Generative Face Video Compression: Promises and Challenges

Yan Ye

Alibaba Cloud Intelligence, Alibaba Group

Outline

Introduction

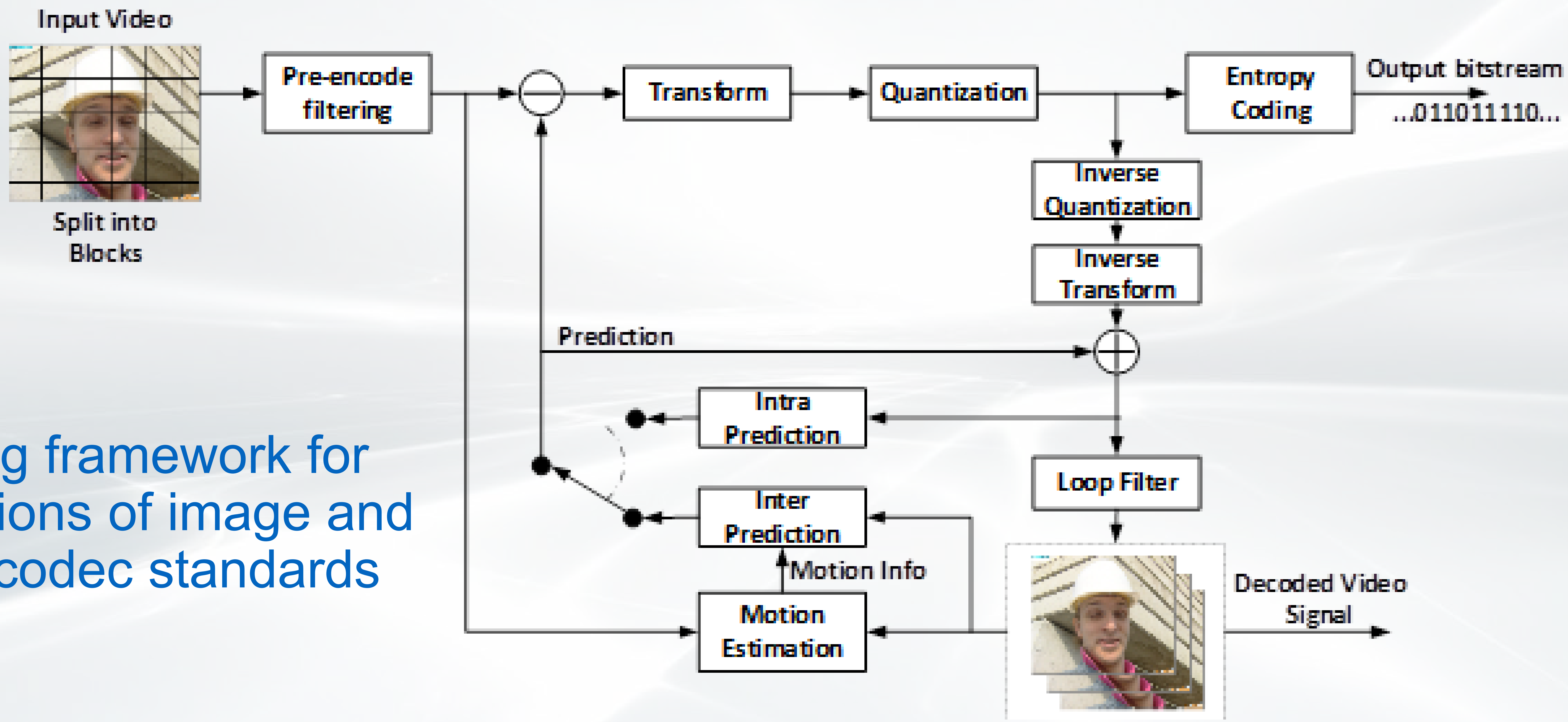
Part 1: the promise

Part 2: the challenge

Concluding remarks

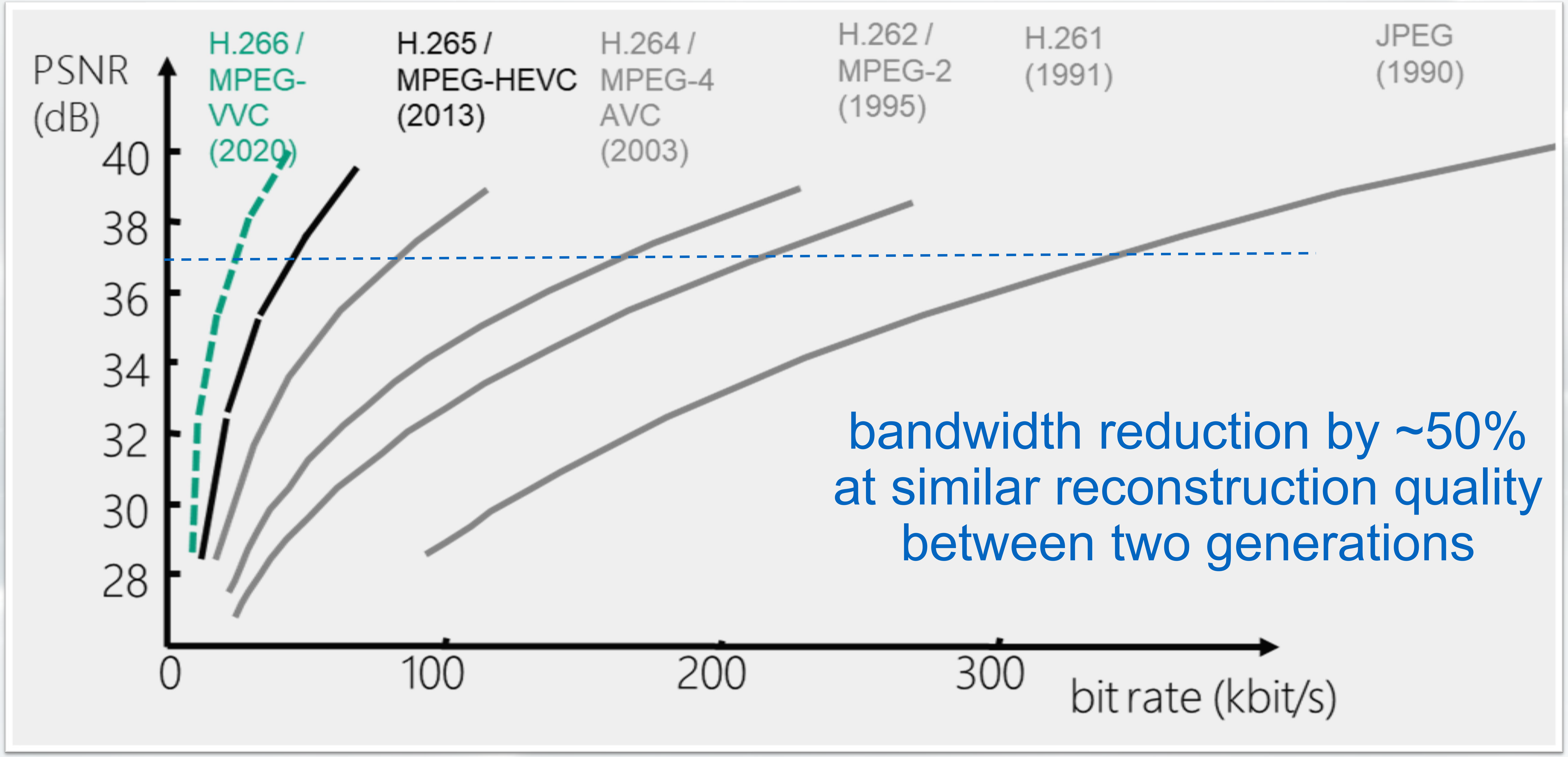
INTRODUCTION

Block-based hybrid video coding



Coding framework for generations of image and video codec standards

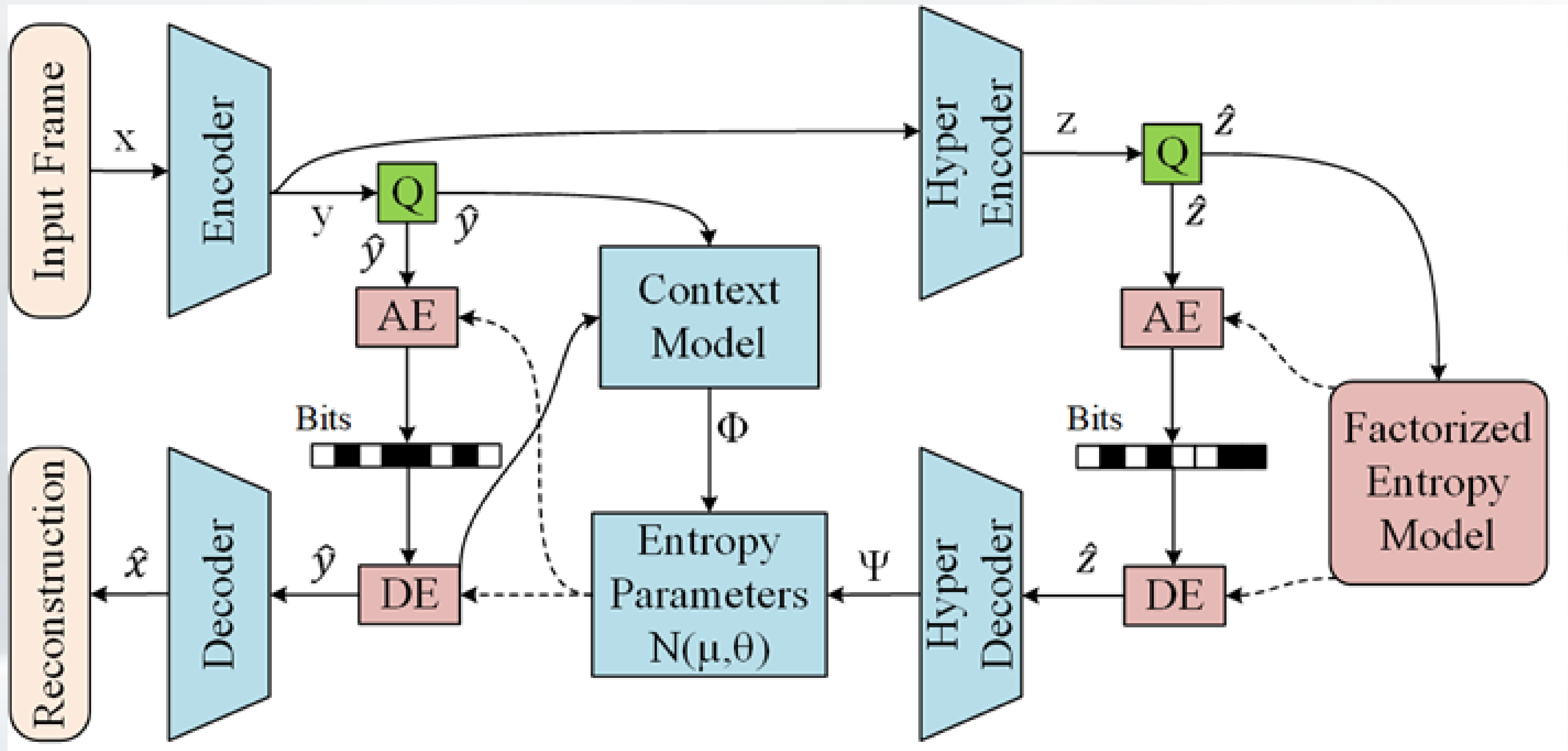
Evolution of compression efficiency

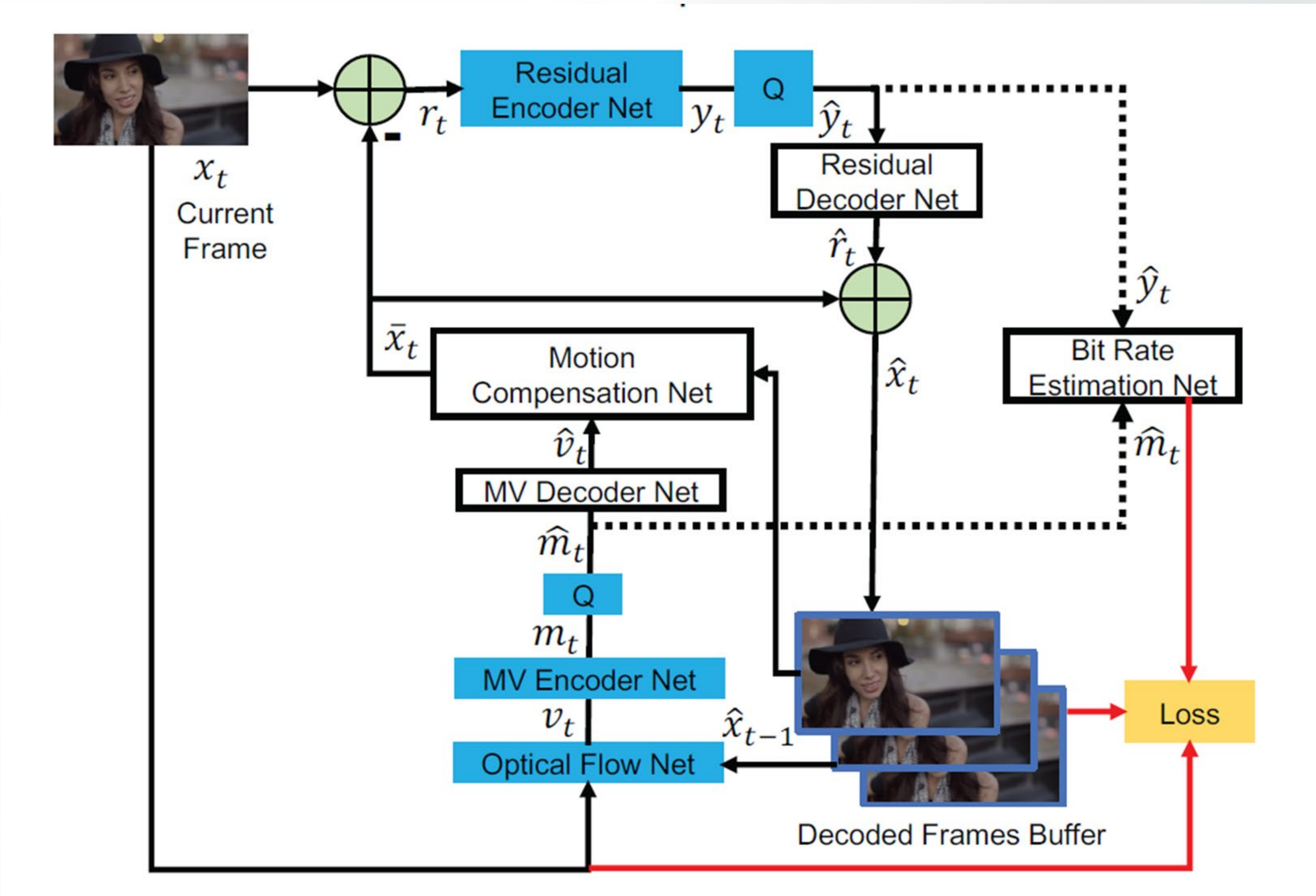


Slide courtesy of B. Bross, “Versatile video coding (VVC) on the final stretch”, ITU Workshop on “The future of media,” Geneva, Switzerland, 8 October 2019

