Ultra-low bitrate Video Conferencing Using Deep Animation models

Goluck KONUKO
PhD Student: **Laboratoire des Signaux et Systèmes - CNRS - Centrale Supélec - Université Paris Saclay**

Context and Motivation

01

Ultra-low bitrate compression for real-time chat applications.

Coding Frameworks

02

Deep animation modelling with sparse motion representation

Results

03

RD and Subjective comparison versus conventional frameworks.

Conclusions

04

Performance gains, limitations and ongoing work.

Introduction

- My Thesis - **Deep generative compression for Low Bitrate Video Conferencing**
  - Applied deep learning in video compression
  - Target low-latency, low bitrate application
  - Primarily modelling human face and associated coding scene elements

Context and Motivation

Model-based video compression

Apply deep neural networks as the building blocks of a compression framework.

Low-bitrate communication

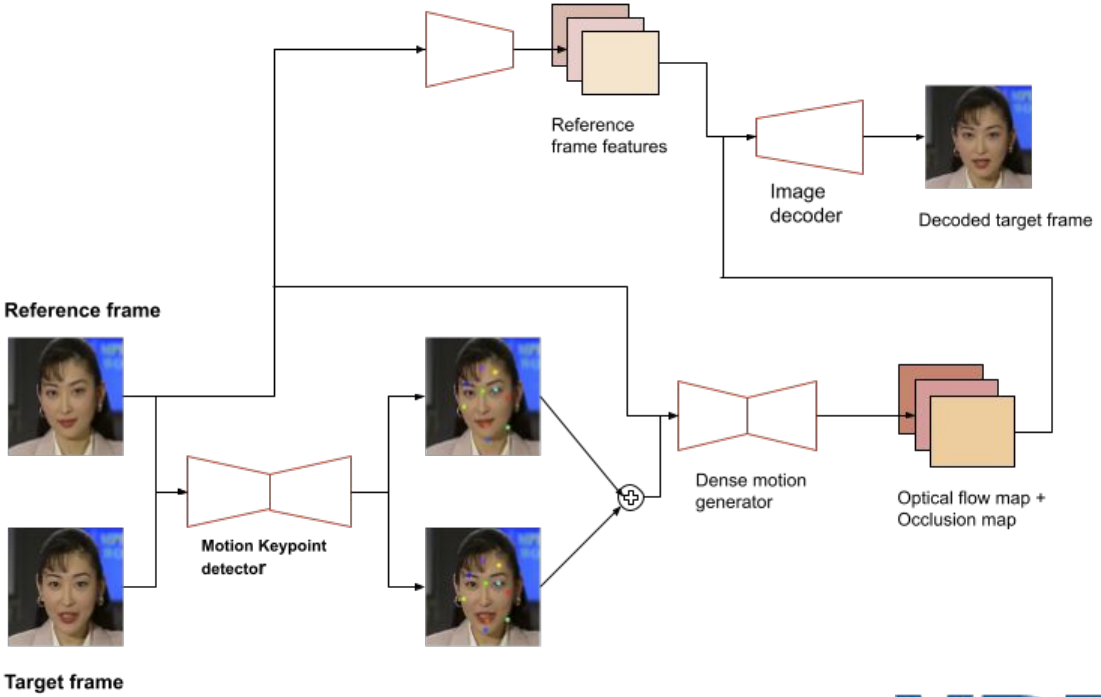
Use sparsely coded information for video reconstruction.

High reconstruction quality

Use generative capabilities of GANs to achieve high perceptual quality.

Framework: DAC Components

- 3 neural networks jointly trained with a linear combination of **feature matching, equivariance, GAN and perceptual losses.**
- Motion keypoint detector
  - Pixel coordinate, R^2
  - Local affine transform ($R^{2\times2}$)
- Dense motion generator: Predicts a **dense optical flow map** representing motion between the reference and target frames.

