

Page de garde

1 Présentation du projet

1.1 Objectif général

L'objectif général de notre projet de fin d'études, dit *PFE*, est d'étudier et de mettre en place une chaîne de transmission vidéo. Nous ferons une étude technique et la réalisation de l'ensemble des étapes : de la station émettrice au récepteur, en passant par les techniques de mise en réseau.

Nous avons décidé de nous intéresser plus particulièrement à l'amélioration de la qualité d'une zone géographique d'un flux vidéo, au détriment du reste de la vidéo.

Ce principe peut se révéler intéressant dans des cas où le consommateur d'un flux vidéo ne serait intéressé que par une zone précise d'une image capturée par une caméra, tout en étant susceptible de changer de zone d'intérêt à tout instant. La zone d'intérêt est choisie et commandée par le client récepteur du flux vidéo, afin que l'émetteur puisse adapter la zone d'intérêt pour la suite de la transmission.

Nous essaierons de faire des choix technologiques les plus simples et compréhensibles possibles et d'assurer une interopérabilité entre différentes plateformes matérielles et logicielles.

1.2 Cas d'application

Afin d'illustrer ce système en pratique, nous l'implémenterons au sein d'une solution de transmission vidéo distante en réalité virtuelle. Une station émettrice mobile capture un flux vidéo provenant d'une caméra à grand angle et transmet un flux qui sera alors reçu sur un autre terminal mobile n'affichant qu'une zone précise de l'image à travers une solution de réalité virtuelle de type Google Cardboard. Le terminal est donc déplacé avec les mouvements de tête du spectateur et la zone de visionnage est donc déplacée en conséquence pour obtenir un effet immersif. La nouvelle région d'intérêt est renvoyée du terminal mobile à l'émetteur pour faire à nouveau coïncider la zone visionnée avec la région d'intérêt à qualité améliorée.

1.3 Contraintes qualitatives

Nous avons la contrainte de transmettre l'ensemble de la vidéo pour pouvoir répondre instantanément aux déplacements de la tête de spectateur. En effet, des solutions existantes comme celle choisie par certains drones tels que celui de Parrot font le choix de ne transmettre que la région d'intérêt. Cela présente l'inconvénient de devoir attendre un aller/retour complet de transmission avant d'avoir l'image demandée. Pour un effet immersif complet, cela ne peut pas être envisagé. D'autres solutions font le choix de simplement transmettre l'image complète non-modifiée, assurant ainsi une image de qualité moyenne sur l'ensemble de la scène.

Notre solution devra donc permettre d'avoir une qualité supérieure de la zone observée au prix d'une dégradation du reste de l'image. Ainsi, lorsque le spectateur bouge la tête, il voit immédiatement la partie de l'image correspondante, mais en qualité dégradée. Il faudra attendre l'aller/retour complet pour obtenir une image de bonne qualité.

Pour pouvoir mesurer une amélioration de la qualité et s'adapter à diverses conditions de transmissions réseau, nous utiliserons un encodeur à débit contraint, de telle sorte que ce soit la qualité qui soit variable. Ce débit pourra d'ailleurs être adapté en fonction des pertes réseaux observées par le récepteur.

2 Recherche

Nous avons sélectionné quelques papiers qui s'intéressent à des techniques s'approchant de notre projet : la compression vidéo avec une qualité variable en fonction de la zone d'intérêt de l'utilisateur.

La solution la plus évidente est d'intervenir directement au niveau de l'algorithme de compression. L'article de Amir Said et William Pearlman?? décrit comment compresser une image avec un débit variable pour chaque partie de l'image.

Cette approche est intéressante mais nous faisons le choix de réutiliser un algorithme de compression, dit *codec*, existant. L'algorithme étant notamment déjà intégré dans les terminaux de réception pour effectuer du décodage matériel, il nous est presque impossible d'intervenir à ce niveau.

L'article de Engin Kurutepe?? montre comment optimiser l'envoi d'un flux provenant de plusieurs caméras. Pour cela, l'équipement récepteur suit les mouvements de la tête de l'utilisateur afin de savoir quelle zone de l'image l'intéresse le plus. Après cela, l'équipement envoie cette information à l'encodeur qui va demander un plus gros débit aux caméras qui filment la zone d'intérêt et un débit plus faible aux autres caméras.

Nous ne pouvons pas réutiliser ce concept car nous n'avons qu'une seule caméra qui filme la scène et le principe décrit par l'article est applicable seulement pour une même scène filmée avec plusieurs caméras. Cela nous a tout de même confortés dans l'idée d'encoder une vidéo sphérique avec une qualité dynamique, mais à l'intérieur d'une même image.

Enfin, l'article de Aditya Mavlankar et Bernd Girod?? explique comment optimiser l'envoi d'une image basse qualité d'un match de football ainsi qu'une zone précise de haute qualité. L'idée ressemble à ce que nous voulons faire, mais les recherches se portent sur l'optimisation de l'encodage pour du broadcast ; comment éviter d'encoder le flux de haute qualité pour chaque téléspectateur. Or, notre projet fonctionne dans un modèle mono-utilisateur, il sera donc plus simple et plus efficace pour notre modèle d'encoder le flux entier en prenant directement en compte la zone d'intérêt de l'utilisateur. Aussi, cet article cite les mécanismes FMO et ASO du standard de compression vidéo H.264 qui permettent respectivement d'avoir un débit différent pour chaque bloc de l'image et de prioriser la qualité des zones de l'image. Malheureusement, ces plugins ne sont pas implémentés par défaut dans des outils libres comme FFmpeg ou la libx264 et peu de récepteurs mobiles ne les comprennent.

Globalement, nous n'avons pas pu trouver d'articles décrivant un procédé permettant d'implémenter directement notre projet, nous partirons donc sur l'étude et la conception d'une solution propre. Cette recherche nous aura tout de même permis d'avoir un bon aperçu des techniques de compression vidéo couramment utilisées.

3 Architecture générale

4 Détails d'implémentation

5 Résultats

6 Conclusion