Union internationale des télécommunications

UIT-R

Secteur des Radiocommunications de l'UIT

Recommandation UIT-R BT.2154-0

(12/2022)

Architecture de système de haut niveau pour la vidéo en immersion pour la présentation sur différents types de dispositifs d'affichage

Série BT

Service de radiodiffusion télévisuelle

Union internationale des télécommunications

Avant-propos

Le rôle du Secteur des radiocommunications est d'assurer l'utilisation rationnelle, équitable, efficace et économique du spectre radioélectrique par tous les services de radiocommunication, y compris les services par satellite, et de procéder à des études pour toutes les gammes de fréquences, à partir desquelles les Recommandations seront élaborées et adoptées.

Les fonctions réglementaires et politiques du Secteur des radiocommunications sont remplies par les Conférences mondiales et régionales des radiocommunications et par les Assemblées des radiocommunications assistées par les Commissions d'études.

Politique en matière de droits de propriété intellectuelle (IPR)

La politique de l'UIT-R en matière de droits de propriété intellectuelle est décrite dans la «Politique commune de l'UIT-T, l'UIT-R, l'ISO et la CEI en matière de brevets», dont il est question dans la Résolution UIT-R 1. Les formulaires que les titulaires de brevets doivent utiliser pour soumettre les déclarations de brevet et d'octroi de licence sont accessibles à l'adresse http://www.itu.int/ITU-R/go/patents/fr, où l'on trouvera également les Lignes directrices pour la mise en œuvre de la politique commune en matière de brevets de l'UIT-T, l'UIT-R, l'ISO et la CEI et la base de données en matière de brevets de l'UIT-R.

Séries des Recommandations UIT-R

(Également disponible en ligne: http://www.itu.int/publ/R-REC/fr)

Séries	Titre
BO	Diffusion par satellite
BR	Enregistrement pour la production, l'archivage et la diffusion; films pour la télévision
BS	Service de radiodiffusion sonore
BT	**Service de radiodiffusion télévisuelle**
F	Service fixe
M	Services mobile, de radiorepérage et d'amateur y compris les services par satellite associés
P	Propagation des ondes radioélectriques
RA	Radio astronomie
RS	Systèmes de télédétection
S	Service fixe par satellite
SA	Applications spatiales et météorologie
SF	Partage des fréquences et coordination entre les systèmes du service fixe par satellite et du service fixe
SM	Gestion du spectre
SNG	Reportage d'actualités par satellite
TF	Émissions de fréquences étalon et de signaux horaires
V	Vocabulaire et sujets associés

***Note**: Cette Recommandation UIT-R a été approuvée en anglais aux termes de la procédure détaillée dans la Résolution UIT-R 1.*

Publication électronique
Genève, 2023

RECOMMANDATION UIT-R BT.2154-0

Architecture de système de haut niveau pour la vidéo en immersion pour la présentation sur différents types de dispositifs d'affichage

(Questions UIT-R 140-1/6 et UIT-R 143-2/6)

(2022)

Domaine d'application

La présente Recommandation définit une architecture de système de haut niveau pour les vidéos en immersion destinées à être présentées sur différents types de dispositifs d'affichage. L'architecture comprend des objets vidéo, une description de scène, un système de restitution et un lecteur, qui constituent l'ensemble minimal de ses composantes. La présente Recommandation identifie également les informations à transférer entre le système de restitution et le lecteur.

Mots clés

Vidéo en immersion, 6DoF, description de scène, vidéo volumétrique, adaptation aux dispositifs

L'Assemblée des radiocommunications de l'UIT,

considérant

a) que les vidéos en immersion, qui permettent aux utilisateurs finals de se déplacer dans un espace vidéo et de regarder la vidéo de façon omnidirectionnelle, depuis tout point de vue, offrent une nouvelle expérience visuelle améliorée;

b) que la vidéo en immersion est représentée moyennant l'organisation d'objets vidéo tels que la vidéo volumétrique, la vidéo omnidirectionnelle et la vidéo bidimensionnelle dans un espace tridimensionnel;

c) que divers types de dispositifs d'affichage sont à la disposition des utilisateurs finals (par exemple, des visiocasques, des smartphones et des tablettes) et doivent être pris en compte pour la présentation de vidéos en immersion;

d) qu'un nombre croissant de plates-formes de diffusion interactives sont disponibles pour distribuer des contenus au public;

e) que des serveurs situés sur des réseaux, y compris en nuage et en périphérie, dotés d'une puissance de calcul accrue, seront utilisés de manière efficace pour restituer la vidéo en immersion en s'adaptant aux divers types de dispositifs d'affichage possédant différentes capacités de calcul et d'affichage;

