**Impact des ROI sur le streaming vidéo**

Encadrante : M FOURMAUX Olivier

Etudiants : **DUAN Jitong**

**OBLETTE Edgar**

**Table des matières**

[1 Cahier des charges 1](#_Toc62604573)

[1.1 Contexte de projet 2](#_Toc62604574)

[1.2 Objectifs à atteindre 2](#_Toc62604575)

[2 Plan de développement 2](#_Toc62604576)

[3 Bibliographie 3](#_Toc62604577)

[3.1 Vidéo VR à trois degrés de liberté 3](#_Toc62604578)

[3.1.1 Architecture de communication 3](#_Toc62604579)

[● Vidéo panoramique monoculaire 3](#_Toc62604580)

[● Vidéo panoramique binoculaire 3](#_Toc62604581)

[● Vidéo VR180 3](#_Toc62604582)

[3.1.2 Six degrés de liberté de la vidéo RV (degrés de liberté, DoF) 4](#_Toc62604583)

[3.2 Stratégie de transmission DASH 4](#_Toc62604584)

[3.2.1 Stratégie de contrôle des caches 5](#_Toc62604585)

[3.2.2 Stratégies de contrôle de la bande passante 5](#_Toc62604586)

[4 Analyse 7](#_Toc62604587)

[4.1 Technologie de transport dynamique adaptative basée sur HTTP 8](#_Toc62604588)

[4.2 Technologie de transmission dynamique adaptative basée sur DASH pour la vidéo panoramique 9](#_Toc62604589)

[5 Conception 10](#_Toc62604590)

[5.1 Sélection de la couche de carrelage 10](#_Toc62604591)

[6 Etat d'avancement 12](#_Toc62604592)

[7 Référence 13](#_Toc62604593)

# Cahier des charges

Nous réalisons ce projet dans le cadre l’UE PRES du premier semestre du master informatique de l’université de la Sorbonne. Le projet est composé de deux livrables, le premier est le présent document qui a pour but de définir le périmètre du projet, un état de l’art sur l’existant dans la littérature et une solution proposée. Durant ce projet nous effectuant des points réguliers avec notre encadrant pour corriger les erreurs et améliorer la qualité de notre rendu, les réunions sont effectuées à distance afin de respecter les mesures sanitaires actuelles.

## Contexte de projet

La vidéo 360 attire de plus en plus l'attention dans le contexte de la réalité virtuelle (RV). En raison de ses exigences en matière de très haute résolution, les services professionnels de diffusion en continu de vidéo 360 existants souffrent de graves inconvénients. Par utiliser un nouveau système de diffusion en continu de bout en bout, du codage à l'affichage, pour transmettre des vidéos 360 de résolution 8K et offrir une expérience de RV améliorée grâce aux écrans montés sur le chauffage (HMD). Les principales contributions du système proposé concernent le carrelage, l'intégration du flux dynamique adaptatif MPEG sur la norme HTTP (DASH) et la sélection du niveau de débit en fonction du port d'affichage. Le tuilage et le streaming adaptatif permettent au système proposé de fournir une vidéo à 360 degrés de très haute résolution et de bonne qualité visuelle.

## Objectifs à atteindre

Nous avons quatre objectifs principaux à atteindre durant ce projet, on peut les regrouper en deux catégories: documentation et réalisation .

* Etudier les codec supportant l'encodage différencié de région d'Intérêt (ROI) pour mettre en oeuvre un système codeur/décodeur adaptatif au niveau spacial
* Analyser la transmission vidéo dynamique via HTTP et son extension à la ROI
* Démonstration de l'encodage différencié avec un casque de VR ou émulation sur un écran classique et de son impact sur le réseau

# Plan de développement

Pour le plan de développement, nous discuterons des différentes phases de notre traitement du sujet, tout d'abord, nous avons mené une étude documentaire sur les techniques de perception des points de vue en RV, nous avons passé par l'apprentissage des techniques clés du traitement et de la transmission vidéo en RV. Ensuite, nous avons travaillé sur la compréhension de la projection équirectangulaire, la puissance de calcul d'affichage des HMD et la définition du nombre de pixels par unité d'angle.

Ensuite, six semaines ont été consacrées à la rédaction de la documentation et nous avons fait des recherches approfondies qui nous ont permis de comprendre comment utiliser le concept de tuilage pour diviser chaque image vidéo en trames afin de les coder, transmettre et décoder efficacement, mais aussi d'apprendre les bases du langage de programmation python, ce qui nous a été très utile pour le reste de notre projet.

Après deux mois de recherche, nous avons réussi à faire des recherches et à étudier la littérature sur les techniques et les concepts pertinents.

Toutes ces recherches ont permis d'obtenir la base nécessaire pour simuler les techniques de l'encodage différencié de région d'Intérêt (ROI).