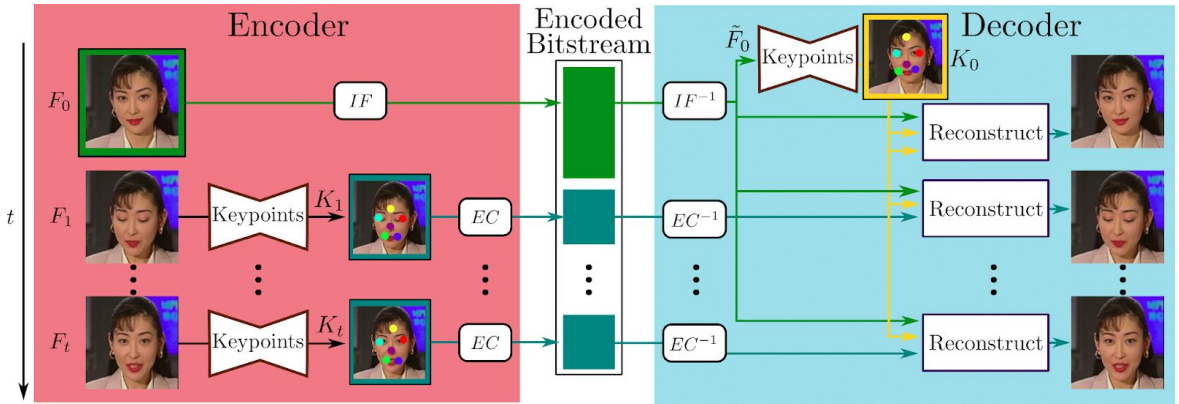


Framework: Deep Animation Coding

Motion representation with sparse keypoints

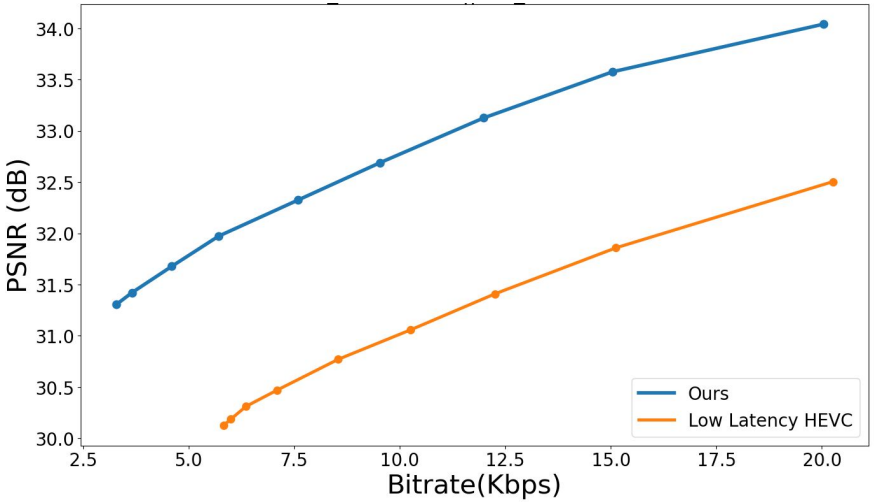
Dense optical flow prediction with a deep neural network.

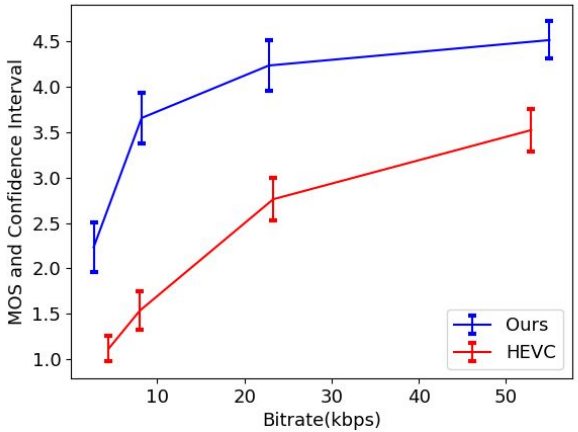
Apply Adversarially trained GAN to reconstruct P frames



Visual Comparison

Results





	VoxCeleb BD quality / BD rate	Xiph.org BD quality / BD rate
PSNR	1.66 / -62.18	2.23 / -76.86
SSIM	0.06/ -56.34	0.1/ -67.43
MS-SSIM	0.04 / -52.33	0.06 / -51.73
VIF	0.02 / -41.46	0.05 / -49.99