f) qu'une architecture commune pour les vidéos en immersion destinées à être présentées sur divers types de dispositifs d'affichage facilitera le développement de systèmes et d'applications vidéo en immersion,

reconnaissant

a) la série de normes ISO/IEC 23090 – Technologie de l'information – Représentation codée de médias immersifs;

b) la Recommandation UIT-R BT.2123 – Valeurs de paramètres vidéo des systèmes audiovisuels en immersion évolués pour la production et l'échange international de programmes de radiodiffusion;

c) le Rapport UIT-R BT.2420 – Ensemble de scénarios d'utilisation et état actuel des systèmes audiovisuels en immersion évolués,

recommande

que les systèmes de vidéo en immersion destinés à divers types de dispositifs d'affichage soient conçus conformément à l'architecture de système de haut niveau décrite dans l'Annexe.

Annexe

Architecture de système de haut niveau pour la vidéo en immersion pour la présentation sur différents types de dispositifs d'affichage

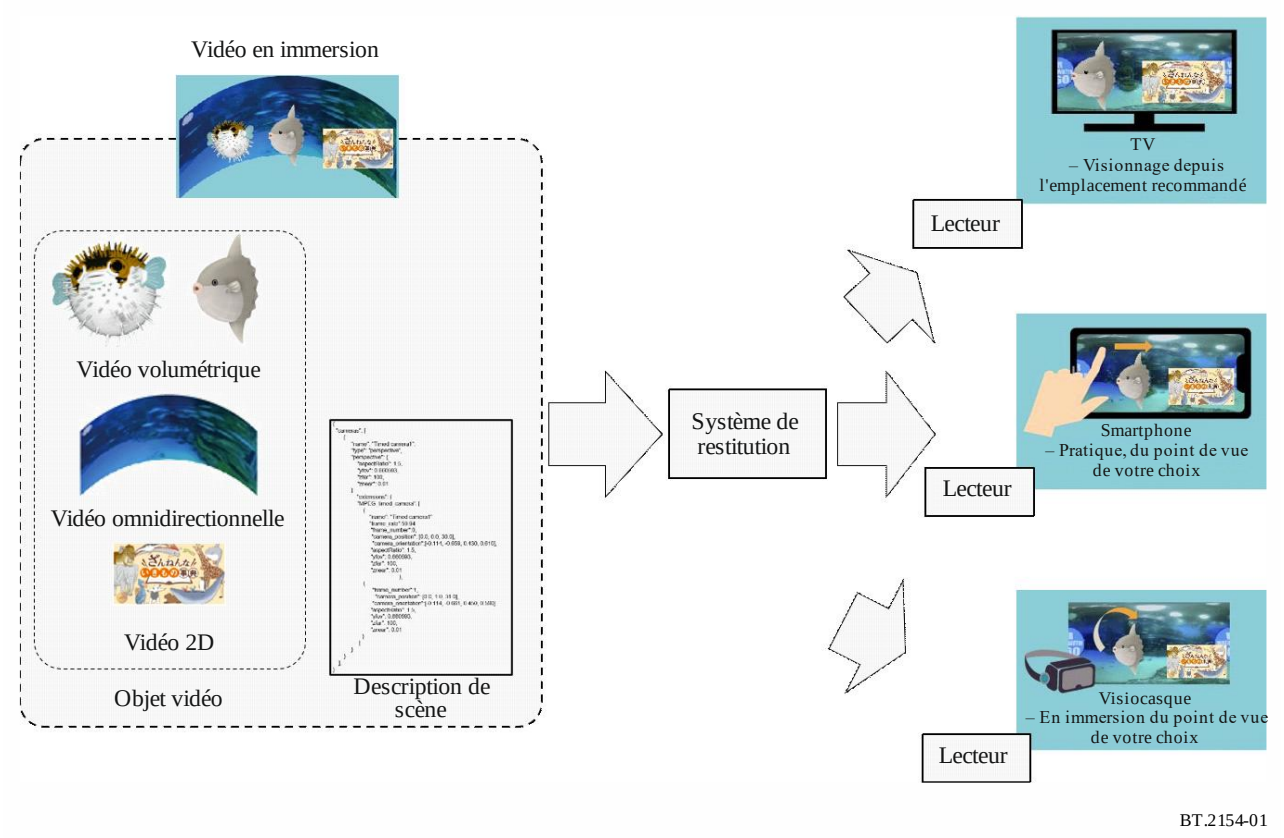
1 Vue d'ensemble

La Fig. 1 donne un aperçu de l'architecture de système pour la vidéo en immersion, de sa composition à sa présentation sur divers types de dispositifs d'affichage.