Pour la suite, nous envisageons de travailler sur les deux points primordiaux suivants :

* L'achèvement du codec sur le retour sur investissement du projet.
* La démonstration de la fonctionnalité du codec.

# Bibliographie

## Vidéo VR à trois degrés de liberté

### Architecture de communication

 Les réseaux véhiculaires peuvent être classés en trois types d’architectures :

### Vidéo panoramique monoculaire

Filmé et assemblé hors de l'image est une sphère, le film sphérique est une sorte de panoramique monoculaire (il manque juste une petite partie ci-dessous), à l'heure actuelle dans le réseau la plupart des vidéos RV sont des panoramiques monoculaires, c'est-à-dire que l'œil gauche et l'œil droit voient le même contenu.

### Vidéo panoramique binoculaire

La vidéo panoramique binoculaire utilise le principe de la parallaxe binoculaire pour présenter des images avec une certaine différence de perspective dans les yeux gauche et droit du casque VR, ce qui n'est pas différent des films en 3D en principe. Ce n'est qu'au cours du processus de génération de la vidéo panoramique binoculaire qu'elle peut être générée par la prise de vue parallèle avec une caméra à double objectif, ou par la prise de vue avec une caméra à objectif unique avec des informations de profondeur et la présentation de la parallaxe de l'œil gauche et de l'œil droit par le calcul de la profondeur.

### Vidéo VR180

Récemment, YouTube et Microsoft ont lancé conjointement un nouveau format vidéo appelé VR180, qui est équivalent à la moitié avant de la vidéo panoramique. Le haut et le bas du panorama sont également abrégés, et peuvent être visionnés sur le VR HMD ou sur un écran de télévision/PC ordinaire ou un terminal mobile. conférences, talk-shows, etc. Mais le VR180 ne peut pas remplacer les vidéos panoramiques/VR qui nécessitent un plus fort sentiment d'immersion. Le VR180 peut réduire le coût du tournage, faciliter le tournage, réduire le risque d'infraction, rendre la post-production ou la production instantanée plus pratique, abaisser le seuil du contenu VR et promouvoir la génération d'UGC. En même temps, le VR180 nécessite une faible largeur de bande de réseau à la même résolution, une faible puissance de calcul des terminaux et une grande accessibilité des terminaux, ce qui favorise le développement du commerce de la RV sur grand écran. Dans l'ensemble, le VR180 est un compromis sage et pragmatique du niveau technologique restrictif à ce stade, qui favorisera la génération de contenu de RV.

### Six degrés de liberté de la vidéo RV (degrés de liberté, DoF)

Indique la dimension mobile d'un objet, la plupart des HMD et des vidéos VR ne supportent actuellement que trois degrés de liberté, c'est-à-dire la rotation azimutale de la tête (lacet), la rotation en tangage (tangage) et la rotation panoramique (roulis) à une position de visualisation spécifique. Six degrés de liberté est basé sur trois degrés de liberté, plus la variation spatiale de la position de visualisation : X, Y, Z. La vidéo à six degrés de liberté est générée de trois façons : enregistrement par caméra à champ lumineux, modèle 3D et capture spatiale. Six degrés de liberté, s'ils sont enregistrés avec une caméra à champ clair (une matrice de plusieurs caméras), constituent un pic de trafic (et un énorme défi pour la compression vidéo multi-angle) et ont une plage de mouvement limitée pour un visionnage libre. Six degrés de liberté dans le modèle 3D, similaires au point de vue subjectif et aux changements de perspective dans les jeux 3D, se situent en fait dans la plage de la RV CG plutôt que dans la plage de la RV vidéo, et permettent également d'obtenir un point de vue libre. L'approche dite de capture spatiale est une fusion des deux premiers moyens. Un modèle 3D du milieu environnant est généré par balayage LIDAR pendant que l'appareil photo est en train de filmer, et la cartographie est effectuée par la suite. Il peut être un peu plus facile et plus rapide de générer des vidéos à six degrés de liberté, mais le modèle d'environnement qu'il génère est basé sur un point de vue, de sorte que le modèle d'environnement 3D et le modèle d'objet présentent une certaine déviation. Bien que les six degrés de liberté n'en soient qu'à leurs débuts, le développement de la RV à six degrés de liberté est la direction future inévitable d'une plus grande immersion.

