

الجزائرية الديمقراطية الشعبية الجمهورية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

المدرسة العليا للإعلام الآلي - 08 ماي 1945 - بسيدي بلعباس
Ecole Supérieure en Informatique
-08 Mai 1945- Sidi Bel Abbès



Mémoire de Fin d'étude

En Vue de l'obtention du diplôme de **Master**

Filière : **Informatique**

Spécialité : **Intelligence Artificielle et Sciences de Données (IASD)**

Thème

360° Videos Streaming Optimization and Saliency Prediction

Présenté par :

ABDALLAOUI Walid

Soutenu le : **00/00/2024**

Devant le jury composé de :

- M/Mme/Mlle XXX
- M/Mme/Mlle XXX
- M/Mme/Mlle XXX
- M/Mme/Mlle XXX

- Président
- Encadreur
- Examineur
- Examineur

Année Universitaire : 2023 / 2024

Abstract

In recent years there has been a significant interest in Virtual Reality applications, notably 360° videos. A typical 360° video viewing arrangement involves a user interacting with the scene through a Head Mounted Device (HMD) with the field of view (FoV) which is the region the user sees at any moment.

360° videos provide immersive experiences with potential applications in diverse fields such as education, medicine, and entertainment. However, streaming these videos over a network is bandwidth-intensive, as their file sizes can be up to six times larger than traditional videos. Consequently, this technology faces unique challenges, including high bandwidth requirements and the need for low motion-to-photon latency to prevent user's cybersickness.

To address these challenges, optimization models are crucial. These models rely on understanding the content of 360° videos, particularly through saliency prediction. Saliency prediction identifies the areas within the video that capture users' attention, enabling optimization techniques to prioritize the delivery of these regions.

Several optimization techniques are employed in 360° video streaming. Tile-based streaming divides the video into smaller tiles, allowing for adaptive streaming of only the tiles within the viewer's field of view. Field of View (FoV) prediction anticipates the viewer's gaze direction, enabling proactive fetching of relevant tiles.

In this thesis, we conduct a study of 360° videos streaming technology diving into the streaming framework of this videos at different levels from the content server to the user HMD. We give an overview of the challenges associated with it and the optimization techniques used to overcome them with a focus on the potential visual saliency prediction can offer for such techniques. We conduct an analysis of deep learning based 360° visual saliency prediction models in the literature leading to a better understanding of the problem of human visual attention exploring this immersive content.

Keywords: VR, 360° Videos, HMD, Saliency, FoV, Streaming

Résumé

Ces dernières années ont vu un intérêt croissant pour les applications de Réalité Virtuelle, notamment les vidéos à 360°. Une configuration typique de visionnage de vidéos à 360° implique un utilisateur interagissant avec la scène via un casque de réalité virtuelle (HMD) avec un champ de vision (FoV) qui est la région que l'utilisateur voit à chaque instant.

Les vidéos à 360° offrent des expériences immersives avec des applications potentielles dans divers domaines tels que l'éducation, la médecine et le divertissement. Cependant, la diffusion en continu de ces vidéos sur un réseau est gourmande en bande passante, car leurs tailles de fichier peuvent être jusqu'à six fois supérieures à celles des vidéos traditionnelles. Par conséquent, cette technologie fait face à des défis uniques, notamment des exigences élevées en matière de bande passante et la nécessité d'une faible latence mouvement-photon pour prévenir le mal des transports chez l'utilisateur.

Pour relever ces défis, les modèles d'optimisation sont cruciaux. Ces modèles reposent sur la compréhension du contenu des vidéos à 360°, notamment grâce à la prédiction de saillance. La prédiction de saillance identifie les zones de la vidéo qui captent l'attention des utilisateurs, permettant aux techniques d'optimisation de prioriser la diffusion de ces régions.

Plusieurs techniques d'optimisation sont utilisées dans la diffusion de vidéos à 360°. La diffusion en mosaïque divise la vidéo en tuiles plus petites, permettant une diffusion adaptative uniquement des tuiles dans le champ de vision du spectateur. La prédiction du champ de vision (FoV) anticipe la direction du regard du spectateur, permettant une récupération proactive des tuiles pertinentes.

Dans cette thèse, nous menons une étude sur la technologie de diffusion de vidéos à 360°, en plongeant dans le cadre de diffusion de ces vidéos à différents niveaux, du serveur de contenu au casque HMD de l'utilisateur. Nous donnons un aperçu des défis associés et des techniques d'optimisation utilisées pour les surmonter, en mettant l'accent sur le potentiel que la prédiction de saillance visuelle peut offrir pour de telles techniques. Nous effectuons une analyse des modèles de prédiction de saillance visuelle à 360° basés sur l'apprentissage profond dans la littérature, menant à une meilleure compréhension du problème de l'attention visuelle humaine explorant ce contenu immersif.

Mots-clés: RV, Vidéos 360°, HMD, Saillance, Champ de vision, Streaming