Table 1: User preference score vs HEVC

Bitrate (kbps)	No. of votes (Ours/HEVC)	PREFERENCE (%)
5	268 / 0	**100.00**
10	229 / 31	88.08
15	218 / 40	84.50
20	193 / 59	76.59
25	160 / 92	63.49
30	164 / 98	62.60

Some Issues

Quality Scalability

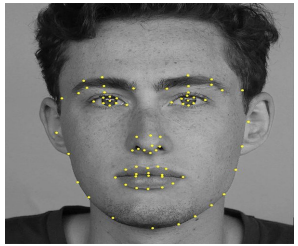
The animation model converges to a single RD point

Keypoint Quantization

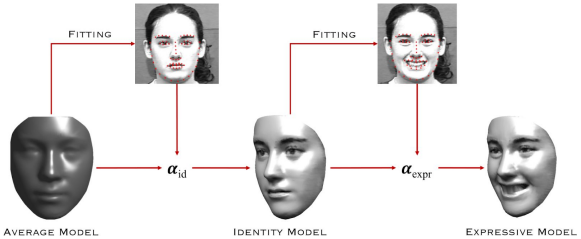
Quantization of keypoints degrades optical flow prediction at test time.

Supervised vs Unsupervised Keypoints

Supervised KPs



3DMMs + Supervised KPs



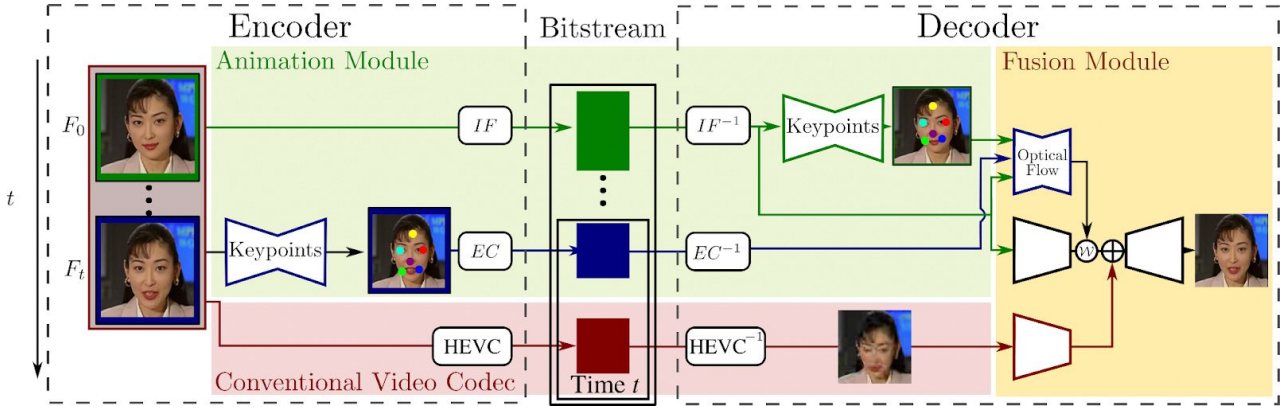
Unsupervised KPs

Framework: Hybrid Coding

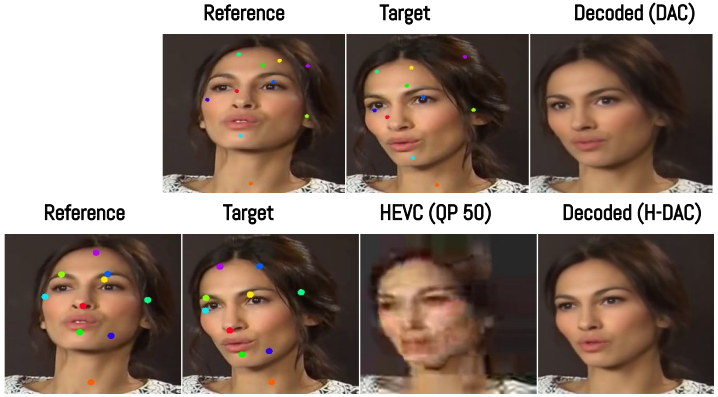
Two-stream hybrid coding for quality scalability.

