

# **AHG 16: Proposed Common Software Tools and Testing Conditions for Generative Face Video Compression**

**Bolin Chen, Jie Chen, Ru-Ling Liao, Yan Ye and Shiqi Wang**

**Alibaba Group and City University of Hong Kong**

**Jan. 2024**

# Proposal Review

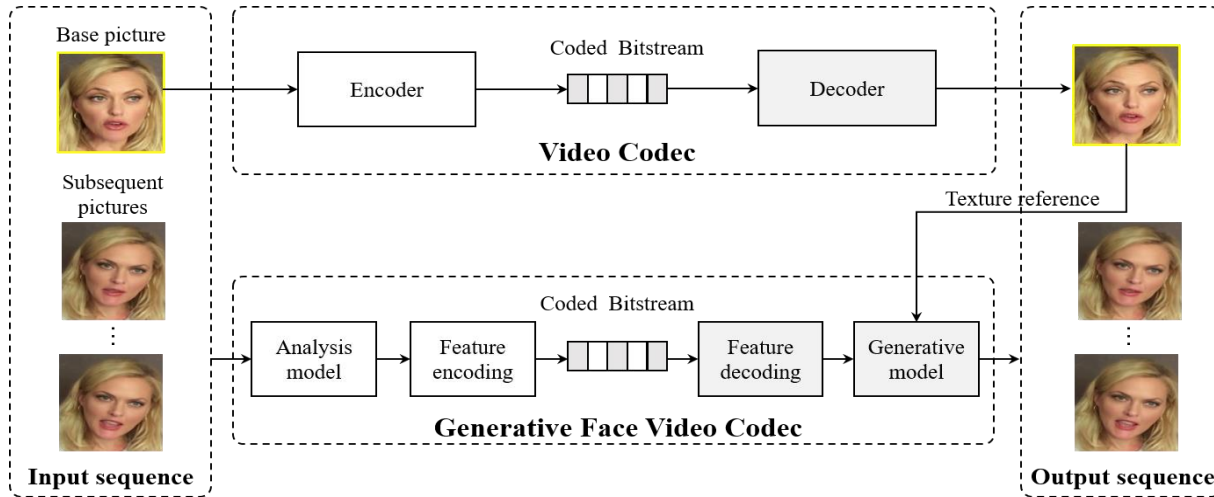


Figure: General block diagram of generative compression codec.

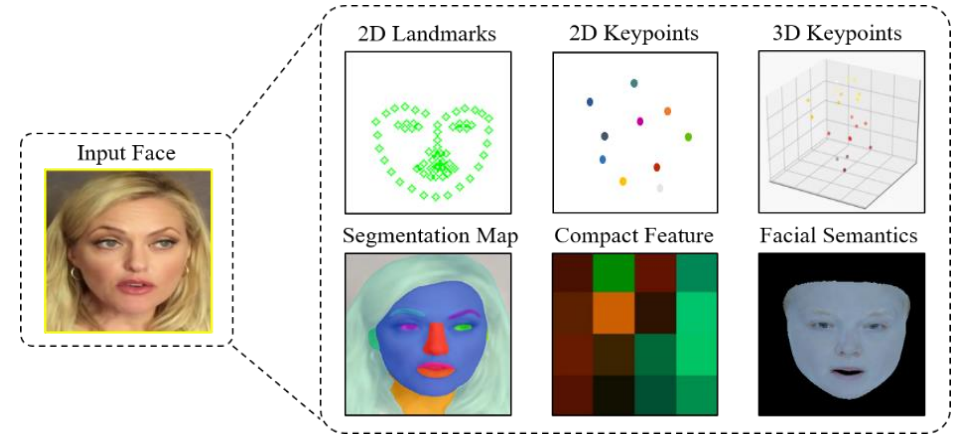


Figure: Examples of facial representations used in generative face video compression.

## ➤ 1<sup>st</sup> Conf Call on 7/Dec/2023

- ❑ We propose **a unified software package** within different face video representations.
- ❑ We propose **test data with suitable copyright license** and corresponding **test conditions** for experimentation.
- ❑ We **provide rate-distortion performance** based on the software and CTC.