## Stratégie de transmission DASH

La diffusion en continu adaptative est basée sur la stratégie de transmission HITTP MPEG proposée en 2012, qui convient aux médias en continu segmentés. Dans la stratégie DASH, la source vidéo est d'abord découpée en tranches de longueur égale et de qualité différente et stockée dans le serveur, puis le mécanisme "pulls the stream" du client est utilisé pour combiner la capacité de la bande passante disponible avec les informations de débit moyen des tranches enregistrées et la tranche de représentation des médias du chemin de téléchargement du serveur (Media La fonction PresentationDescription (MPD) ajuste de manière adaptative la demande de tranche de code et sélectionne la tranche de code avec le taux de code adapté à la capacité de transport de la bande passante à télécharger. L'application pratique montre que cette méthode de déploiement est plus simple et plus facile à utiliser que le déploiement côté serveur. Les politiques DASH existantes selon le mécanisme du principe de contrôle peuvent être divisées en trois catégories : politique de contrôle du cache, politique de contrôle de la bande passante et politique de contrôle hybride. L'objectif de la politique de contrôle du cache est de garantir une lecture continue et fluide des flux vidéo et une commutation de débit binaire aussi fluide que possible pour éviter les débordements et les sous-débordements du cache. L'objectif d'optimisation de la politique de contrôle de la bande passante est d'augmenter le débit binaire autant que possible en sélectionnant de manière adaptative le débit binaire qui correspond à la capacité de transport de la bande passante disponible. La stratégie hybride est un mélange des deux précédentes.

### Stratégie de contrôle des caches

La mise en cache est importante dans la stratégie DASH. C'est la présence d'une file d'attente en cache qui donne au client DASH la possibilité de faire face au risque de fluctuations des canaux. S'il n'y a pas de cache, dans le cas où le débit de la vidéo en continu est supérieur à la capacité de la bande passante disponible, le goulot d'étranglement de la bande passante sera l'encombrement et la perte de paquets, ce qui entraînera un téléchargement vidéo intempestif, suivi d'un débordement du cache, d'un décalage de la lecture. La politique de contrôle du cache consiste à fixer une valeur seuil pour le cache, en enregistrant et en prédisant la tendance des modifications du cache, et en prenant différents plans de réponse à différents niveaux du cache pour garantir un débit binaire élevé tout en évitant les retards. Dans la littérature [2], des auteurs ont proposé une stratégie DASH et une amélioration de la qualité de l'environnement. La stratégie exposée dans le document fixe un seuil de réservation pour la longueur du cache afin d'éviter les retards et de minimiser les fluctuations du débit binaire vidéo pendant la lecture. La politique est principalement divisée en deux parties, le module d'estimation de la largeur de bande et le module de sélection du débit binaire. Le premier est chargé d'estimer la capacité de transport de la bande passante disponible et de transmettre les résultats au second. Ce dernier est lancé après le téléchargement de chaque film en code. L'estimation de la largeur de bande est équilibrée entre l'estimation de la largeur de bande basée sur la RTT et la stratégie d'estimation de la largeur de bande basée sur la tranche de code en utilisant une stratégie de prédiction par étapes. Le module de sélection du taux de codage est mis en œuvre dans le but de minimiser le plongeon de la sélection du taux de codage. Dans la littérature [3], les auteurs proposent une stratégie d'adaptation du taux de code pour construire un mécanisme de gestion du cache de diffusion basé sur le débit disponible afin d'améliorer la qualité d'expérience de l'utilisateur en éliminant les tranches de code à faible taux de code pour remplir les tranches à taux de code de haute qualité, et le biais du client vers l'état du réseau en présence de contention de flux TCP est atténué par la randomisation des demandes de tranches de code. Le débit disponible dans la politique présentée dans le document est une valeur lissée du débit historique observé.

### Stratégies de contrôle de la bande passante

La bande passante est également très importante dans la stratégie DASH. Le changement adaptatif du taux de codage est basé sur le changement de la largeur de bande, une largeur de bande élevée sélectionne un taux de codage élevé pour améliorer l'expérience de l'utilisateur, une largeur de bande faible sélectionne un taux de codage faible pour éviter le décalage. Dans le DASH traditionnel, les informations sur la bande passante proviennent principalement du taux de téléchargement de la tranche de code du client. Selon la définition standard et la simplification dans l'application pratique, l'estimation de la largeur de bande existante est basée sur le fait que le client ne fournit que les informations sur le taux de code moyen pour une tranche de code au moment de la décision, et n'enregistre que la taille de la tranche de code et le temps de téléchargement après la fin du téléchargement, de sorte que seule la moyenne du taux de téléchargement de la tranche de code peut être obtenue. Dans la littérature [4], les auteurs proposent une stratégie adaptative dynamique pour l'estimation de la largeur de bande avec des fenêtres pondérées basées sur le Sliding Percentile (SP). Les auteurs suivent de près la bande passante historique pour s'assurer que la bande passante estimée peut être rapidement convergée lorsque le réseau change soudainement. Les données historiques sur la bande passante sont une moyenne mobile des estimations de la bande passante. Bien entendu, certains chercheurs ont proposé des optimisations pour ce système, en faisant valoir que l'enregistrement de la seule valeur moyenne est inexact. Dans la littérature [5], les auteurs proposent ensuite une stratégie d'adaptation du taux de codage avec un mécanisme de décision qui inclut une fonction objective locale d'amélioration de la qualité de l'environnement pour atteindre un équilibre entre les différentes demandes. Le fichier MPD est étendu pour enregistrer le taux de code instantané, de sorte que la sélection de la tranche de code peut suivre la fluctuation instantanée du taux de code. Les résultats expérimentaux présentés dans le document démontrent l'importance de la prédiction instantanée du taux de codage et de la tranche de code. Les résultats expérimentaux montrent que la stratégie présentée dans le document est plus robuste et qu'elle permet une prédiction plus précise de la largeur de bande par rapport à d'autres stratégies. Toutefois, ce système modifie le cadre DASH et le format de transmission des signaux existants, et la complexité est considérablement accrue.

