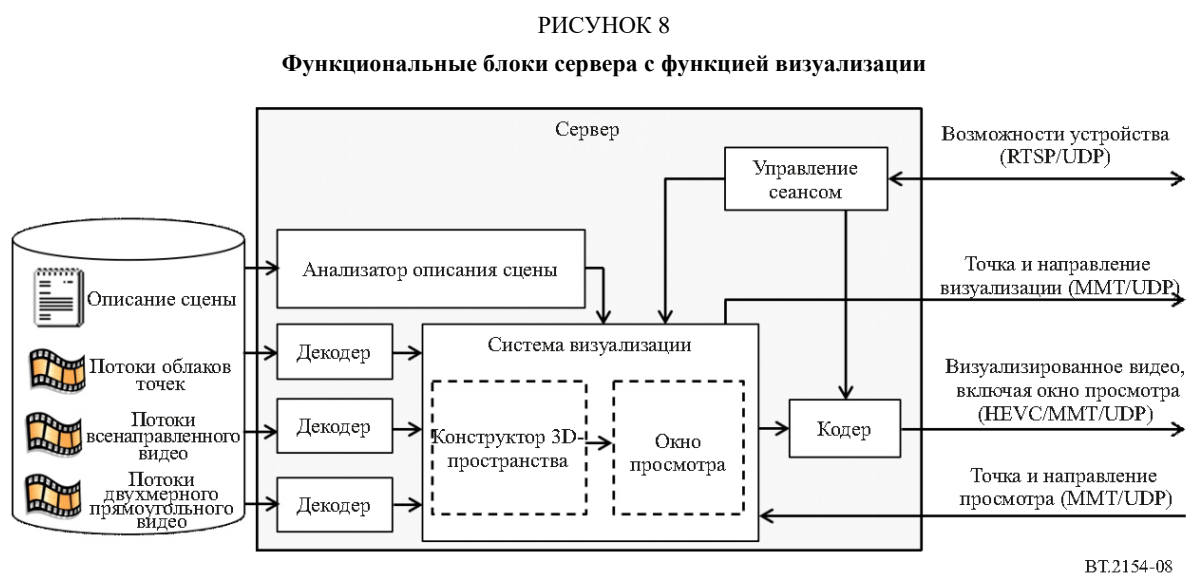
ВТ.2154-07

Средство воспроизведения использует метод SETUP (запрос установки транспортного механизма для медиа-контента) в рамках протокола потоковой передачи в реальном времени (RTSP, IETF RFC 7826), для установления сеанса с сервером и передачи ему информации о своих возможностях, включая разрешение отображения, частоту кадров и поле просмотра, а также о методе кодирования, доступном для сжатия видео, включая окно просмотра.

Серверу сообщается информация о точке и направлении просмотра пользователя в формате сообщения ММТ ("Транспортирование данных MPEG", стандарт ИСО/МЭК 23008-1). В случае с головным дисплеем точка и направление просмотра определяются исходя из собственного движения пользователя, а в случае со смартфоном/планшетом они определяются на основе действий пользователя с экраном.

### 3.2 Реализация системы визуализации

Сервер с функцией визуализации разрабатывается отдельно от средства воспроизведения. Для разных типов устройств требуются разные типы средства воспроизведения, в то время как сервер остается одним и тем же вне зависимости от типа средства воспроизведения. На рисунке 8 показаны функциональные блоки сервера с функцией визуализации.



Сервер анализирует описания сцен, в режиме реального времени декодирует требуемые визуальные объекты и располагает их в трехмерном пространстве в соответствии с описаниями сцен. Затем система визуализации из этого трехмерного пространства создает видео в виде окна просмотра с разрешением отображения, соответствующим точке и направлению просмотра, указанным в сообщении от средства воспроизведения. Создается также еще одно окно просмотра на основе включенной в описания сцен информации о рекомендованном окне просмотра для устройств, которые не отправляют сообщений о возможностях устройства и не предполагают смены точки/направления просмотра.

Созданное системой визуализации видео, включая окно просмотра, сжимается путем высокоэффективного кодирования видеоизображений (HEVC, стандарт ИСО/МЭК 23008-2 | Рек. МСЭ-T Н.265) до двумерного видео и передается на средство воспроизведения в формате ММТ. Параллельно с этим средству воспроизведения передается сообщение в формате ММТ с параметрами визуализации, которые использовались при создании окна просмотра.