AI-based image and video coding

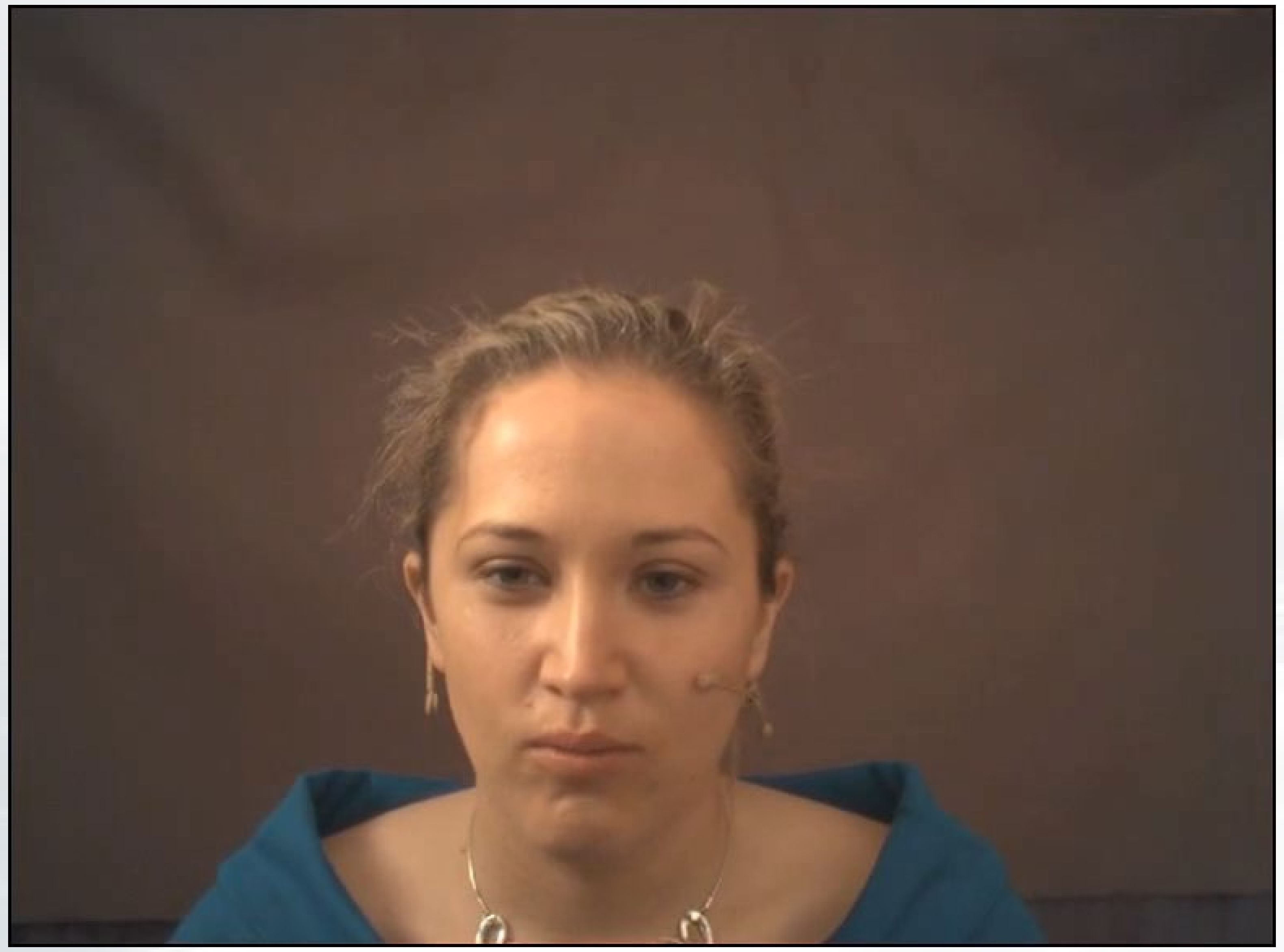
- Enhancing/replacing a coding tool within the hybrid framework
  - ✓ Intra coding, inter coding, loop filtering, etc.
- **End-to-end learning-based image and video compression**





D. Minnen, J. Ballé, and G. Toderici. "Joint autoregressive and hierarchical priors for learned image compression." In Advances in Neural Information Processing Systems, pages 10771–10780, 2018.
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, pp. 11006–11015, 2019.