La vidéo en immersion se compose de la description de scène et de multiples objets vidéo auxquels il est fait référence à partir de la description de scène, y compris la vidéo volumétrique, qui peut représenter la forme et la texture des objets en trois dimensions, la vidéo omnidirectionnelle autour des objets, et la vidéo bidimensionnelle afin que les utilisateurs puissent regarder la vidéo à partir de n'importe quelle position dans n'importe quelle direction. La vidéo omnidirectionnelle et la vidéo 2D peuvent contenir des informations de profondeur sur ces objets. Les descriptions de scène fournissent une représentation tridimensionnelle en séries chronologiques de la vidéo en immersion, qui comprend la position, l'orientation et la taille de chaque objet, ainsi que la disposition spatiale et temporelle dans l'espace tridimensionnel.

FIGURE 1

Vue d'ensemble de la vidéo en immersion, de la composition à la présentation



Un système de restitution construit un espace tridimensionnel à partir de divers objets vidéo indiqués par la description de scène et produit une vidéo qui sera visionnée depuis la position et dans la direction choisies par l'utilisateur. Le lecteur de chaque dispositif d'affichage présente la vidéo à partir de la vidéo restituée de la façon la plus adaptée au dispositif, en tenant compte de la position et de la direction de visionnage choisies par l'utilisateur.

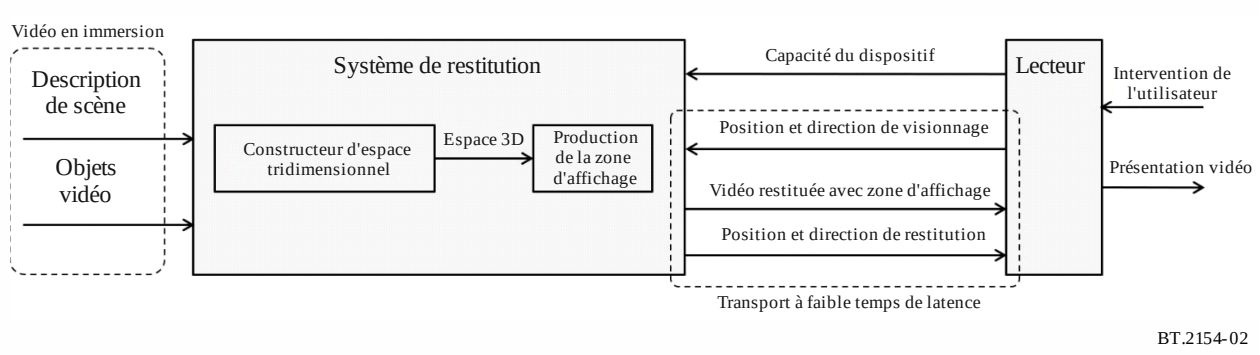
2 Architecture de système de haut niveau

2.1 Définition

L'architecture de système de haut niveau décrite dans la Fig. 2 est définie pour présenter la vidéo en immersion sur différents types de dispositifs d'affichage.

FIGURE 2

Architecture de système de haut niveau pour la vidéo en immersion



2.2 Vidéo en immersion

La vidéo en immersion représente un espace tridimensionnel en séries chronologiques et se compose d'objets vidéo et de la description de scène.

Les objets vidéo comprennent la vidéo volumétrique, qui représente la forme et la texture des objets en trois dimensions, la vidéo omnidirectionnelle autour des objets, et la vidéo rectangulaire bidimensionnelle (2D). La vidéo omnidirectionnelle et la vidéo 2D peuvent être associées à des informations de profondeur.

La description de scène définit un espace tridimensionnel en séries chronologiques en faisant référence aux multiples objets vidéo et en identifiant la position, l'orientation et la taille de chaque objet ainsi que sa disposition spatiale et temporelle dans cet espace tridimensionnel.

La description de scène peut également contenir des informations sur la position et la direction de visionnage que le créateur du contenu recommande à l'utilisateur de choisir, c'est-à-dire la zone d'affichage recommandée, pour la présentation sur le dispositif d'affichage.

2.3 Système de restitution et lecteur

Un système de restitution construit un espace tridimensionnel à partir de divers objets vidéo et de la description de scène. Il produit également la vidéo à visionner compte tenu de la position et de la direction choisies par l'utilisateur.

Le lecteur présente la vidéo à partir de la vidéo restituée de la façon la plus adaptée au dispositif, en tenant compte de la position et de la direction de visionnage choisies par l'utilisateur.