* **Politique de contrôle des hybrides**

C'est une stratégie de contrôle qui combine les deux politiques ci-dessus pour gérer le cache et estimer la bande passante en même temps, et les deux travaillent ensemble pour maximiser l'expérience de l'utilisateur. Dans la littérature [6], une stratégie adaptative tenant compte de la tranche de code est proposée pour obtenir un meilleur mécanisme de gestion de la qualité de l'environnement pour les systèmes DASH en considérant le débit mesuré, l'occupation du cache et la variation de la taille de la tranche de code. Dans le document [7], un mécanisme DASH basé sur un contrôle de qualité MPD et de tranche de code amélioré qui peut améliorer la qualité de l'environnement est proposé. Dans l'article, les auteurs proposent que lorsque la longueur du cache de lecture n'est pas suffisante pour permettre au client de conserver une lecture fluide jusqu'au téléchargement de la prochaine tranche de code au taux de code actuellement sélectionné, le taux de code sélectionné pour la prochaine tranche de code sera dégradé. Lorsque la durée du cache de lecture est suffisamment longue, même une baisse soudaine de l'état du réseau n'affecte pas la fluidité de la lecture jusqu'à la tranche de code suivante, le taux de code de la sélection de la tranche de code suivante sera mis à niveau. Les résultats expérimentaux présentés dans cet article montrent que la stratégie d'optimisation proposée peut garantir une lecture fluide et un changement de débit binaire en douceur avec une faible latence. De plus, lorsque la largeur de bande du réseau diminue, la stratégie proposée peut atteindre un débit binaire plus élevé par rapport aux autres stratégies car sa sélection de débit binaire dépend du cache de lecture. Une politique adaptative d'optimisation de la qualité de l'environnement basée sur la méthode Q-learningl19 est proposée dans la littérature [8]. Le Q-learning est une méthode d'apprentissage qui permet à un système dynamique avec un ensemble d'états possibles et un algorithme de choisir l'action la plus appropriée en fonction de l'état dans lequel il se trouve. Chaque processus de décision génère une récompense qui influence les décisions futures. Dans le document, le Q-learning est utilisé comme un processus de décision dans un système DASH pour déterminer le taux de codage de la prochaine tranche de code à télécharger. Les résultats expérimentaux présentés dans le document montrent que la stratégie minimise le risque de décalage, sélectionne le taux de codage plus efficacement et a le moins de changement de taux de codage par rapport aux autres stratégies. Ses objectifs d'optimisation de la qualité de l'environnement prennent principalement en compte le changement de taux de code, le décalage et la qualité de la tranche de code demandée. Une stratégie d'optimisation en fonction du contenu est proposée dans la littérature [9]. La stratégie dans le document dépend de la variation de la longueur du cache de lecture,

# Analyse

La plupart des vidéos de VR sont actuellement diffusées en segments (segments) basés sur le Dynamic Adaptive Streaming over HTTP (DASH). Les segments sont généralement d'environ 1 à 15 secondes, et dans le mécanisme de diffusion de la vidéo VR basé sur des tuiles, la vidéo entière est divisée en tuiles de plus petite granularité, chaque tuile peut être codée et décodée indépendamment, et chaque tuile a plusieurs versions de qualité pour une diffusion adaptative à différents points de vue. En plus du fichier de description du streaming MPD, le fichier de description de la vidéo VR (virtual reality videodescriptor,VRD) et le fichier de description de la relation spatiale (SRD) peuvent être utilisés pour décrire la transmission vidéo.