Face video compression for video chat

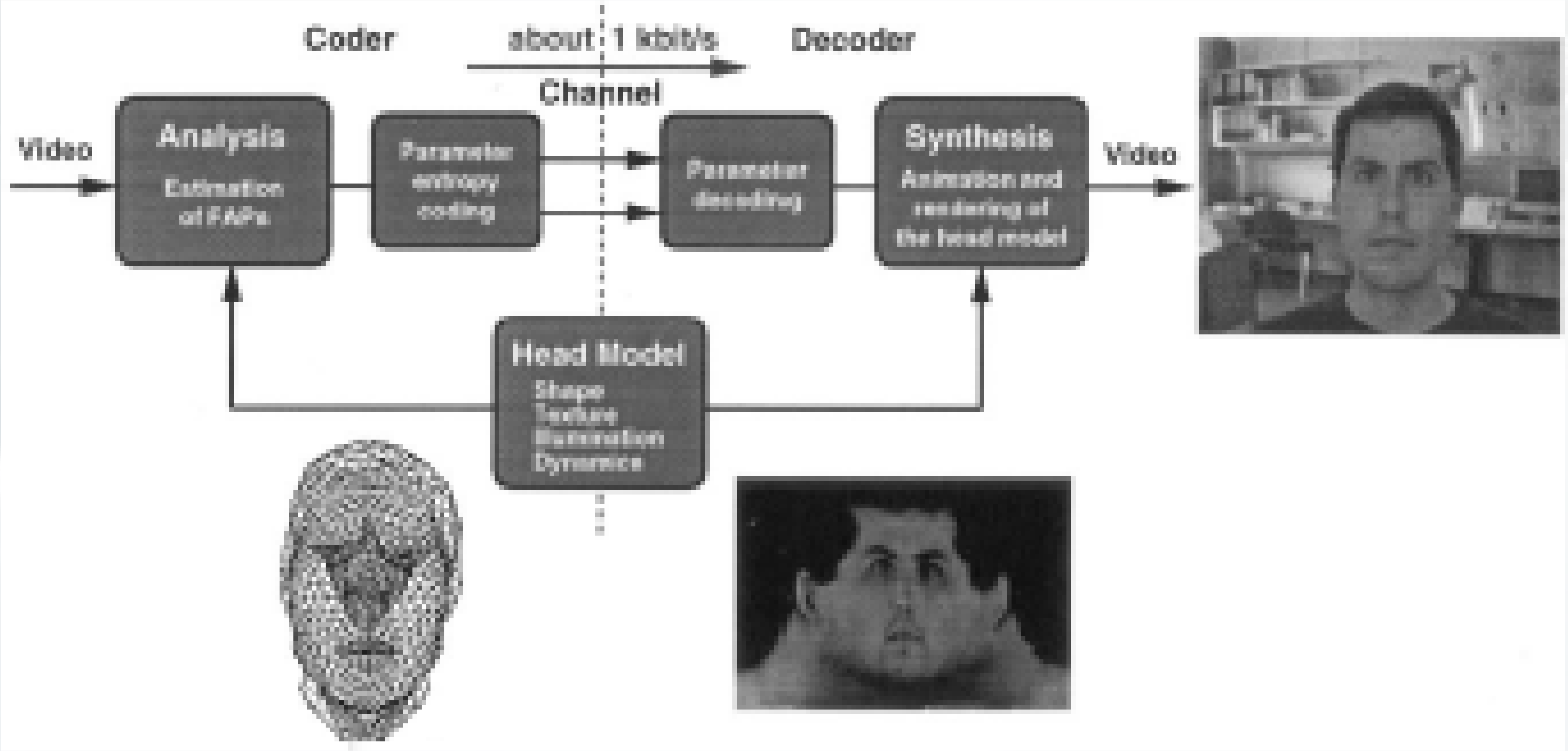


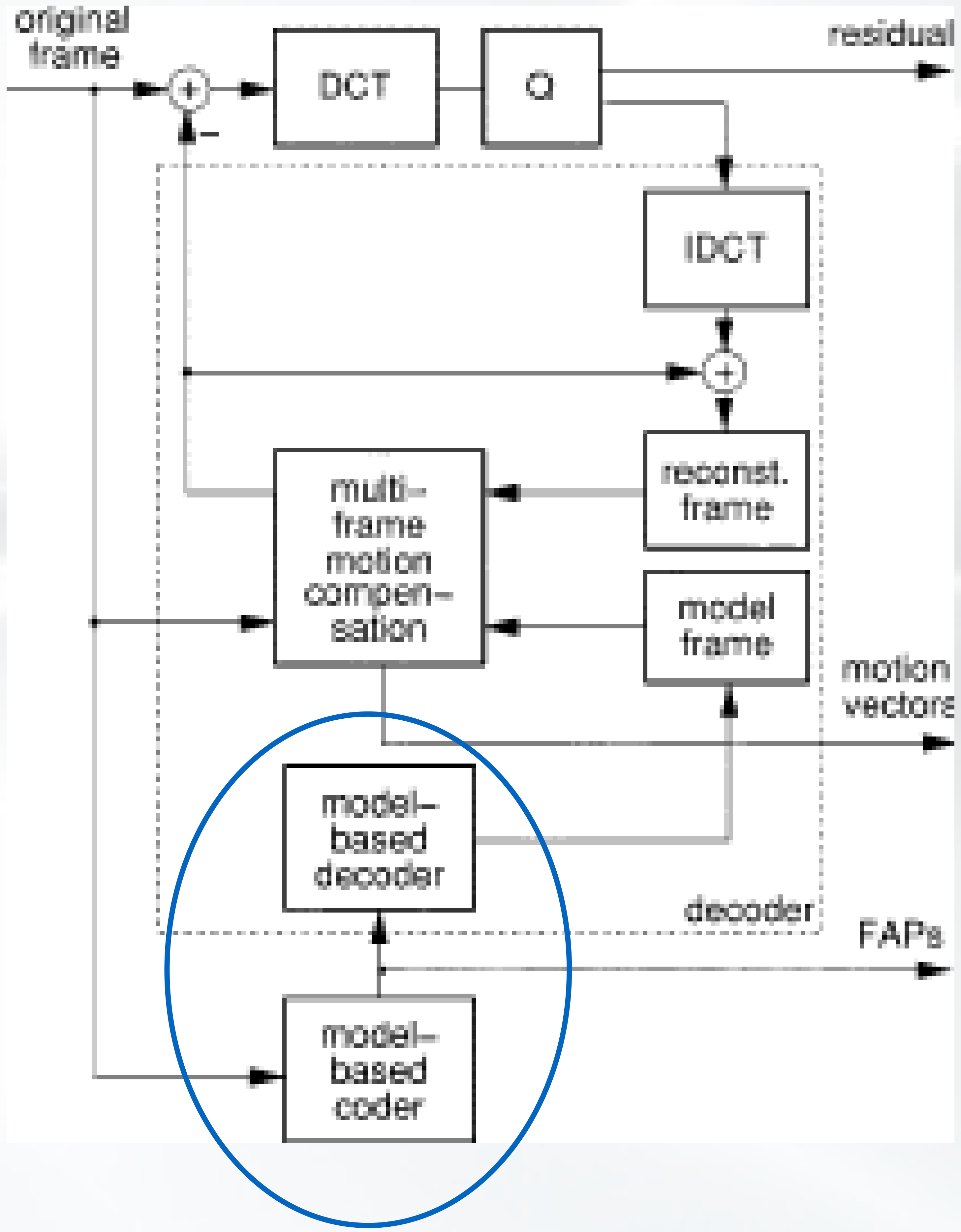
We focus on coding of human face video, where we find much inherent structure and prior knowledge, such as their shape, composition, and movement

Model-based video compression

In the 1990's, model-based video compression was studied for video telephony:

- parameterized 3-D head model specifies shape and color of a person
- facial animation parameters (FAP's) specifies motion and deformation in the temporal domain





P. Eisert, T. Wiegand, and B. Girod, "Model-aided coding: a new approach to incorporate facial animation into motion-compensated video coding," IEEE Transactions on CSVT, vol. 10, no. 3, pp. 344–358, 2000

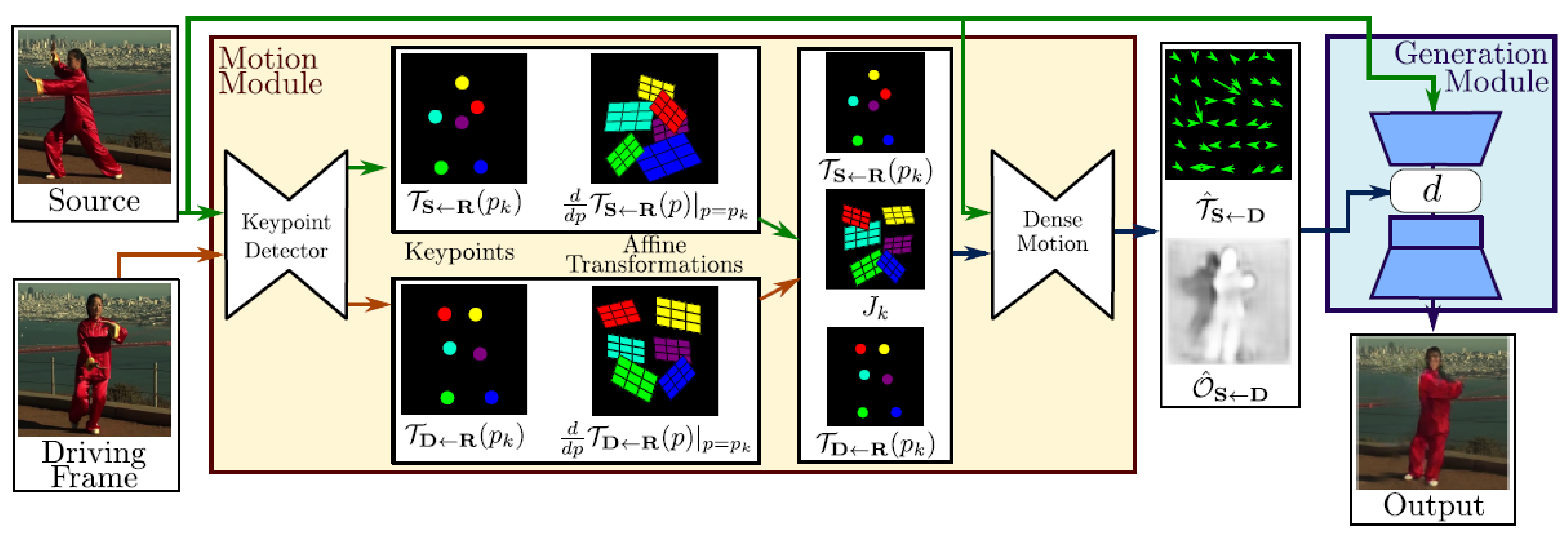
Alibaba Cloud
Worldwide Cloud Services Partner