# Proposal Update

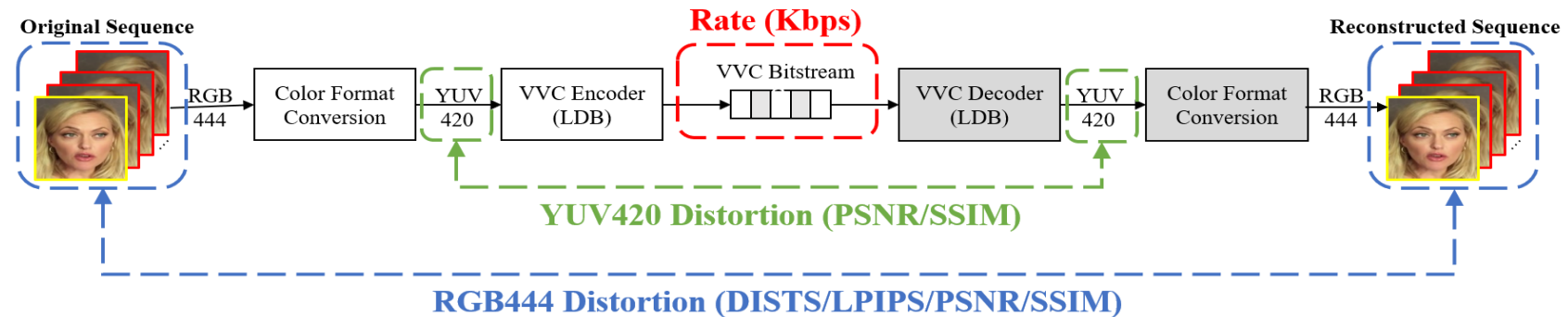
---

- ❑ The distortion measurement has been amended according to the discussion at the AHG conference call by **adding PSNR and SSIM calculation in the YUV420 domain**.
  - The github repo has been updated with the latest metric calculation functionality.
  - Rate distortion performance in terms of PSNR and SSIM (in RGB 444 and in YUV 420) has been added to this contribution.
  
- ❑ The xls attachment has also been revised to show **per-sequence rate-distortion performance and rate-distortion plots**.
  
- ❑ Extra **Bad cases** are also tested and provided except the currently proposed testing data.

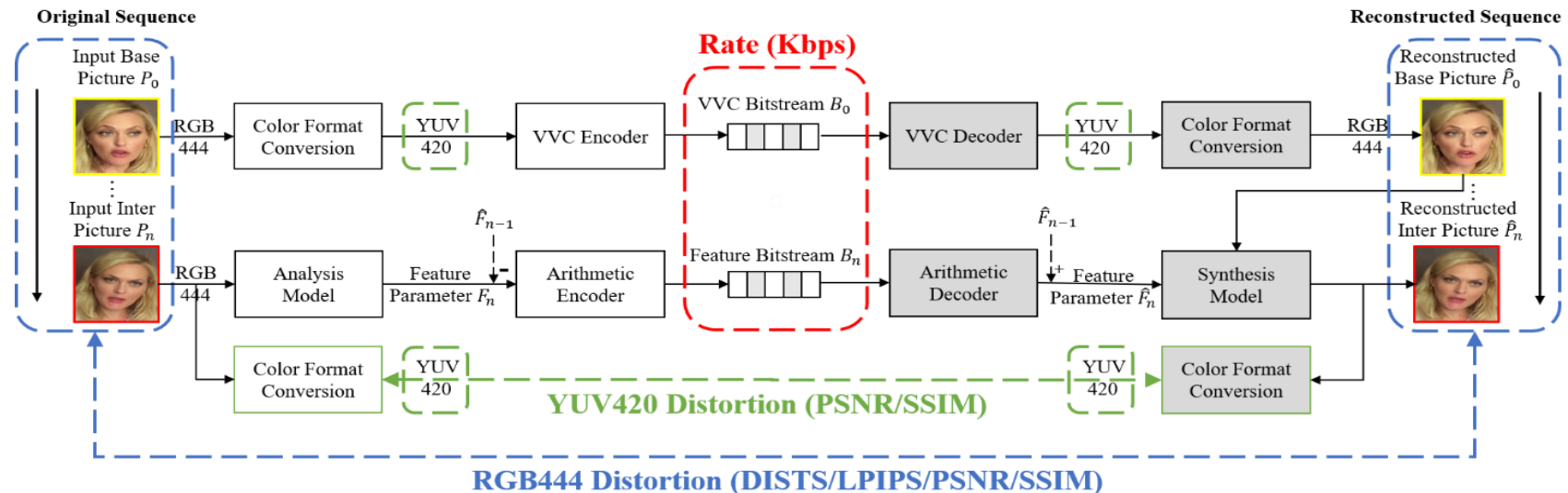
# Updated Testing Pipelines

- As shown in below Figures, coding is performed in **YUV 420 domain** to conform to **Main 10 Profile**, whereas distortion is measured in both **RGB 444 domain** and **YUV 420 domain**.