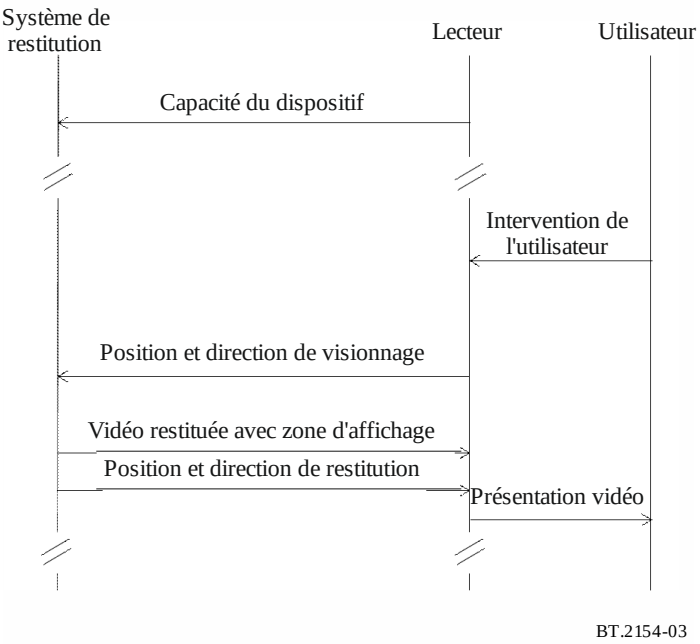
Les fonctions du système de restitution et du lecteur devraient être séparées de sorte que le lecteur n'ait pas à traiter toutes les informations dans l'espace tridimensionnel, mais seulement une partie qui doit être présentée sur le dispositif. Cette séparation réduit le volume des données des objets vidéo que le lecteur doit traiter et la charge de traitement du lecteur, ce qui permet la mise en œuvre de lecteurs dont la capacité de traitement est plus légère. Même lorsque de nouveaux types d'objets vidéo seront introduits, seul le système de restitution devra être mis à jour pour les prendre en charge, sans mise à jour du lecteur.

2.4 Informations à transférer entre le système de restitution et le lecteur

La Fig. 3 montre le flux d'informations à transférer entre le système de restitution et le lecteur.

FIGURE 3

Informations à transférer entre le système de restitution et le lecteur



1) Avant de commencer le processus de restitution, le système de restitution construit un espace tridimensionnel sur la base de la description de scène et des objets vidéo, et le lecteur notifie au système de restitution la capacité du dispositif, notamment sa résolution d'affichage, son champ de vision et sa fréquence d'image.

2) Lorsqu'un utilisateur commence à visionner le contenu, le lecteur notifie au système de restitution la position et la direction de visionnage choisies par l'utilisateur, qui peuvent changer au cours du visionnage de la vidéo, en fonction des interventions réalisées par l'utilisateur.

3) À partir de l'espace tridimensionnel, le système de restitution produit la vidéo incluant la zone d'affichage qui sera présentée en fonction de la position et la direction de visionnage choisies par l'utilisateur notifiées. Le système de restitution peut produire une vidéo pour un éventail de zones plus étendu que la zone d'affichage dans le but de prendre en charge le déplacement rapide de la position et de la direction de visionnage. En outre, le système de restitution peut produire une vidéo dans la zone d'affichage recommandée sur la base des informations relatives à la zone d'affichage recommandée présentes dans la description de scène, le cas échéant.

4) La vidéo restituée est transférée au lecteur avec l'indication de la position et de la direction de restitution dans l'espace tridimensionnel utilisé lorsque la vidéo a été produite. Un transport à faible temps de latence doit être utilisé pour le transfert de la vidéo et des informations concernant la position et la direction de restitution.

5) Le lecteur présente la totalité ou une partie de la vidéo transférée compte tenu de la position et de la direction de visionnage choisies par l'utilisateur.

Pièce jointe
à l'Annexe
(informative)

Exemple de mise en œuvre d'une architecture de système de haut niveau

1 Vue d'ensemble

La présente Pièce jointe décrit un exemple de système qui met en œuvre l'architecture de système de haut niveau définie dans l'Annexe pour la vidéo en immersion pour la présentation sur différents types de dispositifs d'affichage.