CityU
香港城市大學
City University of Hong Kong
專業 創新 胸懷全球
Professional · Creative
For The World

PART 1:
THE PROMISE

Related work

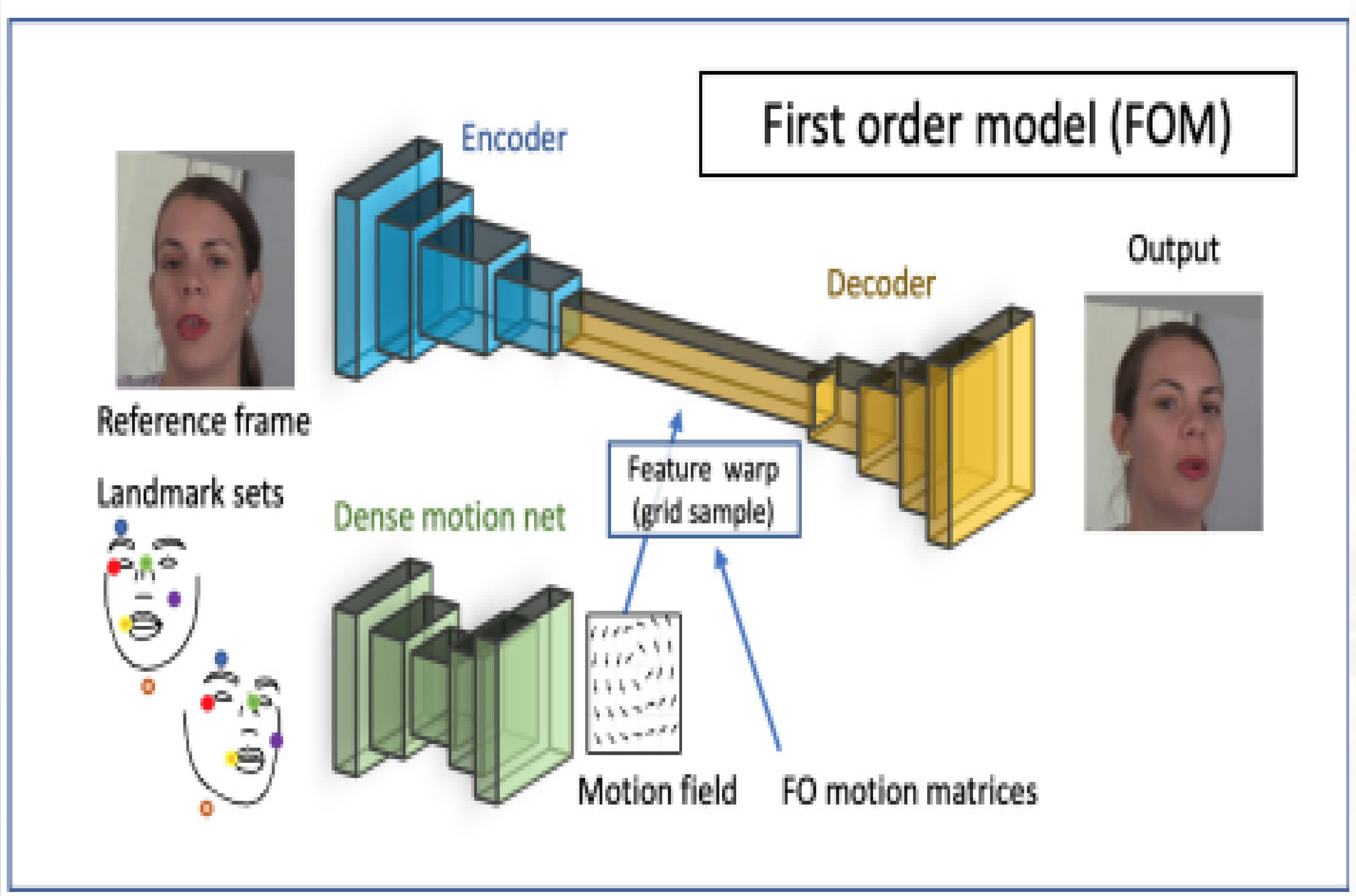
First order motion model (FOMM)

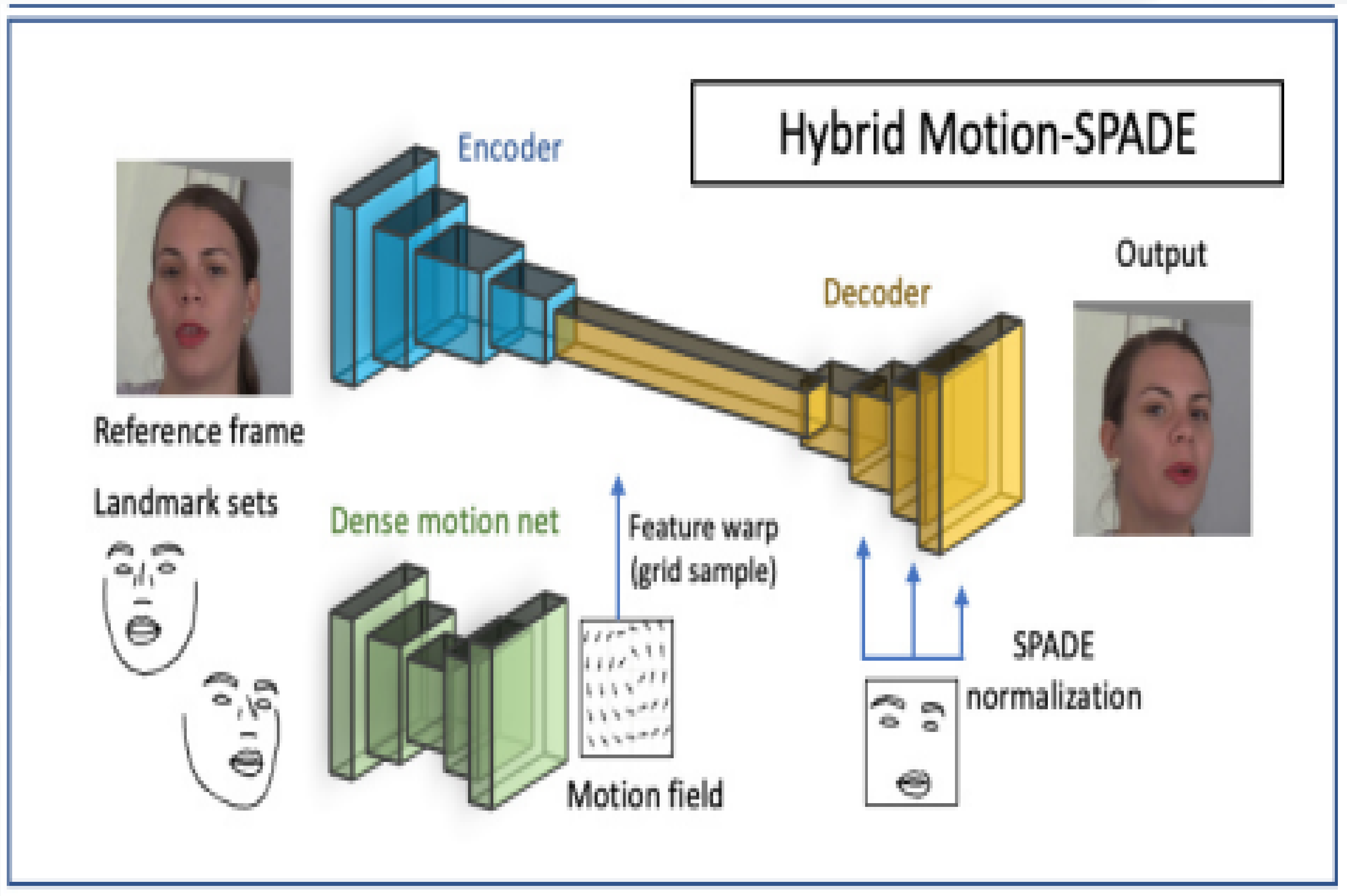


- *Complex motions are represented using a set of keypoints & corresponding affine transformations*
- *Generator network combines the source image and the motion derived from the driving video*
- *Object in the source image is animated according to the motion of driving video*

A. Siarohin, et. al., "First order motion model for image animation," Advances in Neural Information Processing Systems, vol. 32, pp. 7137–7147, 2019.

