Temporal Extrapolation for Zero-Latency Video Transmission

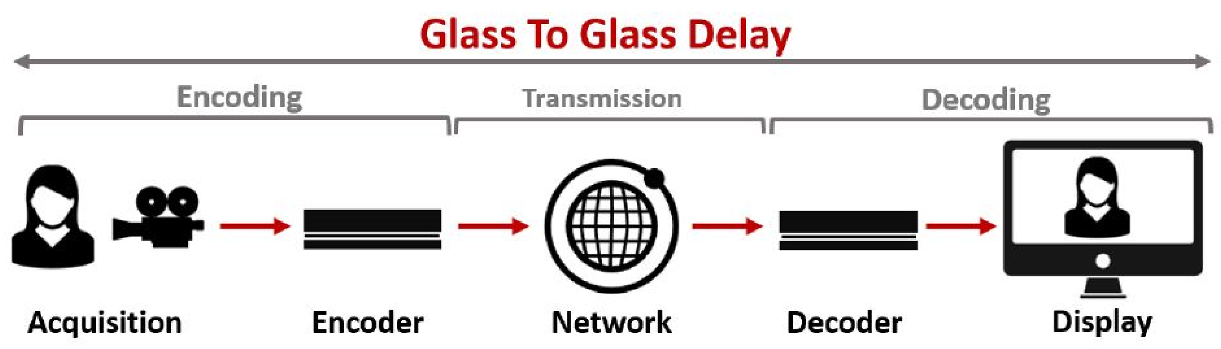
Melan Vijayaratnam[1], Marco Cagnazzo[1], Giuseppe Valenzise[2], Michel Kieffer[2], Anthony Trioux[3]

[1]Telecom Paris, [2]CentraleSupelec, [3]IEMN

*Contact: melan.vijayaratnam@telecom-paris.fr

Introduction

Video-based services involved in applications such as teleoperation of vehicles, telepresence, tactile internet, etc., demand nowadays information exchange with low to ultra-low latency. In this work, we leverage recent progress in the field of video frame extrapolation using deep learning techniques to propose a framework to further reduce latency in video transmission. We will use current extrapolation networks and investigate their usefulness to serve such purpose.



Methodology

We propose to compensate latency using frame extrapolation. This extrapolation can be applied at either the encoder or at the decoder side, as illustrated in Figure 1. For the sake of simplicity, we assume here a low-delay encoder configuration, where the encoding order of frames is identical to the display order. Nevertheless, the proposed scheme generalizes to any GOP structure.