➤ VVC



➤ GFVC



**Table - Average bit-rate savings of 30 face video sequences for different classes compared with the traditional VVC codec**

Dataset	FOMM		FV2V		CFTE	
	Rate-DISTS	Rate-LPIPS	Rate-DISTS	Rate-LPIPS	Rate-DISTS	Rate-LPIPS
Class A	-35.56%	-30.99%	-51.86%	-50.71%	-64.76%	-63.66%
Class B	-32.51%	-29.37%	-43.18%	-40.05%	-55.73%	-52.52%
Average	-34.04%	-30.18%	-47.52%	-45.38%	-60.24%	-58.09%

- In comparison with the latest VVC codec, the **GFVC** codec (i.e., FOMM, FV2V and CFTE) can achieve **superior BD-rate savings** in terms of Rate-DISTS and Rate-LPIPS.

**Table - Average bit-rate savings of 30 face video sequences for different classes compared with the traditional VVC codec in RGB 444 domain**

Dateset	FOMM		FV2V		CFTE	
	Rate-PSNR	Rate-SSIM	Rate-PSNR	Rate-SSIM	Rate-PSNR	Rate-SSIM
Class A	#VALUE!	-10.60%	3.00%	-21.00%	-5.35%	-28.10%
Class B	16.90%	-0.80%	0.81%	-12.10%	-3.31%	-23.30%
Average	#VALUE!	-5.70%	1.91%	-16.60%	-4.33%	-25.70%

- ❑ For RGB 444, there are gains even for traditional metrics.
- ❑ “#VALUE!” means the curve overlap for some sequences so that the BD rate cannot be calculated.

**Table - Average bit-rate savings of 30 face video sequences for different classes compared with the traditional VVC codec in YUV 420 domain**

Dateset	FOMM		FV2V		CFTE	
	Rate-PSNR	Rate-SSIM	Rate-PSNR	Rate-SSIM	Rate-PSNR	Rate-SSIM
Class A	#VALUE!	1.70%	0.60%	-20.00%	-2.20%	#VALUE!
Class B	-0.10%	1.90%	-0.90%	-9.60%	-3.10%	#VALUE!
Average	#VALUE!	1.80%	-0.20%	-14.80%	-2.70%	#VALUE!

- ❑ For YUV 420, due to additional color space conversion, some loss in terms of traditional metrics can be observed compared with RGB 444.



# Extra Testing Data Introduction



- ❑ These extra sequences have **more complex head motion and camera motion**.
- ❑ They are also sourced from CFVQA dataset (Class B), but not included in the currently proposed test data.



# Extra Bad Cases

Original



VVC @ 4.28k



FOMM @ 3.92k



Distorted  
Face



Writhing  
BG



Occuled  
BG



FV2V @ 4.96k

CFTE @ 4.18k

# Extra Bad Cases

Original



VVC @ 6.25k



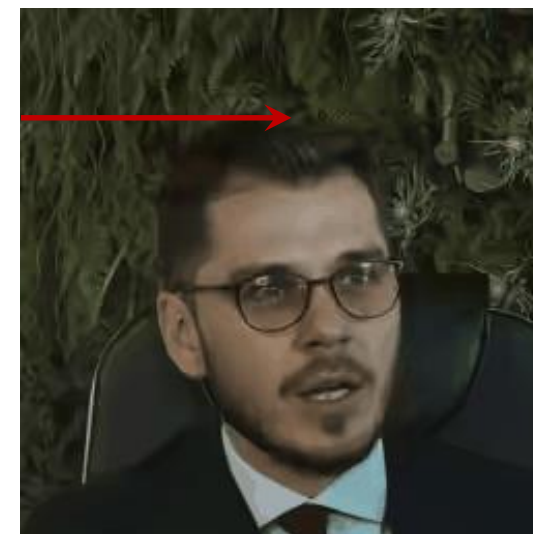
Distorted  
Mouth/Nose

FOMM @ 7.19k



FV2V @ 6.67k

Lack of  
Reference



CFTE @ 6.10k

Writhing  
BG

# Extra Bad Cases

Original



VVC @ 4.79k



FOMM @ 4.09k



Dislocated  
Face &  
Shoulder



Distorted  
Face



Writhing  
BG



FV2V @ 3.90k

CFTE @ 4.72k



Q & A