Low bandwidth video-chat compression

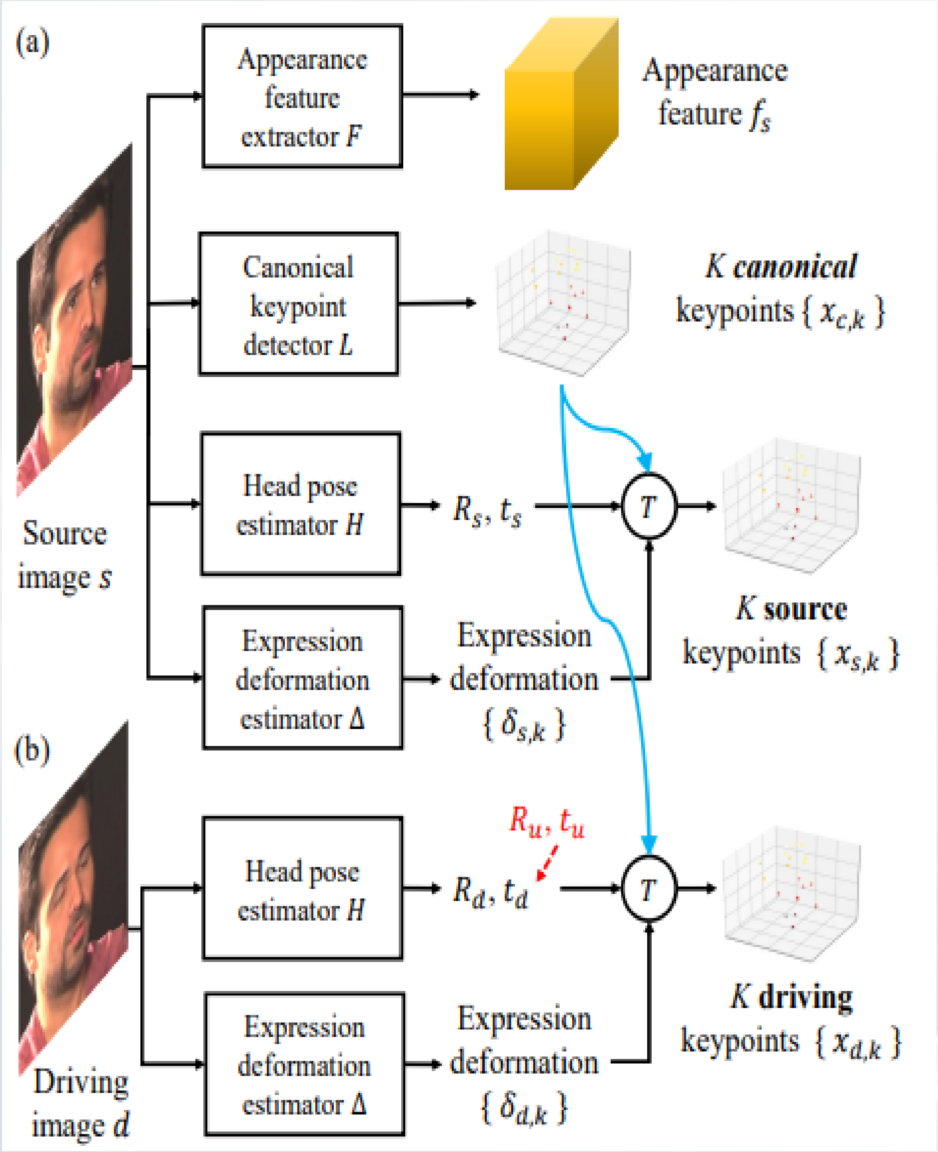


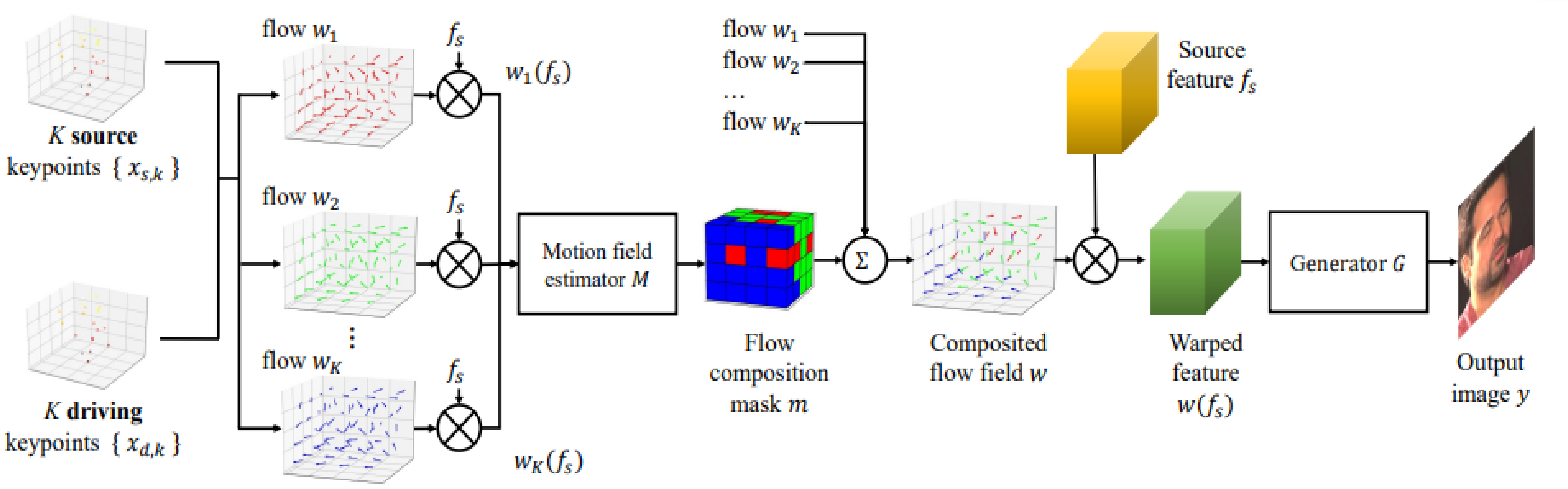


- *Apply FOMM towards talking-head video compression*
- *Explore quality and bandwidth trade-offs for static landmarks (i.e., keypoints), dynamic landmarks or segmentation maps*
- *Runs real-time on mobile platform*

M. Oquab, et. al., "Low bandwidth video-chat compression using deep generative models," in IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshop, 2021.

Free-view neural talking-head synthesis



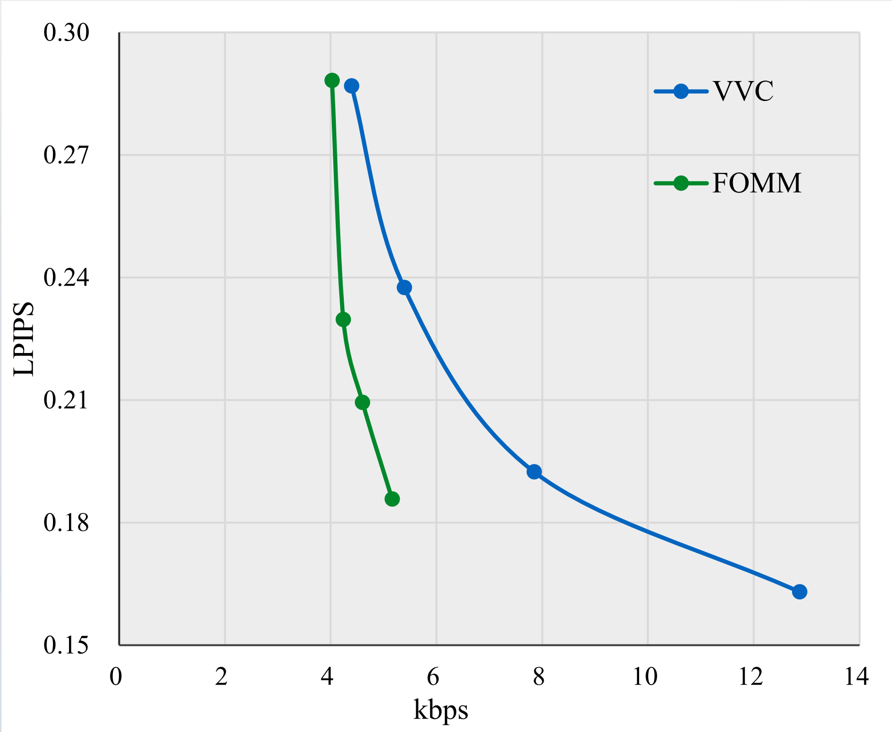


- *Motion information represented using compact 3D keypoints*
- *Source image containing the target person's appearance and driving video dictates the motion in the output*
- *3D keypoints allows to rotate the head during synthesis*

T.C. Wang, et. al., "One-shot free-view neural talking-head synthesis for video conferencing," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 10039–10049.