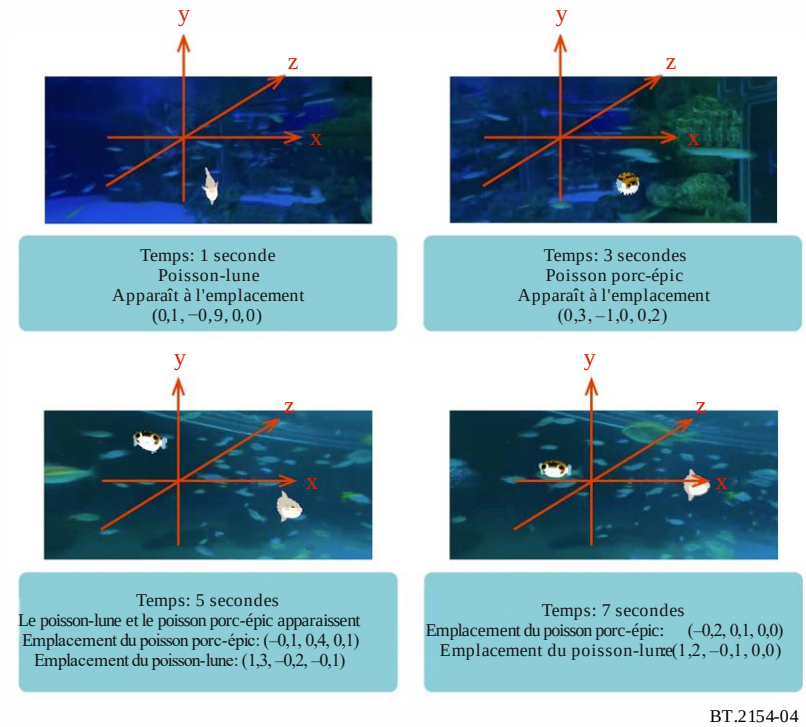
2 Vidéo en immersion

2.1 Description de scène

Le concept de description de scène est illustré à la Fig. 4. Comme indiqué dans cette Figure, à 1 seconde, un objet poisson-lune apparaît à l'emplacement (0,1, −0,9, 0,0) dans l'espace tridimensionnel. Deux secondes plus tard, soit à 3 secondes, un poisson porc-épic apparaît à l'emplacement (0,3, −1,0, 0,2). À cet instant, l'objet poisson-lune a disparu. Ainsi, la description de la scène spécifie la position, l'orientation et la taille des objets dans l'espace tridimensionnel à chaque instant.

FIGURE 4

Disposition d'objets en séries chronologiques dans un espace tridimensionnel utilisant la description de scène



BT.2154-04

Dans cet exemple, le format étendu du format GL Transmission (glTF2), dont les spécifications figurent à l'adresse https://github.com/KhronosGroup/glTF/tree/master/specification/2.0, est utilisé pour la description de scène. On trouve dans la Figure 5 un exemple de description de scène.

FIGURE 5

Exemple de description de scène


```
{
    "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, -50.611990661056071],
    "visible": 1
}.
```


BT.2154-05

2.2 Objet vidéo

À titre d'objets vidéo pour la vidéo volumétrique, on utilise des flux de nuage de points obtenus par compression de la vidéo volumétrique au format nuage de points selon la norme ISO/CEI 23090-5 «Codage basé sur la vidéo volumétrique (V3C) et compression de nuage de points basée sur la vidéo».

À titre de vidéo omnidirectionnelle, on utilise la vidéo à 360 degrés convertie par projection équirectangulaire obtenue selon la norme ISO/IEC 23090-2 «Format de média omnidirectionnel».

De plus, la vidéo rectangulaire bidimensionnelle est utilisée pour la présentation en superposition.

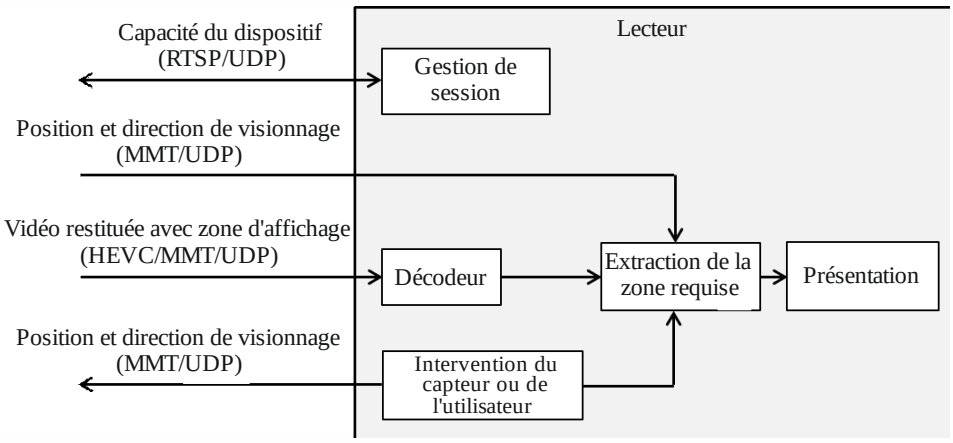
3 Mise en œuvre du système de restitution et du lecteur

3.1 Mise en œuvre du lecteur

On a conçu des lecteurs pour visiocasque et smartphone/tablette et des lecteurs pour téléviseur traditionnel, respectivement. Le lecteur utilisé pour les téléviseurs traditionnels ne permet pas à l'utilisateur de modifier sa position et sa direction de visionnage. Les blocs fonctionnels de ces dispositifs sont décrits dans les Fig. 6 et 7.