Le VRD contient le format de projection et les informations d'orientation. le SRD contient la description de la qualité de chaque région de la tuile dans le cadre de projection. le MPD comprend la taille du segment vidéo, l'emplacement du fichier vidéo et la description des informations d'encodage. Sur la base des informations de description de MPD, VRD et SRD, le client réorganise et rend le fichier source vidéo reçu pour obtenir finalement le contenu vidéo VR complet. Par conséquent, le choix de la qualité du segment vidéo pour répondre à la demande de l'utilisateur du moment suivant est un autre défi plus difficile à relever pour la transmission vidéo RV. Les recherches actuelles sur la transmission vidéo adaptative de la VR sont divisées en trois catégories principales : la transmission vidéo adaptative de la VR basée sur les tuiles ; la transmission vidéo adaptative de la VR basée sur le point de vue ; la transmission vidéo adaptative de la VR basée sur la prédiction du point de vue.

Selon le schéma de transmission adaptative FoV de l'utilisateur, les premières tuiles multiples sont divisées sur l'espace de la séquence vidéo à 360 degrés, et la description de la relation spatiale MPEG-DASH SRD est utilisée pour déterminer les informations spatiales de chaque tuile par rapport à l'ensemble de la séquence vidéo rectangulaire à 360 degrés. Ensuite, la sphère 3D est divisée en carreaux à mailles multiples (mesh), à chacun desquels est attribué un identificateur unique, et ils divisent l'ensemble de la sphère 3D en six carreaux à mailles, les carreaux supérieurs et inférieurs aux pôles, et la partie centrale en quatre carreaux à mailles.

Enfin, en fonction de certaines relations de cartographie, différentes tuiles sont mappées dans les tuiles de grille correspondantes, et la résolution des tuiles de grille est modifiée dynamiquement en fonction de la valeur de référence de l'utilisateur, et les tuiles vidéo des tuiles de grille dans la valeur de référence de l'utilisateur fournissent des résolutions plus élevées (chaque tuile fournit des résolutions différentes). La littérature propose un schéma de transmission vidéo adaptative à 360 degrés, en considérant encore une fois que la transmission de la totalité de la vidéo haute résolution à 360 degrés gaspille les ressources de la bande passante, et en considérant le retard de transmission au HMD et la variation des conditions du réseau de l'utilisateur, ils servent côté serveur pour générer plusieurs versions de la vidéo à 360 degrés, chacune avec différentes régions d'amélioration de la qualité, c'est-à-dire le centre d'accentuation de la qualité (QEC), où la vidéo à l'intérieur du QEC a une haute qualité, tandis que la vidéo à l'extérieur du QEC a une qualité dégradée, et l'utilisateur demande la vidéo de son FoV actuel, et le serveur, sur la base des informations du FoV de l'utilisateur, envoie le Le serveur envoie la vidéo de la version QEC la plus proche du FoV de l'utilisateur selon les informations du FoV de l'utilisateur, de sorte que les ressources de bande passante consommées sont beaucoup plus faibles par rapport à l'ensemble de la vidéo 360 de haute qualité, et les auteurs fournissent deux débits binaires différents pour la même version QEC en réponse aux changements des conditions du réseau de l'utilisateur. Les utilisateurs peuvent choisir la version QEC à haut ou bas débit de la vidéo en fonction des conditions de leur réseau. Le système d'amélioration de la qualité vidéo par la transmission de la partie FoV de l'utilisateur actuel permet d'économiser de la bande passante dans une certaine mesure, mais il augmente la dimension spatiale (FoV de l'utilisateur), ce qui rend difficile la mise en cache préalable des futurs clips vidéo, et une fois qu'il y a un processus de recapture, il est facile d'avoir un décalage vidéo, ce qui affecte la qualité d'expérience de l'utilisateur. (SVC), ils proposent un schéma de transmission vidéo adaptative à 360 degrés, dont l'idée principale est de mettre en cache une granularité à plus long terme en exploitant les caractéristiques de la couche de base et de la couche d'amélioration du SVC, dans laquelle la couche de base de la vidéo est nécessaire pour ignorer le point de vue de l'utilisateur et mettre en cache le et comme le point de vue de l'utilisateur change relativement rapidement, la couche d'amélioration peut mettre en cache une granularité de temps plus courte, ce qui permet d'atteindre l'objectif d'améliorer la qualité d'expérience de l'utilisateur, d'économiser la bande passante et de réduire le stockage côté serveur.