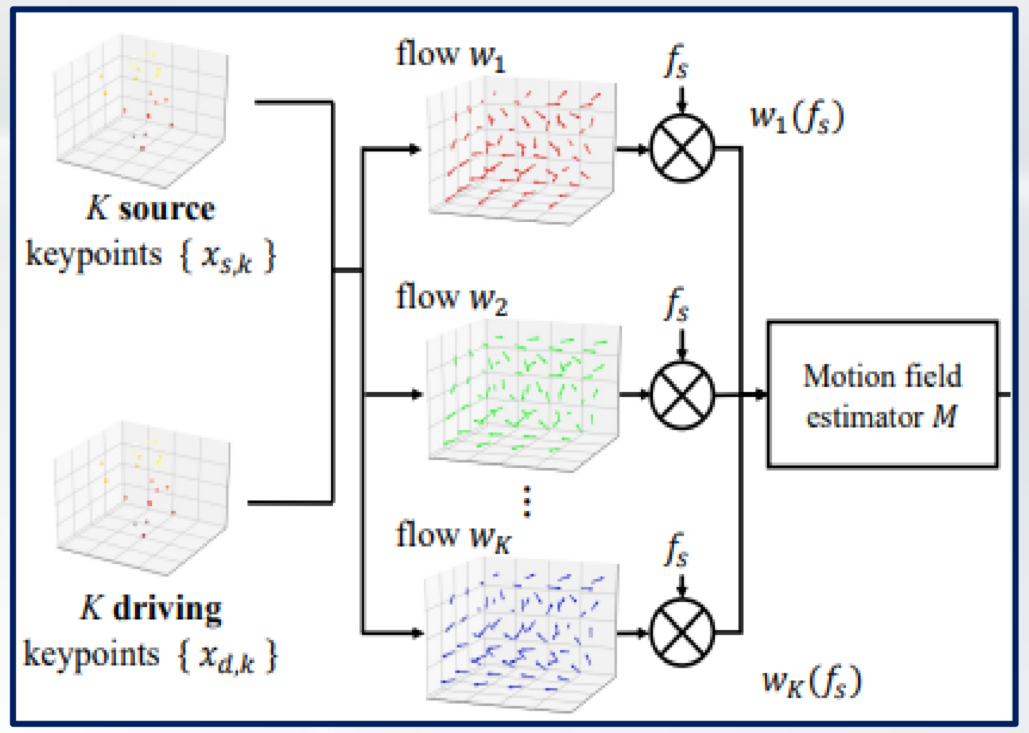
Going beyond keypoints

Loosely correlated w/ facial features



Limited operation range

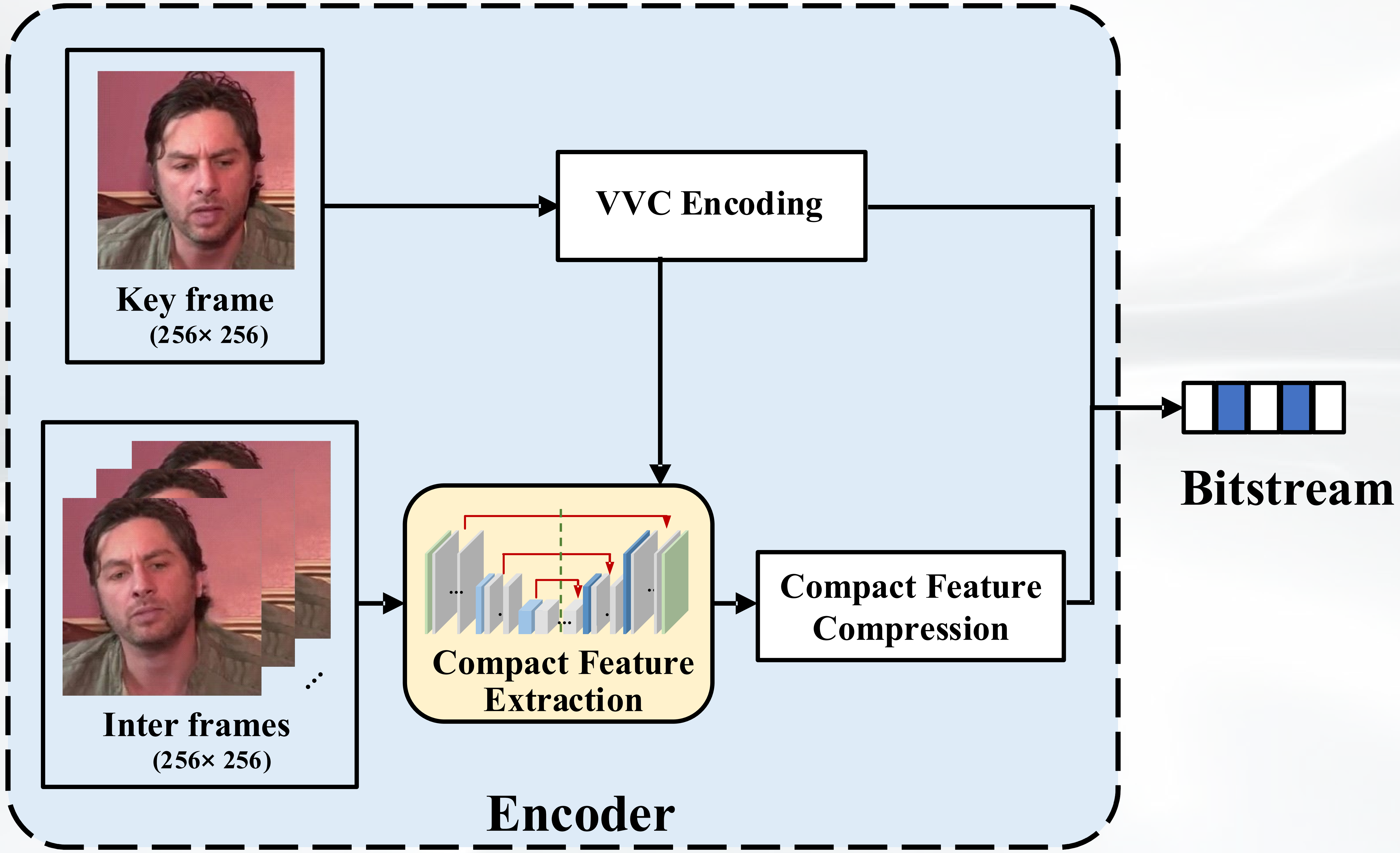
Separately drives motion flow



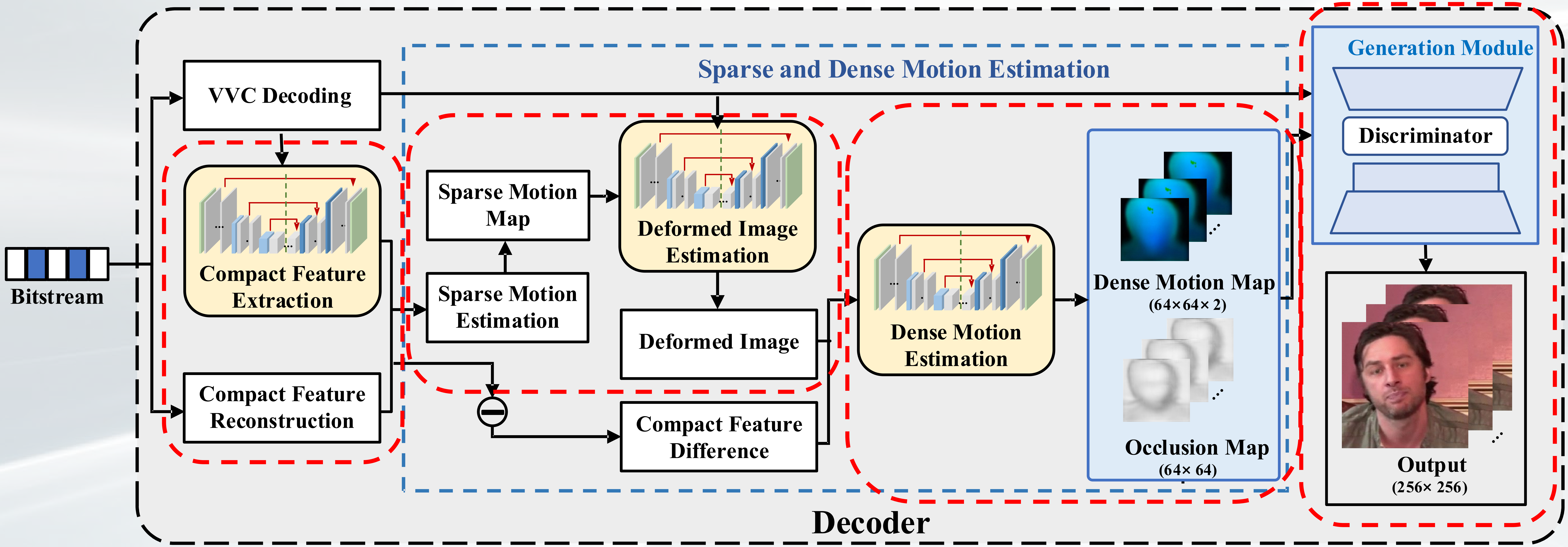
We aim to represent motion more efficiently and generate it more reliably

Compact feature for temporal evolution (CFTE)

CFTE encoder



CFTE decoder



CFTE work flow

Feature extraction

$$F_{comp} = g_{(Conv,GDN)}(f_{U-Net}(\phi(X,s)))$$

Sparse motion

$$M_{sparse} = GF_{flow}(\tilde{F}^{K}_{comp}, \tilde{F}^{I}_{comp}) \Rightarrow F_{cdf}$$

Coarse deformed frame

Dense motion & occlusion

$$M_{dense} = P_1(f_{U-Net}(concat(F_{cdf}, Diff_{<I,K>})))$$

$$M_{occlusion} = P_2(f_{U-Net}(concat(F_{cdf}, Diff_{<I,K>})))$$

where $Diff_{<I,K>} = \varphi(\tilde{F}^{I}_{comp}) - \varphi(\tilde{F}^{K}_{comp})$

Video frame generation

$$\hat{I} = M_{occlusion} \odot f_{U-Net}(K, M_{dense})$$

Training loss

Perceptual loss

$$L_{per-initial} = \sum_{n=1}^{i} \frac{1}{C_i \times H_i \times W_i} \left\| VGG_i(F_{cdf}) - VGG_i(\phi(I)) \right\|$$

$$L_{per-final} = \sum_{n=1}^{i} \frac{1}{C_i \times H_i \times W_i} \left\| VGG_i(\hat{I}) - VGG_i(I) \right\|$$

Adversarial loss

$$L_G(\hat{I}) = -\sum_{i=1}^{k} E_{\hat{I}\sim P_g}(D_i(\hat{I}))$$

$$L_D(\hat{I}, I) = \sum_{i=1}^{k} E_{\hat{I}\sim P_g}\left(D_i(\hat{I})\right) - \sum_{i=1}^{k} E_{\hat{I}\sim P_r}(D_i(I))$$

Total loss

$$L_{total} = \lambda_{inti\,al} \cdot L_{per-init\,ial} + \lambda_{fina\,l} \cdot L_{per-fin\,al} + \lambda_{adv} \cdot (L_G + L_D)$$

CFTE decoding flow visualization

Key frame | Current frame | CFTE map | Coarse deformed frame | Dense motion map | Occlusion map | Final output

CFTE entropy coding

CFTE map residual = Current CFTE map − Previous CFTE map

$$\begin{matrix} 0, & -1, & 0, & 1, \\ 0, & -1, & 1, & 1, \\ -1, & -1, & 0, & 0, \\ 0, & 0, & 1, & 0 \end{matrix} = \begin{matrix} 18, & -11, & 58, & 36, \\ 25, & -21, & 19, & -36, \\ 18, & -23, & 48, & -33, \\ 8, & 3, & 55, & -20 \end{matrix} - \begin{matrix} 18, & -10, & 58, & 35, \\ 25, & -20, & 18, & -37, \\ 19, & -22, & 48, & -33, \\ 8, & 3, & 54, & -20 \end{matrix}$$

↓

First-order Exp-Golomb binarization → ***CABAC*** → ***bitstream***

Experimental results

Experimental settings

VVC anchor	Generative methods
• VTM-10.0, LDB configuration • QPs {37, 42, 47, 52}	• First frame coded by VTM-10.0, QPs {37, 42, 47, 52} • FOMM based on https://github.com/AliaksandrSiarohin/first-order-model • Face_vid2vid from https://github.com/zhanglonghao1992/One-Shot_Free-View_Neural_Talking_Head_Synthesis • *Entropy coding of FOMM and Face_vid2vid keypoints are aligned with that of CFTE*

Test sequences

Resolution: 256x256

Frame rate: 25 fps

Duration: 10 sec

Cropped from open source database: https://ibug.doc.ic.ac.uk/resources/300-VW/ in RGB format

