

TIPE - La compression vidéo en temps réel

Introduction

En 2022, la vidéo représenterait environ 82% du trafic IP total sur Internet. Cela inclut le streaming vidéo, les vidéos à la demande, les vidéos en direct, les appels vidéo, et autre forme de contenu visuel. [1] Dans ce contexte, la transmission d'une vidéo non compressée en haute définition, avec un taux de rafraîchissement de 30 images par seconde, nécessiterait un débit de 1,49 Gbps. Cette exigence de bande passante est significativement supérieure à la vitesse moyenne mondiale de connexion, évaluée à 75,4 Mbps. [1] Ainsi, la compression vidéo permet d'ajuster la taille des données tout en préservant une qualité visuelle acceptable, facilitant ainsi le stockage, la diffusion et le partage de contenus vidéo sur l'ensemble du réseau.

L'ensemble des différents algorithmes de compression vidéo, appelés codecs vidéo, reposent tous sur deux principales techniques : la suppression des redondances spatiales et temporelles. Autrement dit, l'objectif est d'exploiter la corrélation entre plusieurs images successives dans le temps, et la corrélation entre les pixels spatialement proches les uns des autres sur la même image. Pour comprendre le fonctionnement d'un codec vidéo, il faut s'intéresser à la division du flux d'images. Tous les codecs vidéo, et essentiellement ceux reposant sur la norme H.264, la plus répandue, divisent ce flux en trois types de trame :

- des images de type I, images de référence, compressées un minimum et ne dépendent d'aucune autre trame
- des images de type P, images prédites et interpolées à partir des images précédents, notamment des images de type I
- des images de type B, images dit bidirectionnelles car interpolées à partir des images précédentes mais également à partir des images futures [3]

Une séquence d'images consécutives de ces trois types permet alors de définir un GOP (ou groupe d'images). Il est généralement constitué en premier d'une image de référence suivie de plusieurs images intermédiaire, il prend la forme I-B-B-P-B-B-P-B-B

Dans la pratique, l'encodeur stocke, ou transmet au décodeur, uniquement les images de référence I et quelques données sur les images P et B qu'il n'est alors pas nécessaire de transmettre complètement.

Ces données, notamment constituées des vecteurs de mouvement des blocs de pixels et des résidus, sont moins volumineuses que les images en elles même mais sont suffisantes pour reconstruire entièrement une image P ou B à partir de l'image de référence, moyennant une certaine perte, convenable, de détails et de qualité.

La construction de ces vecteurs de mouvement se fait par la méthode d'estimation et de compensation de mouvement. Cette technique consiste à détecter le déplacement des éléments d'une image à l'autre. [7] Par exemple, dans le cadre de la diffusion d'un événement sportif, de simples vecteurs décrivant le mouvement des pixels représentant une balle

peuvent suffire à caractériser les images intermédiaires. Elle fait partie de la prédiction inter-images dont le but est de tirer parti de la redondance temporelle en exploitant les similarités entre chaque trame ou parties de trame voisines.

Une autre approche de la compression video consiste à faire usage de la redondance spatiale. Cette compression intra-frame, se concentrant sur une seule image, est très utilisée dans le domaine de la compression d'image. Une technique fondamentale employée est la transformée en cosinus discrète (DCT) qui est au cœur de tous les standards de compression JPEG et MPEG depuis 30 ans. [5] Sachant que l'oeil humain perçoit plus facilement les différences de contraste lorsque la fréquence est faible, les coefficients correspondant aux hautes fréquences ont moins d'importance et ces coefficients peuvent alors être quantifiés de manière moins précise voire ignorés. [5,6] Ces coefficients sont ensuite compressés à l'aide de l'algorithme de Huffman avant d'être transmis. [4]

Bibliographie

- [1] https://www.cisco.com/c/dam/m/en_us/solutions/service-provider/vni-forecast-highlights/pdf/Global_Device_Growth_Traffic_Profiles.pdf (https://www.cisco.com/c/dam/m/en_us/solutions/service-provider/vni-forecast-highlights/pdf/Global_Device_Growth_Traffic_Profiles.pdf)
- [2] https://openaccess.thecvf.com/content_ICCV_2019/papers/Rippel_Learned_Video_Compression_ICCV_2019_paper.pdf (https://openaccess.thecvf.com/content_ICCV_2019/papers/Rippel_Learned_Video_Compression_ICCV_2019_paper.pdf) (p.2)
- [3] https://theses.hal.science/tel-00522618/file/PhD_Olivier_Brouard.pdf (https://theses.hal.science/tel-00522618/file/PhD_Olivier_Brouard.pdf) (GOP p.29, DCT p.31)
- [4] <https://eprints.nottingham.ac.uk/13447/1/thesis.pdf> (<https://eprints.nottingham.ac.uk/13447/1/thesis.pdf>) (Huffman p.25-28)
- [5] <https://interstices.info/de-fourier-a-la-compression-dimages-et-de-videos/> (<https://interstices.info/de-fourier-a-la-compression-dimages-et-de-videos/>)
- [6] <https://www.youtube.com/watch?v=Q2aEzeMDHMA> (<https://www.youtube.com/watch?v=Q2aEzeMDHMA>) (DCT explained on JPG images)
- [7] https://course.ece.cmu.edu/~ee899/project/deepak_mid.htm (https://course.ece.cmu.edu/~ee899/project/deepak_mid.htm) (Algo de block-matching)