

PhD opening announcement

Title: Deep learning-based compression of dynamic 3D point clouds

Background, context and objectives of the thesis:

A point cloud (PC) is a set of points in 3D space represented by spatial coordinates (x, y, z) and associated attributes, such as the color and reflectance of each point. Point clouds provide a surface or volumetric representation of objects, as well as free navigation of the scene with six degrees of freedom. Hence, they are an essential data structure in several domains, such as virtual and mixed reality, immersive communication, perception in autonomous vehicles, etc [1]. Since point clouds easily range in the millions of points and can have complex sets of attributes, efficient point cloud compression (PCC) is particularly relevant. The non-regular sampling of point clouds makes difficult to use conventional signal processing and compression tools, which have been traditionally designed to work on regular discrete spaces such as a pixel grid. As a result, compression of point clouds is currently a matter of research and standardization. In particular, the Moving Picture Expert Group (MPEG) has launched a standardization activity for point cloud coding, which has resulted in two recent standards (G-PCC and V-PCC) for geometry-based and video-based point cloud compression, respectively [2]. Specifically, while V-PCC employs a 2D projection principle for coding and relies on off-the-shelf conventional video codecs, G-PCC tackles the problem directly in the 3D space, but uses relatively simple tools such as octrees and hand-crafted models of point dependencies.

Deep point cloud compression (D-PCC) is a recent research avenue exploring the use of deep neural networks for PCC [3]. For lossy geometry coding, voxel-based D-PCC methods have been shown to outperform traditional methods significantly [4,5]. For lossless geometry coding, deep neural networks have been used to improve entropy modeling [6]. Also, D-PCC for attributes has been explored by interpreting point clouds as a 2D discrete manifold in 3D space [7]. Recently, sparse tensor representations have been shown to provide significant advantages in the coding of point clouds [8]. For what concerns dynamic point clouds, the mainstream compression approach is the one of V-PCC, i.e., using 2D projections and conventional video codecs. Nevertheless, recent work has shown that a D-PCC approach in the 3D domain could perform much better [9].

In this thesis, we will study new D-PCC approaches to code dynamic point clouds. Specifically, we will consider the following objectives:

- We will first investigate how to compress dynamic point clouds in the voxel domain, by jointly learning motion estimation and compensation within the coding loop, similarly to what has been done for 2D video [10]. Since the number of points can change from a frame to another, motion estimation needs to be done in a proper feature space, which departs from conventional methods based on regular 2D grids.
- We will explore more general representations for 3D point clouds, in particular spatio-temporal graphs and recent neural representations based on NeRFs and Gaussian Splatting. Compressing point clouds using these representations has been less explored and has the potential to bring significant novel methodological contributions and performance gains.
- Finally, we will also specialize the compression of dynamic point clouds to specific applications, e.g., telepresence applications where 3D human models or avatars are used. In this case, the availability of a prior knowledge of the kind of signal to compress enables the use of domain-specific modeling, with potentially significant coding gains [11].

Supervision, location and funding conditions:

The thesis is a *cotutelle* (co-supervision) between the Université Paris-Saclay, France, and the Ecole de Technologie Supérieure (ETS) of Montreal, Canada, in the context of the International Laboratory on Learning Systems (ILLS) of the CNRS (<https://www.centralesupelec.fr/fr/ills-international-laboratory-learning-systems>).

The PhD candidate will be co-supervised by Giuseppe Valenzise and Pierre Duhamel (Laboratoire des Signaux et Systèmes, Université Paris-Saclay, CNRS, CentraleSupélec), and by Stéphane Coulombe (ETS Montréal). The PhD candidate is expected to spend at least one year at each of the two institutions, and will obtain a double diploma from UPSaclay and ETS upon successful defense of the thesis.

The PhD student will be funded by a Canadian scholarship during the stay at ETS, and from a French MESR scholarship during the stay in France. While the Canadian scholarship is already granted, the candidate will have to *apply* to the STIC doctoral school of UPSaclay and *obtain* the MESR scholarship for the remaining funding (https://www.universite-paris-saclay.fr/sites/default/files/2023-11/adi_flyer_2024_fr_en.pdf).

Profile of the candidates:

We are looking for candidates with strong programming skills, especially in Python. Familiarity with deep learning frameworks (Pytorch, TensorFlow) is highly recommended. The candidate should have a solid background in either mathematical modeling or signal processing. Previous research experience (e.g., one or more research internships) are a plus. Fluency in both oral and written English is essential. Knowledge of French is not required for the application.

Contact information:

If you are interested to apply to this PhD position, please send your CV, transcript of academic records (including bachelor and master) and a letter of motivation to

- Giuseppe Valenzise (giuseppe.valenzise@l2s.centralesupelec.fr),
- Pierre Duhamel (pierre.duhamel@l2s.centralesupelec.fr)
- Stéphane Coulombe (Stephane.Coulombe@etsmtl.ca).

The deadline for applications is 15th March, 2024.

References:

- [1] G. Valenzise, M. Alain, E. Zerman, and C. Ozcinar, "Immersive Video Technologies", Elsevier, 2022.
- [2] C. Cao, M. Preda, V. Zakharchenko, E. S. Jang, and T. Zaharia, "Compression of Sparse and Dense Dynamic Point Clouds—Methods and Standards," Proceedings of the IEEE, pp. 1–22, 2021
- [3] M. Quach, J. Pang, D. Tian, G. Valenzise, and F. Dufaux, "Survey on Deep Learning-based Point Cloud Compression," Frontiers in Signal Processing, vol. 2, 2022. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-03579360>
- [4] M. Quach, G. Valenzise, F. Dufaux. "Learning Convolutional Transforms for Lossy Point Cloud Geometry Compression." IEEE International Conference on Image Processing (ICIP'2019), Sep 2019, Taipei, Taiwan.
- [5] M. Quach, G. Valenzise, F. Dufaux. "Improved Deep Point Cloud Geometry Compression." IEEE International Workshop on Multimedia Signal Processing (MMSp'2020), Sep 2020, Tampere, Finland
- [6] D. T. Nguyen, M. Quach, G. Valenzise, P. Duhamel. "Lossless Coding of Point Cloud Geometry using a Deep Generative Model." IEEE Transactions on Circuits and Systems for Video Technology, Institute of Electrical and Electronics Engineers, 2021, 31 (12), pp.4617 – 4629

- [7] M. Quach, G. Valenzise, F. Dufaux. "Folding-based Compression of Point Cloud Attributes." IEEE International Conference on Image Processing (ICIP'2020), Oct 2020, Abu Dhabi, United Arab Emirates.
- [8] J. Wang, D. Ding, Z. Li and Z. Ma, "Multiscale Point Cloud Geometry Compression," Data Compression Conference (DCC), 2021, pp. 73-82
- [9] T.Fan, L.Gao, Y.Xu, Z.Li, and D.Wang, "D-DPCC:Deep Dynamic Point Cloud Compression via 3D Motion Prediction," in Proceedings of the 31st International Joint Conference on Artificial Intelligence (IJCAI), Jul. 2022.
- [10] G. Lu, W. Ouyang, D. Xu, X. Zhang, C. Cai, and Z. Gao, "DVC: An End-To-End Deep Video Compression Framework," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2019.
- [11] G. Konuko, G. Valenzise, and S. Lathuiliere, "Ultra-low bitrate video conferencing using deep image animation," in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, Toronto, Canada, Jun. 2021