Distortion metrics

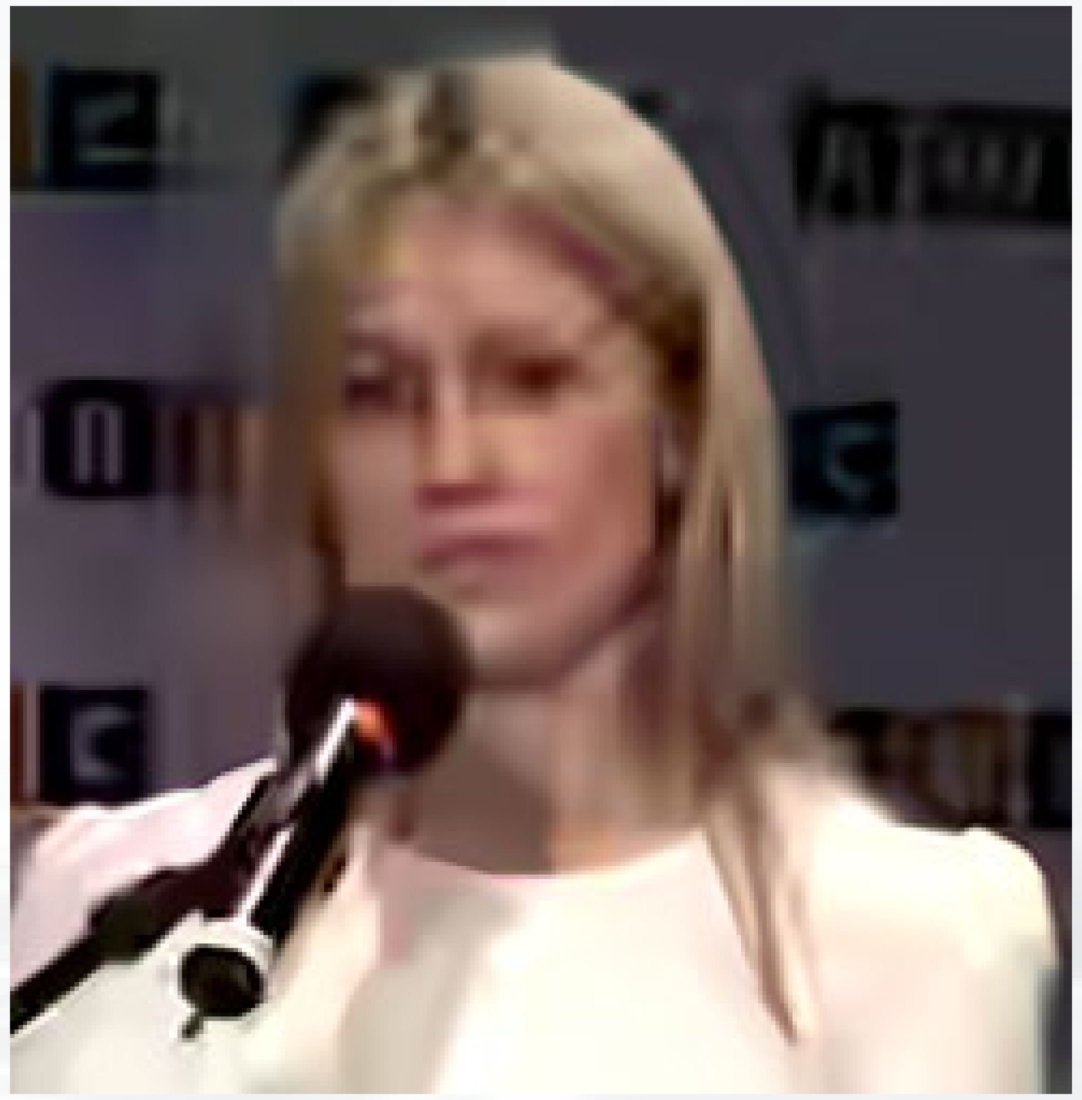
- Conventional metrics: PSNR, SSIM
- Learning-based distortion metrics:
  - LPIPS: *Learned Perceptual Image Patch Similarity*
  - DISTS: *Deep Image Structure and Texture Similarity*
- All metrics calculated with the open-source implementation from https://github.com/dingkeyan93/IQA-optimization

R. Zhang, et al. "The unreasonable effectiveness of deep features as a perceptual metric." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
K. Ding, et. al., "Image quality assessment: Unifying structure and texture similarity," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020
K. Ding, et al. "Comparison of full-reference image quality models for optimization of image processing systems." International Journal of Computer Vision 129.4 (2021): 1258-1281

Distortion metrics: a visualization

Original, frame #2

VTM-10.0, QP 47



VTM-10.0, QP 52

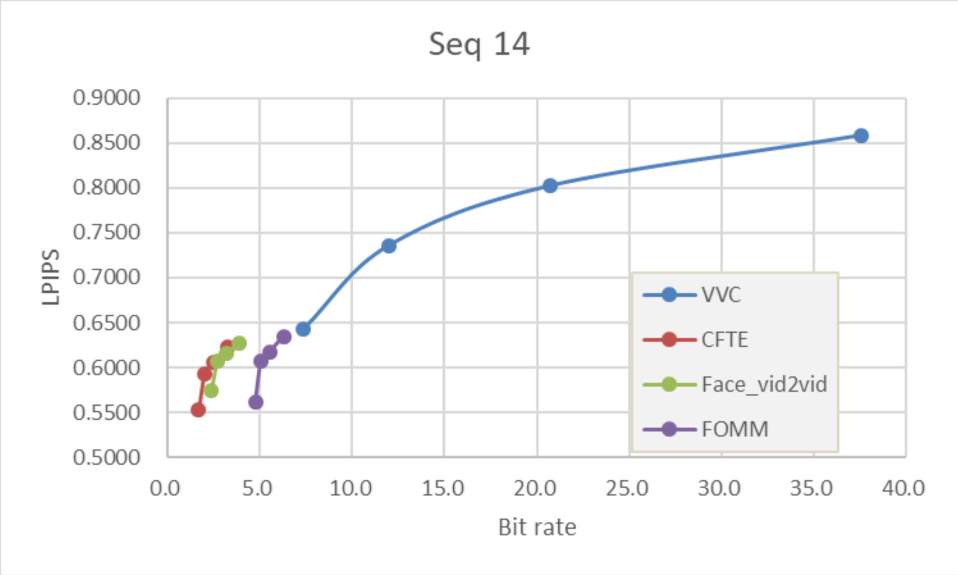
FOMM, I-QP 42

PSNR (↑)	30.18	27.31	24.36
SSIM (↑)	0.8816	0.8107	0.8139
LPIPS (↓)	0.2275	0.3399	0.1637
DISTS (↓)	0.1458	0.2007	0.1092

Generative methods do ***not*** optimize for sample-level fidelity

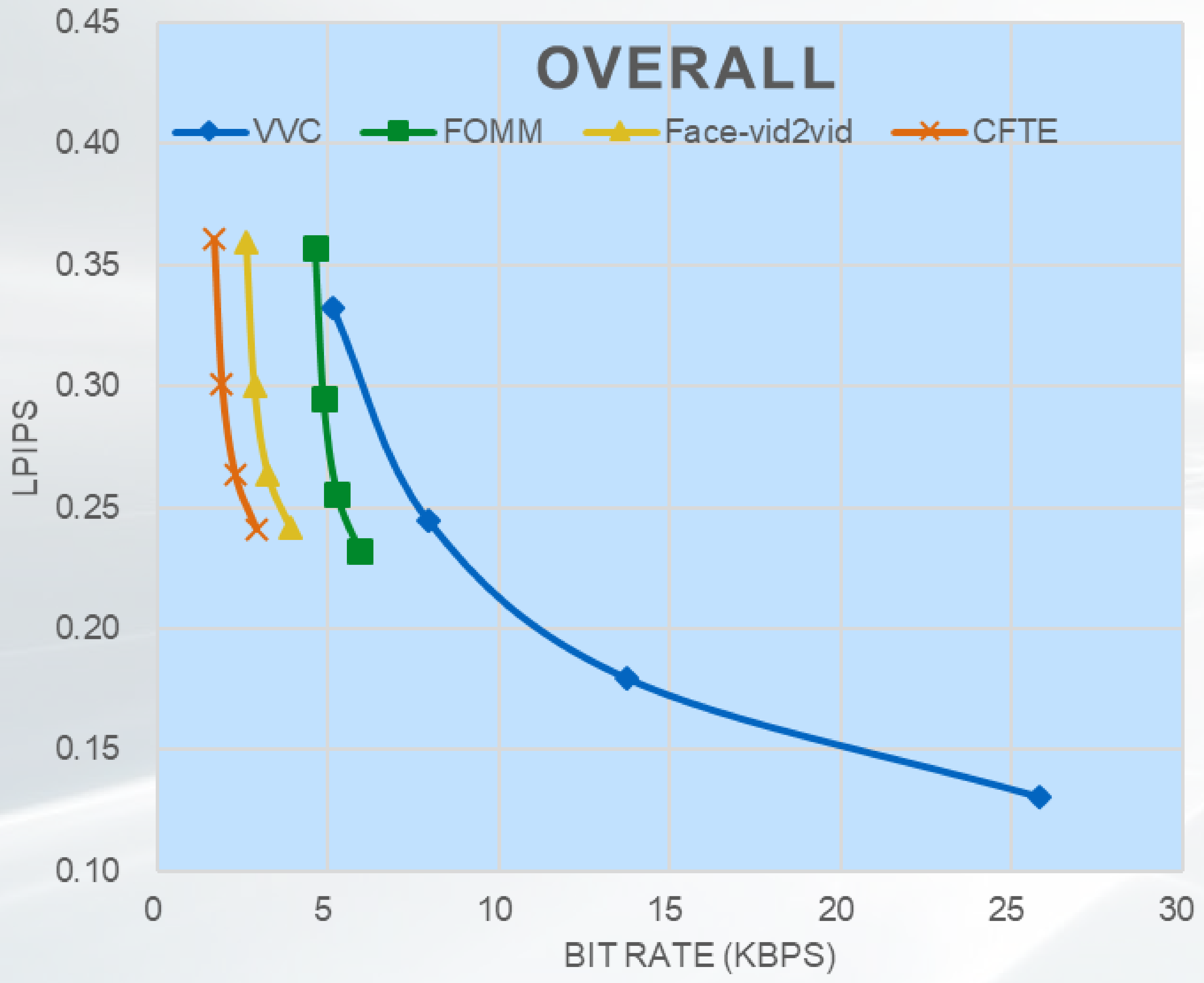
Rate reduction in terms of LPIPS & DISTS