## 4 Представление на устройствах отображения трех разных типов

### 4.1 Головные дисплеи

Как показано на рисунке 9, использование головного дисплея позволяет пользователям смотреть видео в желаемом направлении из желаемой точки, свободно передвигаясь в пространстве с эффектом глубокого погружения. Пользователь может наблюдать объекты не только спереди, но также сзади и сбоку.

РИСУНОК 9

## Просмотр с использованием головного дисплея

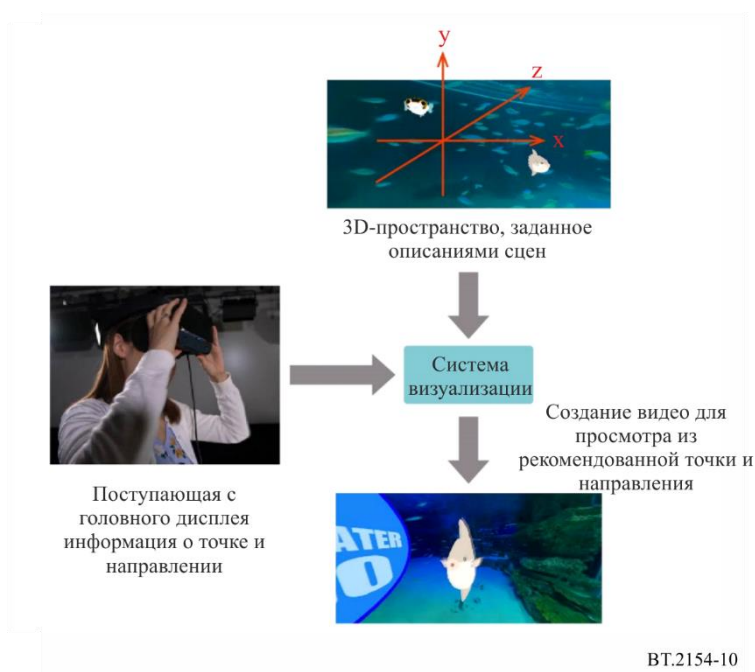


ВТ.2154-09

Система визуализации создает видео с учетом точки и направления просмотра пользователя, которые определяются датчиками головного дисплея, а средство воспроизведения представляет это видео на головном дисплее. Механизм представления на головном дисплее показан на рисунке 10.

РИСУНОК 10

## Механизм представления на головном дисплее



## 4.2 Смартфон

Точка и направление просмотра на смартфоне могут меняться в зависимости от действий пользователя с экраном, что позволяет пользователю смотреть видео в желаемом направлении и с желаемой точки (см. рисунок 11).

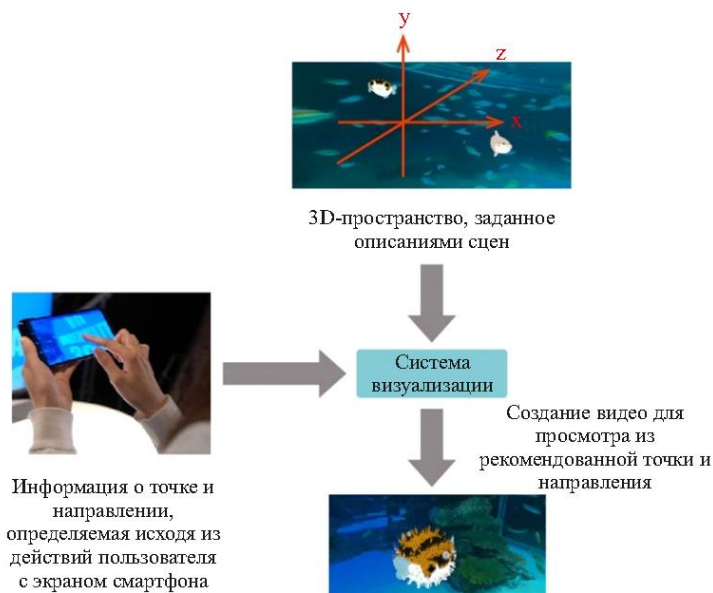
РИСУНОК 11  
Просмотр на смартфоне



ВТ.2154-11

Как и в случае с головными дисплеями, видео для представления на смартфонах создается системой визуализации на основе описаний сцен. Средство воспроизведения на смартфонах представляет видео для просмотра из точки и направления, определяемых исходя из действий пользователя с экраном (см. рисунок 12).