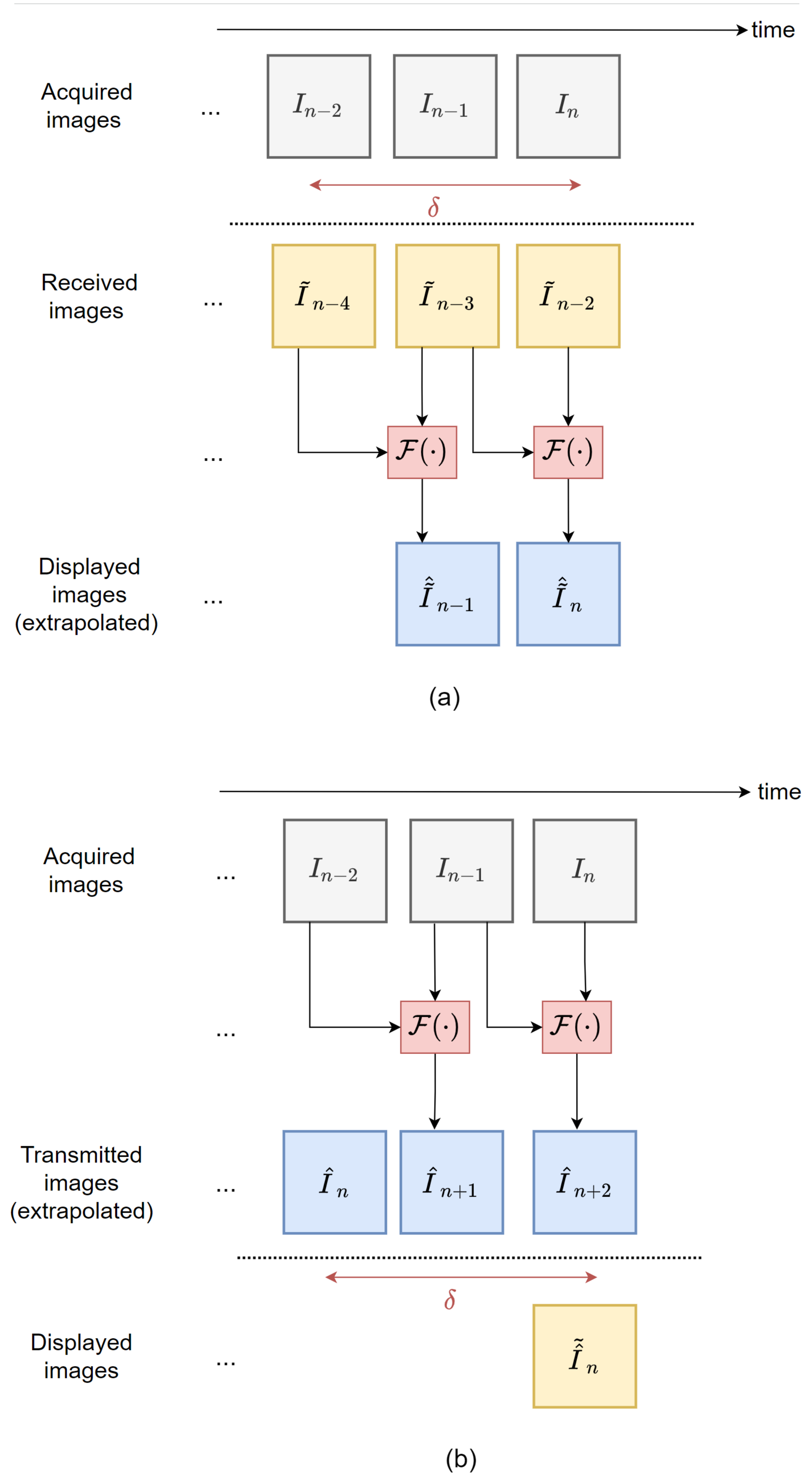


Figura 1. An example of latency compensation with (a) extrapolation at the decoder and (b) extrapolation at the encoder. $\tilde{I}$ and $\hat{I}$ indicate decoded (quantized) and predicted (extrapolated) frames, respectively. $\mathcal{F}$ is the extrapolation function.