## Technologie de transport dynamique adaptative basée sur HTTP

C'est une sorte de technologie de solution de streaming côté serveur et côté client. Dans la technologie DASH, le serveur divise le contenu vidéo en tranches individuelles, et chaque tranche est encodée à différents débits binaires ou résolutions et autres facteurs, tandis que le client peut sélectionner la tranche de média appropriée en fonction de sa propre mémoire tampon et des conditions du réseau, ce qui permet de réaliser une technologie de transmission adaptative. Cela permet d'obtenir un changement transparent de contenu avec une qualité d'image différente et d'offrir une meilleure expérience de visualisation à l'utilisateur. Du côté du serveur HTTP, le fichier de description de la présentation des médias (MPD) et le fichier de segmentation du flux de médias sont stockés dans lequel la segmentation du flux de médias peut être stockée sous la forme de fichiers de segmentation simples ou multiples. Pour effectuer une transmission vidéo complète, le client DASH envoie d'abord une demande au serveur HTTP pour obtenir le fichier MPD, puis le client peut obtenir le contenu vidéo, la fréquence d'images, le débit binaire et d'autres informations en analysant le fichier envoyé par le serveur. Le client utilise ces informations pour sélectionner le segment de média approprié avec ses propres conditions de cache et de réseau. Le client envoie alors une requête HTTP GET au serveur pour obtenir les données de segmentation des médias. Enfin, lorsque le client obtient le segment média, il commence à décoder et à lire le segment média et envoie la demande de segment média suivante. En pratique, l'utilisation de la technologie de transmission DASH permet d'obtenir une transmission vidéo adaptative, l'utilisateur peut regarder la vidéo Le client DASH peut sélectionner la segmentation appropriée des médias en temps réel en fonction des conditions du réseau, une telle méthode de transmission tient compte des conditions de réseau en temps réel du réseau de l'utilisateur

## Technologie de transmission dynamique adaptative basée sur DASH pour la vidéo panoramique

* **Méthode de transmission des images**

Cette méthode de transmission est différente de la méthode de transmission en vue intégrale, qui ne transmet que le contenu vidéo du point de vue de l'utilisateur actuel. Sur la base des informations en temps réel du point de vue de l'utilisateur, la méthode de transmission intra-point de vue ne peut transmettre que les informations à l'intérieur du point de vue en temps réel, et les contenus vidéo à l'intérieur du point de vue sont tous transmis avec la même qualité. Cependant, cette méthode de transmission en direct peut entraîner un décalage ou un vide de l'écran lors de la visualisation de la vidéo panoramique, car les utilisateurs tournent régulièrement la tête ou le corps rapidement pendant la visualisation de la vidéo panoramique. Ce mode de transmission réduit donc également la qualité de l'expérience de l'utilisateur.

* **Méthode de transmission de haute qualité**

Comme la méthode de transmission en vue intégrale et la méthode de transmission en vue subjective entraîneront une dégradation de la qualité de l'expérience de l'utilisateur, de nombreux chercheurs ont proposé une méthode de transmission en vue subjective de haute qualité, ce qui signifie que la vidéo à l'intérieur de la vue de l'utilisateur est transmise en haute qualité, tandis que la vidéo à l'extérieur de la vue est transmise en basse qualité. En général, les vidéos panoramiques sont transmises par blocs, c'est-à-dire que dans le serveur, la vidéo panoramique est divisée en plusieurs blocs, puis chaque bloc est encodé en différents niveaux de qualité. Pendant la lecture, le client peut déterminer les blocs dans le champ de vision de l'utilisateur sur la base des informations de vision de l'utilisateur, puis transmettre ces blocs en haute qualité et les autres blocs en basse qualité. Une telle méthode de transmission peut économiser les ressources de bande passante du réseau sans fil tout en garantissant que la qualité de l'expérience visuelle de l'utilisateur n'est pas affectée. Cependant, cette méthode de transmission de haute qualité en vue a un certain problème, à savoir que le serveur ne peut pas déterminer et ajuster rapidement les blocs dans la zone de vue de l'utilisateur en fonction des informations de vue de l'utilisateur fournies par le client en raison de la rotation fréquente et rapide de la tête de l'utilisateur. Pour résoudre ce problème, de nombreux chercheurs ont proposé une méthode de transmission vidéo panoramique adaptative basée sur la prédiction du point de vue en temps réel, qui permet de prédire le comportement futur de l'utilisateur en matière de rotation de la tête en se basant sur les informations historiques de la rotation de la tête de l'utilisateur. Cela permet au serveur de télécharger à l'avance des blocs dans le champ de vision de l'utilisateur en se basant sur les informations de rotation de la tête de l'utilisateur prévu.

**Technique de transmission multicast dynamique et adaptative pour la vidéo panoramique basée sur DASH:**

Dans les services de vidéo panoramique en ligne, il y a souvent plusieurs utilisateurs qui regardent la même vidéo panoramique en ligne en même temps, et comme les conditions de réseau sans fil des différents utilisateurs diffèrent, le serveur doit sélectionner des blocs vidéo pour chaque utilisateur simultanément et choisir le niveau de qualité approprié pour chaque bloc à transmettre au client DASH. Compte tenu de la qualité de l'expérience de visualisation des utilisateurs, le serveur doit sélectionner le niveau de qualité approprié pour chaque utilisateur afin de maximiser la qualité de l'expérience de visualisation des utilisateurs. Certains chercheurs ont proposé un schéma de transmission vidéo panoramique multi-utilisateurs qui maximise la QoE de tous les utilisateurs en prédisant les informations de point de vue de chaque utilisateur et en allouant les ressources en fonction des informations LTE ou WLAN de chaque utilisateur dans un réseau hétérogène où coexistent les utilisateurs LTE et les utilisateurs WIFI.