FIGURE 6

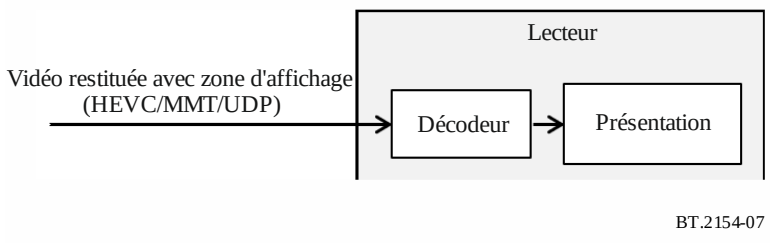
Blocs fonctionnels du lecteur pour visiocasque et smartphone/tablette



BT.2154-06

FIGURE 7

Blocs fonctionnels du lecteur pour téléviseur théoriquement sans dispositif permettant l'intervention de l'utilisateur



Le lecteur utilise la méthode SETUP d'un protocole de transmission en continu et en temps réel (RTSP, IETF RFC 7826) pour établir une session avec le serveur et communiquer au serveur ses capacités, notamment sa résolution d'affichage, sa fréquence d'image, son champ de vision et la méthode de codage disponible utilisée pour compresser la vidéo avec la zone d'affichage.

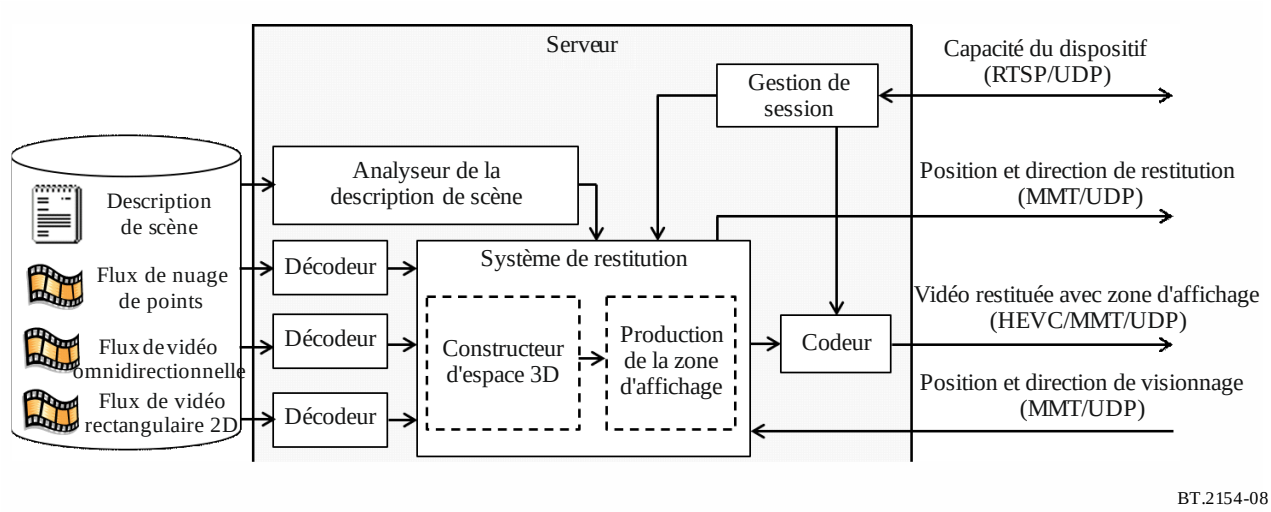
La position et la direction de visionnage choisies par l'utilisateur sont notifiées au serveur dans le format d'un message de transport de médias MPEG (MMT, ISO/IEC 23008-1). Dans le cas d'un visiocasque, la position et la direction de visionnage sont déterminées en fonction du déplacement de l'utilisateur et, dans le cas d'un smartphone ou d'une tablette, en fonction des interventions effectuées sur l'écran par l'utilisateur.

3.2 Mise en œuvre du système de restitution

Un serveur doté d'une fonction de restitution est élaboré séparément du lecteur. Le type de lecteur requis varie en fonction du type de dispositif, mais le serveur est le même, quel que soit le type de lecteur. La Fig. 8 décrit les blocs fonctionnels du serveur doté d'une fonction de restitution.