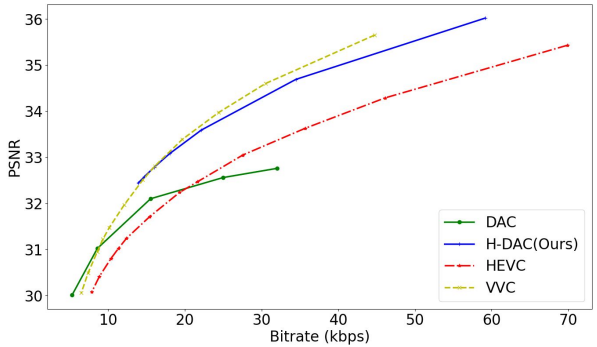
Applying quantization-aware motion keypoints.

A novel fusion module with multi-scale adaptive instance normalization.



Results





DAC(~5kbps)

H-DAC (~9kbps)

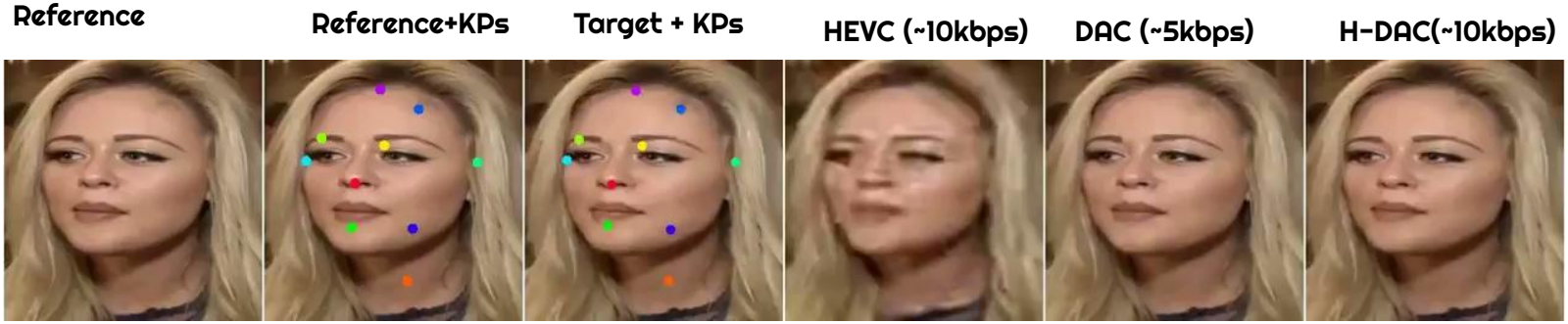
	VoxCeleb BD quality / BD rate	Xiph.org BD quality / BD rate
PSNR↑	1.07 / -33.36	0.97 / -30.7
SSIM↑	0.02/ -33.41	0.02 / -28.33
msVGG ↓	-19.16 / -48.84	-20.04 / -41.64

Visual Comparison

https://drive.google.com/file/d/1kpQ8yn1GVRGp0vNMyLC6mc-mmOssCgPw/view?usp=sharing

https://drive.google.com/file/d/1kpQ8yn1GVRGp0vNMyLC6mc-mmOssCgPw/view?usp=sharing

Visual Comparison



https://drive.google.com/file/d/1aBXLJrJBstSiC_BvaQIQ3nPOI1huIOZ5/view?usp=sharing

https://drive.google.com/file/d/1SI2jUPnerCyavqJZn8zuBegePaXFb7hp/view?usp=sharing

Conclusions

Summary

- We propose the first video conferencing codec that employs a deep generative frame animation scheme with ultra-low bitrate performance.

Ongoing Work

- Ongoing work focuses on quality and resolution scalability as well as model pruning for test on mobile devices.

References

- *G. Konuko, G. Valenzise, and S. Lathuili`ere, “Ultra-low bitrate video conferencing using deep image animation,” in IEEE ICASSP, 2020*
- *Konuko et al. “A hybrid deep animation codec for low-bitrate video conferencing”, preprint. ICIP 2022*
- *A. Siarohin, S. Lathuili`ere, S. Tulyakov, E. Ricci, and N. Sebe, “First order motion model for image animation,” in Neurips, 2019.*