

## Sujet de thèse

(pour septembre 2021)

# Codage vidéo : Réduction de la redondance temporelle basée apprentissage profond pour l'optimisation du compromis débit-distorsion

**Directrice de thèse :** Anissa MOKRAOUI (PR)  
Courriel : anissa.mokraoui@univ-paris13.fr

**Co-encadrant :** Fangchen FENG (MCF)  
Courriel : fangchen.feng@univ-paris13.fr

## I. Contexte général

L'étude réalisée dans [1] montre que la transmission vidéo représente plus de 80 % du trafic Internet et devrait d'ailleurs continuer à augmenter dans les années à venir. Il est donc indispensable d'envisager l'amélioration des performances des systèmes conventionnels actuels de codage vidéo. Le codage vidéo est habituellement considéré comme un problème d'optimisation débit-distorsion sous contrainte abordé de deux manières équivalentes à savoir : (i) minimiser le débit binaire tout en garantissant une distorsion minimale à ne pas dépasser ; ou (ii) minimiser la distorsion tout en respectant un débit binaire à ne pas dépasser.

Les codecs vidéo couramment utilisés et/ou plus récents reposent sur les normes H.264, H.265 et VP9. Tous ces codecs s'appuient sur la même architecture hybride basée sur des blocs en cherchant à réduire la redondance à la fois spatiale et temporelle dans le flux vidéo.

L'état de l'art sur le codage vidéo montre que de nombreuses voies restent encore à explorer notamment avec l'émergence de l'intelligence artificielle. En effet pour la réduction de la redondance spatiale, des autoencodageurs (voir par exemple [2]) basés sur des réseaux de neurones profonds (DNN) offrent des performances comparables, voire meilleures, à celles des codecs d'images traditionnels (JPEG, JPEG2000 ou BPG). L'une des explications possibles à ces résultats serait que ces approches mettent tout en œuvre pour apprendre des transformations non-linéaires qui ne sont pas considérées dans les solutions classiques.

## II. Travaux envisagés

Les travaux envisagés dans cette thèse visent à réduire la redondance temporelle dans une vidéo tout en garantissant des gains significatifs comparés aux schémas classiques. On note d'une part que la majorité des standards de codage vidéo (par exemple H264/AVC, HEVC) ont adopté la solution de mise en correspondance de blocs de différentes tailles combinant à la fois des stratégies de prédiction, de codage des vecteurs de mouvement et selon les cas (type P ou B) le codage de l'erreur résiduelle afin de compenser le mouvement. D'autre part, l'état de l'art montre que les architectures actuelles de réseaux de neurones profonds ([3,4,5,6,7]), s'attachent uniquement à générer un flot optique aussi précis que possible (assez souvent carte dense de mouvement) sans tenir compte du coût binaire engendré. Il s'agira donc dans ce projet de thèse d'apprendre au codeur vidéo non seulement à générer mais aussi à coder l'information de mouvement de manière optimale au sens débit-distorsion qui reste encore un problème ouvert ([8]). Dans cette optique, on envisagera de tirer parti à la fois de l'architecture classique de codage vidéo et des méthodes d'apprentissage notamment des réseaux de neurones profonds dédiés à l'estimation du flot optique. L'objectif étant de proposer une architecture pour une compression vidéo de bout en bout qui optimise conjointement (coût unique) toutes les composantes pour une compression optimale du mouvement au sens débit-distorsion.

## III. Bibliographie

- [1] C. V. networking Index. Forecast and methodology, 2016-2021, white paper. San Jose, CA, USA, 1, 2016.
- [2] L. Theis, W. Shi, A. Cunningham, and F. Huszar. Lossy image compression with compressive autoencoders. arXiv preprint arXiv:1703.00395, 2017.
- [3] Junggi Lee, Kyeongbo Kong, GyuJin Bae, and Woo-Jin Song: BlockNet: A Deep Neural Network for Block-Based Motion Estimation Using Representative Matching, *Symmetry* 2020, 12(5), 840; <https://doi.org/10.3390/sym12050840>.
- [4] P. Fischer, A. Dosovitskiy, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. van der Smagt, D. Cremers, and T. Brox, Flownet: Learning optical flow with convolutional networks CoRR, vol. abs/1504.06852, 2015.
- [5] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, and T. Brox, Flownet 2.0: Evolution of optical ow estimation with deep networks CoRR, vol. abs/1612.01925, 2016.
- [6] D. Sun, X. Yang, M. Liu, and J. Kautz, Pwc-net: Cnns for optical ow using pyramid, warping, and cost volume, CoRR, vol. abs/1709.02371, 2017.
- [7] T. Hui, X. Tang, and C. C. Loy, Lite flownet: A lightweight convolutional neural network for optical ow estimation, CoRR, vol. abs/1805.07036, 2018.
- [8] G. Lu , W. Ouyang , D. Xu, X. Zhang, C. Cai1, and Z. Gao, DVC: An End-to-end Deep Video Compression Framework IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.