РИСУНОК 12  
Механизм представления на смартфоне



ВТ.2154-12

### 4.3 Телевизор

Несмотря на то, что на телевизоре, в отличие от головных дисплеев и смартфонов, пользователи не могут менять точку и направление просмотра, они все равно могут просматривать видео с той точки и из того направления, которые рекомендованы создателем контента (см. рисунок 13).

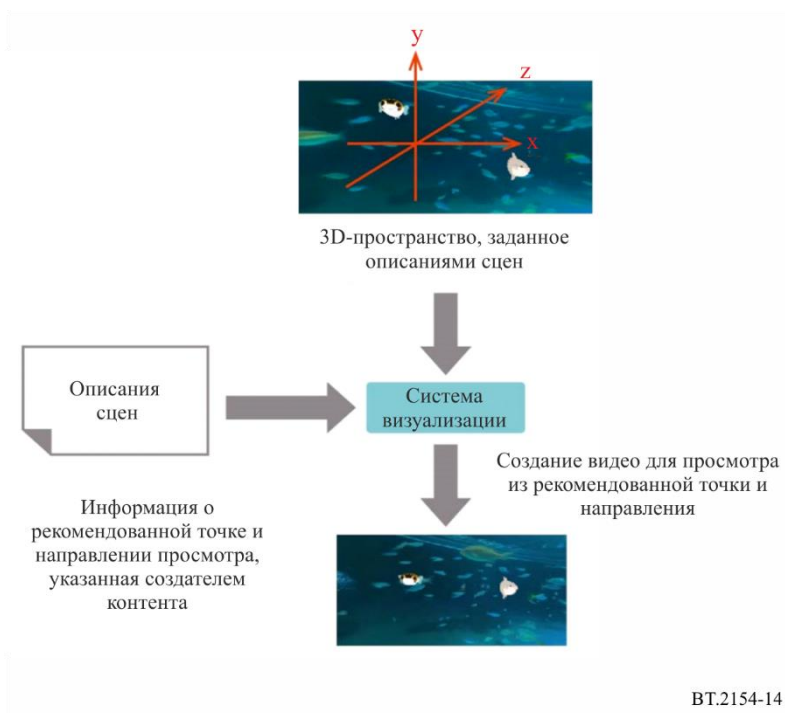
РИСУНОК 13  
Просмотр на телевизоре



ВТ.2154-13

Даже в этом случае система визуализации создает видео на основе информации о трехмерном пространстве, содержащейся в описаниях сцен. Однако, поскольку действий со стороны пользователя не предполагается, информация о точке и направлении просмотра задается в описаниях сцен как информация о рекомендованном окне просмотра. На основе этой информации система визуализации создает видео для представления, как показано на рисунке 14.

РИСУНОК 14  
Механизм представления для дисплея (телевизора),  
не предполагающего действий со стороны пользователя



## 5 Справочные материалы

Презентация по устройствам трех типов доступна по следующему URL:  
<https://www.nhk.or.jp/str/english/open2021/tenji/3/index.html>.

Спецификации для использования при реализации:

Recommendation ITU-T H.265 | ISO/IEC 23008-2 (2020): Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 2: High efficiency video coding.

ISO/IEC 23008-1:2017: Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 1: MPEG media transport.

ISO/IEC 23090-2:2021: Information technology – Coded representation of immersive media – Part 2: Omnidirectional media format.

ISO/IEC 23090-5:2021: Information technology – Coded representation of immersive media – Part 5: Visual volumetric video-based coding (V3C) and video-based point cloud compression (V-PCC).

IETF RFC 7826 (2016): Real-Time Streaming Protocol Version 2.0.

glTF 2.0 Khronos Group, The GL Transmission Format (glTF) 2.0 Specification, URL: <https://github.com/KhronosGroup/glTF/tree/master/specification/2.0/>.

---