# Conception

## Sélection de la couche de carrelage

Le problème du choix de la couche de carrelage sera formulé. Le système doit diffuser en continu une vidéo de 360 images d'une durée de quelques secondes VD et d'une fréquence d'image de FPS d'un serveur à un client sur un réseau de communication dont la largeur de bande varie dans le temps. Sur le serveur, la vidéo est divisée en K segments, chacun ayant une durée de lecture de SD en secondes. Chaque segment est ensuite divisé en M petites parties appelées tuiles. Chaque tuile est encodée en L couches avec différents niveaux de qualité. La couche 0 est appelée couche de base, qui a la plus faible qualité. Les couches {1,1 <IS L-1} sont appelées enhancemen layers, qui fournissent une qualité vidéo supérieure. Une couche d'amélioration ne peut être décodée et lue que si la couche de base et toutes les couches d'amélioration inférieures sont disponibles chez le client.

Supposons qu'à un moment donné, le système doit adapter un segment pour répondre à une contrainte de débit Rc couche l (0 ≤ l ≤L - 1) de la tuile m (1 ≤ m ≤ M) du segment k a un débit de Rmlm et une qualité Qmlm . Soit Vn de note la position de la fenêtre de visualisation au niveau n th (1 ≤ n ≤ N = SD × FPS) trame du segment.

En règle générale, la position d'une fenêtre peut être définie par la longitude et la latitude du point central du point de vue.

Soit lm (0 ≤ lm ≤ L - 1) désigne la couche de qualité maximale sélectionné pour la tuile m (1 ≤ m ≤ M). Le problème de l'adaptation pour le segment k peut être formulé comme suit.

Trouver {1,2,..., ln} pour maximiser un objectif de qualité VQ. qui est une fonction des valeurs de qualité des tuiles {Qlmm} 1<=m<=M,0<=l<=L-1et des positions de la fenêtre de visualisation pendant la lecture du segment {Vn}1<=n<=SDxFPS



et satisfaire à la contrainte du débit binaire.



Dans la méthode que nous proposons, la contrainte de débit binaire Rc est simplement fixée au débit estimé Te, c'est-à-dire Rc = Te. Ici, le débit estimé Te est calculé comme la moyenne des derniers échantillons de débit S comme suit.



Ici, ilast désigne le dernier cycle de téléchargement. L'échantillon de débit Ts d'un cycle de téléchargement est calculé comme le rapport entre les données téléchargées et le temps nécessaire pour les télécharger. La position de la fenêtre de visualisation Vn (1 < n < N) est simplement estimée en utilisant la position de la fenêtre de visualisation que l'utilisateur est en train de regarder. Si Vcur est la position actuelle de la fenêtre et Vne la position estimée de la fenêtre, nous avons : Vne=Vcur,1 <=n<=N. Actuellement, l'objectif de qualité est calculé comme le débit binaire moyen de la fenêtre du segment comme suit. Soit wm (V), le poids de la tuile m étant donné une position de fenêtre V. L'objectif de qualité VQ est donné par



Comme dans [4], le poids wm(V) est calculé comme le rapport entre les pixels visibles de la tuile m et le nombre total de pixels dans la fenêtre d'affichage. {Vn1<= n<= N}, et l'objectif de qualité VQ pour les layers optimaux des tuiles du problème (1) peuvent être trouvés en effectuant une recherche complète sur toutes les sélections possibles. Néanmoins, une recherche complète peut prendre tellement de temps qu'elle ne convient pas à une adaptation en temps réel, en particulier lorsque le nombre de tuiles et/ou le nombre de couches sont importants. Notez que seules les tuiles visibles sont prises en compte dans l'algorithme proposé, car toutes les tuiles invisibles ont un poids de zéro. Le budget de bitrate Rc sera alloué aux tuiles visibles couche par couche. Les détails de l'algorithme proposé sont décrits dans l'algorithme 1.

* **Viewport-aware representation selection**

Compte tenu de la vue actuelle, un représentant optimal du débit binaire est sélectionné pour chaque tuile. Ce travail vise à donner plus d'importance aux tuiles dans la fenêtre, c'est-à-dire un débit plus élevé. Par conséquent, le nombre total de bits à attribuer aux tuiles qui se trouvent en dehors de la fenêtre d'affichage est réduit. La qualité visuelle de chaque tuile est améliorée en augmentant le débit binaire de la fenêtre d'affichage. Le débit binaire des tuiles situées à l'extérieur de la fenêtre de visualisation est réduit progressivement en fonction de la distance entre l'emplacement du centre sphérique de la fenêtre de visualisation et chaque tuile située à l'extérieur de la fenêtre de visualisation.