FIGURE 8

Blocs fonctionnels du serveur doté d'une fonction de restitution



Le serveur analyse les descriptions de scène, décode les objets vidéo requis en temps réel et les dispose dans l'espace tridimensionnel conformément aux descriptions de scène. Puis, à partir de l'espace tridimensionnel, le système de restitution produit une vidéo sous la forme d'une zone d'affichage dont la résolution est conforme à la position et à la direction de visionnage notifiées par le lecteur. Une autre zone d'affichage est produite à partir des informations relatives à la zone d'affichage recommandée contenues dans les descriptions de scène pour le dispositif en l'absence de

notification des informations relatives aux capacités du dispositif n'est communiquée et de modification de la position/direction du dispositif.

La vidéo avec zone d'affichage produite par le système de restitution est compressée selon une norme de codage vidéo à haute efficacité (HEVC, ISO/IEC 23008-2 | Rec. UIT-T H.265) pour obtenir une vidéo bidimensionnelle et transportée vers le lecteur au format MMT. Simultanément, les paramètres de restitution utilisés pour produire la zone d'affichage sont transférés au lecteur dans un format de message MMT.

4 Présentation sur trois types de dispositifs d'affichage différents

4.1 Visiocasque

Comme indiqué dans la Fig. 9, utiliser un visiocasque permet aux utilisateurs d'apprécier une vidéo depuis la position et selon l'angle de leur choix tout en se déplacement librement dans le cadre d'une expérience hautement immersive. L'utilisateur peut ainsi visualiser les objets non seulement depuis l'avant, mais aussi depuis l'arrière et le côté.

FIGURE 9

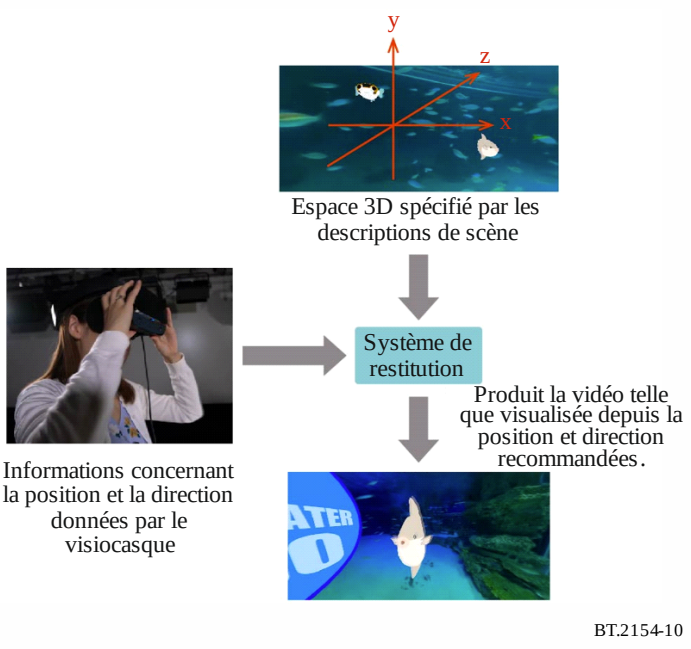
Visionnage avec un visiocasque

BT.2154-09

Dans le visiocasque, le système de restitution produit la vidéo en fonction de la position et de la direction de visionnage de l'utilisateur détectées par les capteurs du visiocasque, et le lecteur présente la vidéo produite sur le visiocasque. Le mécanisme de présentation sur un visiocasque est décrit dans la Fig. 10.

FIGURE 10

Mécanisme de présentation sur un visiocasque



4.2 Smartphone

La position et la direction de visionnage peuvent être modifiées par des interventions réalisées sur l'écran du smartphone, ce qui permet aux utilisateurs de choisir d'où et sous quel angle ils souhaitent visionner une vidéo (voir la Fig. 11).

