UGCS/FDS11/E13/25

Title of Project:

(English)	Multi-hybrid Video Coding via Strategic Framework with Deep Learning and Super-Resolution Context
(Chinese)	基於深度學習和超分辨率前後呼應的多混合視頻編碼框架策略

Video Coding is to compress videos to the possible smallest amount of size yet maintain its quality for transmission or storage. This is a traditional research area, and constantly researchers feel that the research almost comes to a completion, which is not true. It is due to the great demand of higher and higher video quality and larger and larger amount of videos for transmission. We face great pressure to push the compression to even to a higher ratio. Whist the traditional approaches might have come to the physical limit, what can we do? We suggest a new way using the generative capability of deep learning to advance the coding capability, but our suggestion is not a conventional way. We propose **a new hybrid baseline model** for even better compression efficiency, **making the best use of traditional codecs (any best codec) assisted by the generative power of deep learning for super-resolution imaging**. **The concept appear new.**

Traditional Video Coding (codec): For this we refer it to the conventional MPEG1, MPEG2, H.264(2003), and even the H.265(2012) and H.266(2021). A common idea of these codecs is that they use hybrid video coding model, which has the structure of making use of motion estimation, making up deficiency by residue blocks, using inter- and intra-mode coding, and transform and entropy coding. In an over-simplified simple term, the approach just divides the fames into various block sizes for predictive coding, and makes up the differences between the true and the predicted frames by some residual coding techniques. The hybrid video coding is a very successful approach, and we can optimize the coding size using both conventional and learning approaches.

Coding by Deep learning: This is to code or assist coding videos by making use of CNN, autoencoder network, VAE, transformer structure, etc. The approach is to reduce coding bits by further optimizing the motion vectors, mode selection, block size selectin, etc. with deep learning, or constructing directly an end-to-end codec with a deep learning structure. These reduce the bitrate quite substantially. However, these methods still fall within the optimization of physical size/representation of the fames. Can we do anything further?

In this proposal we suggest to fully make use the generative power of deep learning. To our understanding, there is little or no in-depth study of making generative power of deep learning for video coding. Then, **why is generative power useful?** Since we almost come to the physical limit of code reduction, it is good to make use of them for our initial coding, but for video with larger sizes we utilize the generative deep learning to generate the missing pixels which is a special type of deep learning **super-resolution imaging**. Hence we propose a **Multi-Hybrid Approach for building a Hierarchical baseline structure**: **which involves (i) any best (and suitable) available codec, (ii) a specially designed image Super-resolution unit and (iii) a video smoothing (such as Transformer) network for video production.**

Preliminary testing has been done showing that the idea seems working very well with a bitrate comparable or less than the state-of-the-art codecs and more or less the same quality. The major objectives are to investigate the hierarchical structure of this multi-hybrid baseline model for coding, to allocate the simplest down-sampling filter, to choose the best existing video codec with

good quality for this hybrid baseline, to investigate a fast super-resolution model (suitable to super-resolute fames with high quality reference frames available) and a smoothing transformer (or any suitable DL network) for the production of decoded videos.