Le lecteur DASH VR proposé demande la représentation optimale du débit binaire pour chaque tuile en utilisant le MPD conçu. À cette fin, nous définissons V, Sin et Sout pour le viewport, un ensemble de tuiles à l'intérieur du viewport, et un ensemble de tuiles à l'extérieur du viewport, respectivement. Le débit binaire attribué à la ième tuile dans la fenêtre d'affichage est le suivant :



Où la tuile représente i, i∈Z et i ∈[1, N]. N désigne le nombre total de tuiles. est un terme constant défini par le client, qui est ∈ [0, 1]. Dans ce travail, est choisi comme 0.8 empiriquement. Rcur est la largeur de bande actuellement disponible, et wi est le poids de la ième tuile en Sin. Ce poids est calculé pour la ième tuile comme :



Où tot est le nombre total de pixels dans V. Pour répartir progressivement la largeur de bande restante entre les tuiles de la vue extérieure, la distance euclidienne, i est calculée entre le point central de V, et chaque point Fi défini pour chaque tuile de la vue extérieure. soust L'estimation du débit binaire pour la ième tuile dans sout est calculée comme suit



Enfin, le client demande une représentation bitrate pour chaque tuile, qui peut être obtenue comme suit :



où R= R=RSout∪RSin,Jest l'ID de la représentation DASH sélectionnée, J∈ {1,...,}donné est le nombre total de représentations, et rJ est le bitrate de la Jième représentation.

# Etat d'avancement

Dans les premières semaines, nous avons commencé par télécharger des documents sur Google scolar. Nous sommes restés sur ce même site pour comprendre la signification de chaque terme dans les références. Nous avons ensuite passé quelques semaines à travailler sur la documentation et nous avons fait beaucoup de recherches pour comprendre et nous familiariser avec le fonctionnement de la perspective. Nous travaillons actuellement sur des applications pratiques de cette technologie.

Nous examinerons ensuite comment concevoir les codecs, puis nous évaluerons et comparerons la précision des codecs proposés en tenant compte de la perspective.

# Référence

[1] Cagri Ozcinar, Ana De Abreu, Aljoscha Smolic.Viewport-aware adaptive 360° video streaming using tiles for virtual reality ,2017 IEEE International Conference on Image Processing

[2] Suh D, Jang I. Pack S. QoE-enhanced adaptation algorithm over DASH for multimedia streaming[C]//Information Networking (ICOIN), 2014 International Conference on. IEEE, 2014:497-501.

[3] Rahman W U, Yun D, Chung K. A client side buffer management algorithm to improve QoE[J].IEEE Transactions on Consumer Electronics, 2016. 62(4): 371-379

[4] Zhao S. Li Z. Medhi D, et al. Study of user QoE improvement for dynamic adaptive streaming over HTTP (MPEG-DASH)[CJ/Computing, Networking and Communications (1CNC), 2017 International Conference on. IEEE, 2017: 566-570.

[5] Yu L, Tillo T, Xiao J. QoE-driven dynamic adaptive video streaming strategy with future information[J]. IEEE Transactions on Broadcasting, 2017, 63(3): 523-534

[6] Juluri P, Tamarapalli V. Medhi D. QoE management in DASH systems using the segment aware rate adaptation algorithm[C]/Network Operations and Management Symposium (NOMS). 2016IEEE/IFIP. IEEE. 2016: 129-136.

[7] Lee S. Lee E, Lee H. Quality adaptation scheme for improving QoE of MPEG DASHIC]//nformation and Communication Technology Convergence (ICTC), 2016 Internationa Conference on. IEEE. 2016: 368-370

[8] Martin V. Cabrera J, García N. Design, optimization and evaluation of a Q-Learning HTTP Adaptive Streaming ClientjJ. IEE Transactions on Consumer Electronics. 2016, 62(4): 380-388.

[9] Sutton R S, Barto AG. Reinforcement learning: An introduction[M]. MIT press, 2018.

[10] Kim M, Park J, Chung K. Content-aware rate adaptation scheme to improve stability in HTTP Adaptive Streaming[C]/nformation Networking (ICOIN). 2017 International Conference on. IEEE.2017: 401-405

[11]Duc, Nguyen & Hoang, Trung & Hoang, Huong & Huong, Truong & Pham Ngoc, Nam & Cong Thang, Truong. (2019). Scalable 360 Video Streaming using HTTP/2. 10.1109/MMSP.2019.8901805.