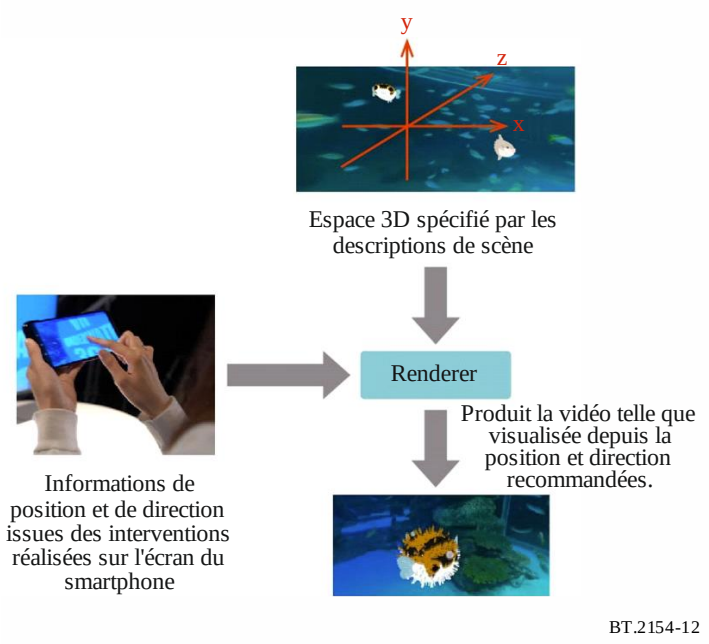
FIGURE 11

Visionnage sur smartphone

Comme lors de la présentation sur des visiocasques, le système de restitution produit une vidéo à présenter sur des smartphones en se basant sur les descriptions de scène. Sur les smartphones, le lecteur présente la vidéo conformément à la position et à la direction spécifiées par l'utilisateur via ses interventions sur l'écran (voir la Fig. 12).

FIGURE 12

Mécanisme de présentation sur smartphone



4.3 Téléviseur

Bien que les utilisateurs ne puissent pas changer l'emplacement et l'angle d'un téléviseur comme c'est le cas avec les visiocasques et les smartphones, ils peuvent toujours profiter facilement de la vidéo en respectant la position et la direction de visionnage recommandées par le créateur du contenu (voir la Fig. 13).

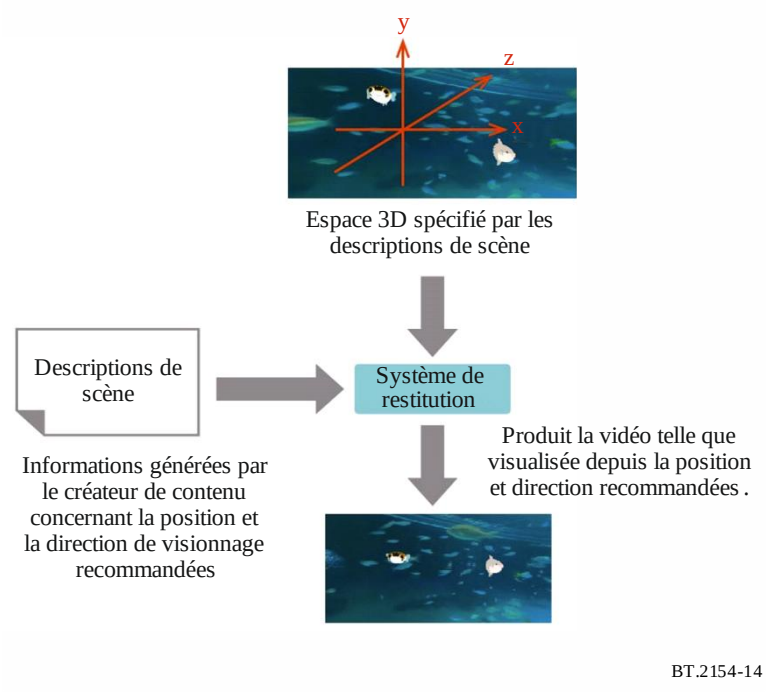
FIGURE 13

Visionnage sur un téléviseur

Même dans ce cas, le système de restitution produit la vidéo sur la base des informations d'espace tridimensionnel spécifiées par les descriptions de scène. Toutefois, l'utilisateur n'intervenant pas, les informations sur la position et la direction de visionnage sont fournies par les descriptions de scène en tant qu'informations sur la zone d'affichage recommandée. Conformément à ces informations, le système de restitution produit la vidéo qui sera présentée comme indiqué à la Fig. 14.

FIGURE 14

Mécanisme pour la présentation sur un écran sans intervention de l'utilisateur (téléviseur)



5 Références

La présentation sur trois types de dispositifs est disponible à l'adresse URL suivante: https://www.nhk.or.jp/strl/english/open2021/tenji/3/index.html

Les spécifications à utiliser lors de la mise en œuvre sont les suivantes:

Recommandation UIT-T H.265 | ISO/CEI 23008-2 (2020): Technologies de l'information – Codage à haute efficacité et livraison des médias dans des environnements hétérogènes – Partie 2: Codage vidéo à haute efficacité.

ISO/CEI 23008-1:2017: Technologies de l'information – Codage à haute efficacité et livraison des médias dans des environnements hétérogènes – Partie 1: Transport des médias MPEG.

ISO/CEI 23090-2:2021: Technologies de l'information – Représentation codée de média immersifs – Partie 2: Format de média omnidirectionnel.

ISO/CEI 23090-5:2021: Technologie de l'information – Représentation codée de médias immersifs – Partie 5: Codage basé sur la vidéo volumétrique (V3C) et compression de nuage de points basée sur la vidéo (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, disponible à l'adresse https://github.com/KhronosGroup/glTF/tree/master/specification/2.0/.