Results

Results

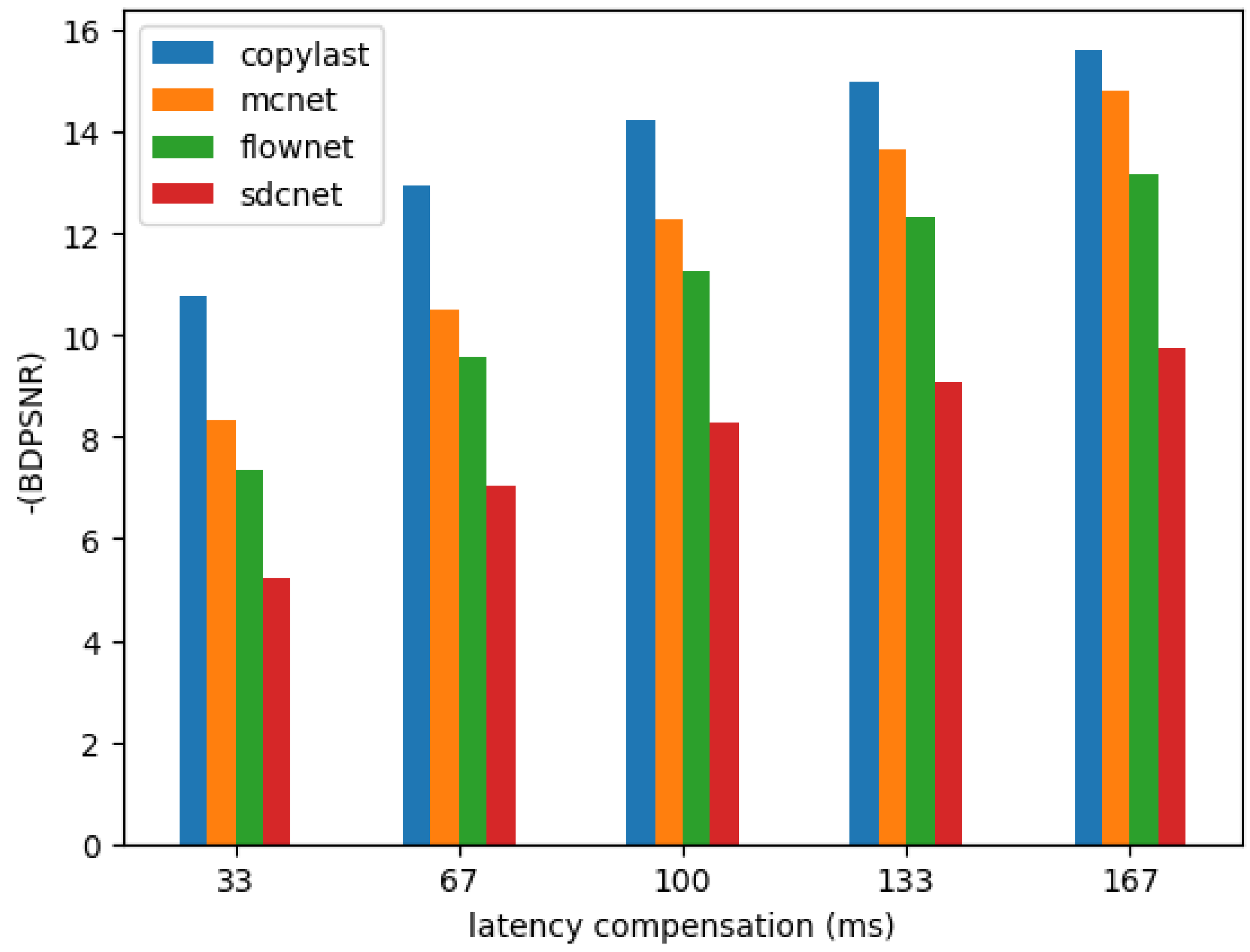


Figura 2. Evolution of the latency compensation-distortion trade-off in terms of PSNR.

Analysis

Figure 2 illustrates visually the latency-distortion trade-off in terms of BDPSNR. As expected, the larger the amount of latency we compensate, the higher the quality loss we incur. For typical G2G latency values of 100 ms observed in remote control applications on the test dataset used in this work, the PSNR loss ranges between 12.3 and 8.3 dB with the considered extrapolation techniques.

Conclusion

This poster introduces a tool that allows us to compensate latency in a video transmission scheme at the price of additional complexity but also of degradation of the frame fidelity. Depending on the application, many configurations are possible extrapolation at the encoder, extrapolation at the decoder, or both at the encoder and decoder. The degradation is essentially caused by the extrapolation approaches. The goal of this paper is not to propose a better extrapolation approach to reduce this loss, but rather demonstrating the applicability of the latency compensation relying on such tool. Future works may concern the improvement of existing extrapolating methods for such task

References

- M. Cordts et al., 2016. *The cityscapes dataset for semantic urban scene understanding*
- F. A. Reda et al., 2018 *dc-net: Video prediction using spatially-displaced convolution*
- E. Ilg et al., 2016 *Flownet 2.0: Evolution of optical flow estimation with deep networks*
- R. Villegas et al., 2017 Decomposing motion and content for natural video sequence prediction