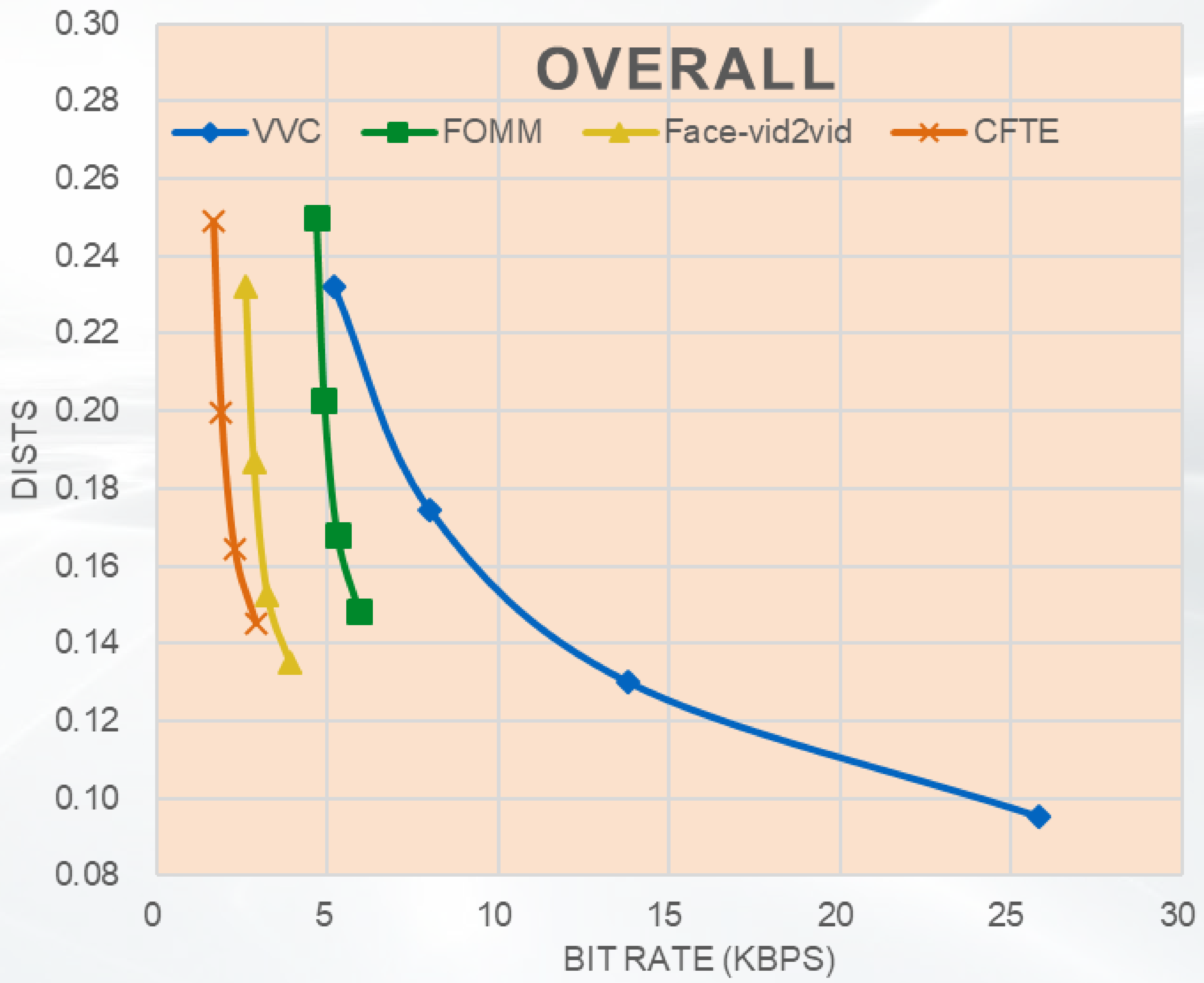
	LPIPS			DISTS		
	FOMM	Face_Vid2Vid	CFTE	FOMM	Face_Vid2Vid	CFTE
Seq 01	-36.2%	-51.4%	-74.3%	-41.9%	-57.3%	-74.2%
Seq 02	-9.4%	-43.8%	-65.5%	-18.0%	-56.3%	-69.7%
Seq 03	-13.6%	-46.3%	-64.6%	-14.2%	-52.5%	-68.5%
Seq 04*	0.0%	-34.6%	0.0%	-2.2%	-63.6%	-65.1%
Seq 05	-4.8%	-47.2%	-62.5%	-14.1%	-57.8%	-67.5%
Seq 06	-34.1%	-62.9%	-71.9%	-38.3%	-66.9%	-73.6%
Seq 07	-43.6%	-60.7%	-74.1%	-59.4%	-75.5%	-82.8%
Seq 08	-27.4%	-56.3%	-69.4%	-28.3%	-58.7%	-69.4%
Seq 09	-15.5%	-48.0%	-67.3%	-15.6%	-50.1%	-67.2%
Seq 10	-19.5%	-50.3%	-67.5%	-19.3%	-53.5%	-68.2%
Seq 11	-24.1%	-58.6%	-71.2%	-21.0%	-63.3%	-71.4%
Seq 12	-13.7%	-47.7%	-64.8%	-17.1%	-50.8%	-66.0%
Seq 13	-18.0%	-48.2%	-68.5%	-16.1%	-52.8%	-68.7%
Seq 14*	0.0%	0.0%	0.0%	-15.9%	-65.2%	-59.2%
Seq 15	-26.9%	-47.6%	-65.1%	-32.1%	-57.4%	-69.5%
Seq 16*	20.9%	-41.2%	0.0%	-12.7%	-65.8%	-62.5%
Seq 17	-40.6%	-58.5%	-71.7%	-46.1%	-64.0%	-73.4%
Seq 18	-21.8%	-49.1%	-68.6%	-26.8%	-53.1%	-70.1%
Seq 19	-11.0%	-28.2%	-58.7%	-12.5%	-45.7%	-62.2%
Seq 20*	0.0%	0.0%	0.0%	21.6%	-57.6%	-65.8%
Average	**-17.0%**	**-44.0%**	**-54.3%**	**-21.5%**	**-58.4%**	**-68.8%**
Average*	**-22.5%**	**-50.3%**	**-67.9%**	**-26.3%**	**-57.2%**	**-70.2%**



* Unreliable BD-rate calculation due to non-overlapped RD curves, removed from average* calculation

Rate-distortion performance: overall





VVC

Bit rate	6.64k
LPIPS	0.3627
DISTS	0.2243
PSNR	25.65
SSIM	0.7631

Face_Vid2Vid

Bit rate	6.18k
LPIPS	0.2135
DISTS	0.1183
PSNR	18.78
SSIM	0.6296

Original

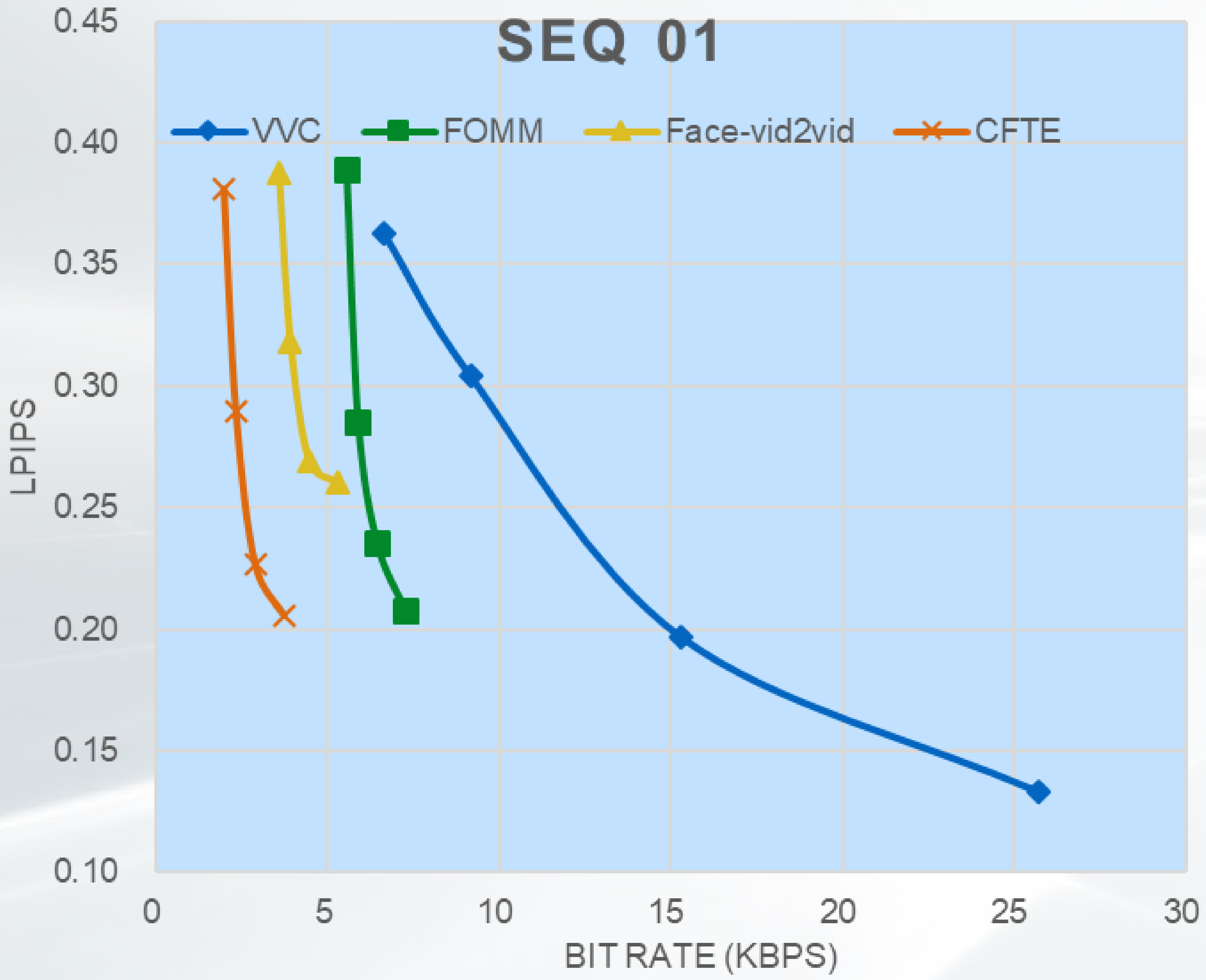
FOMM

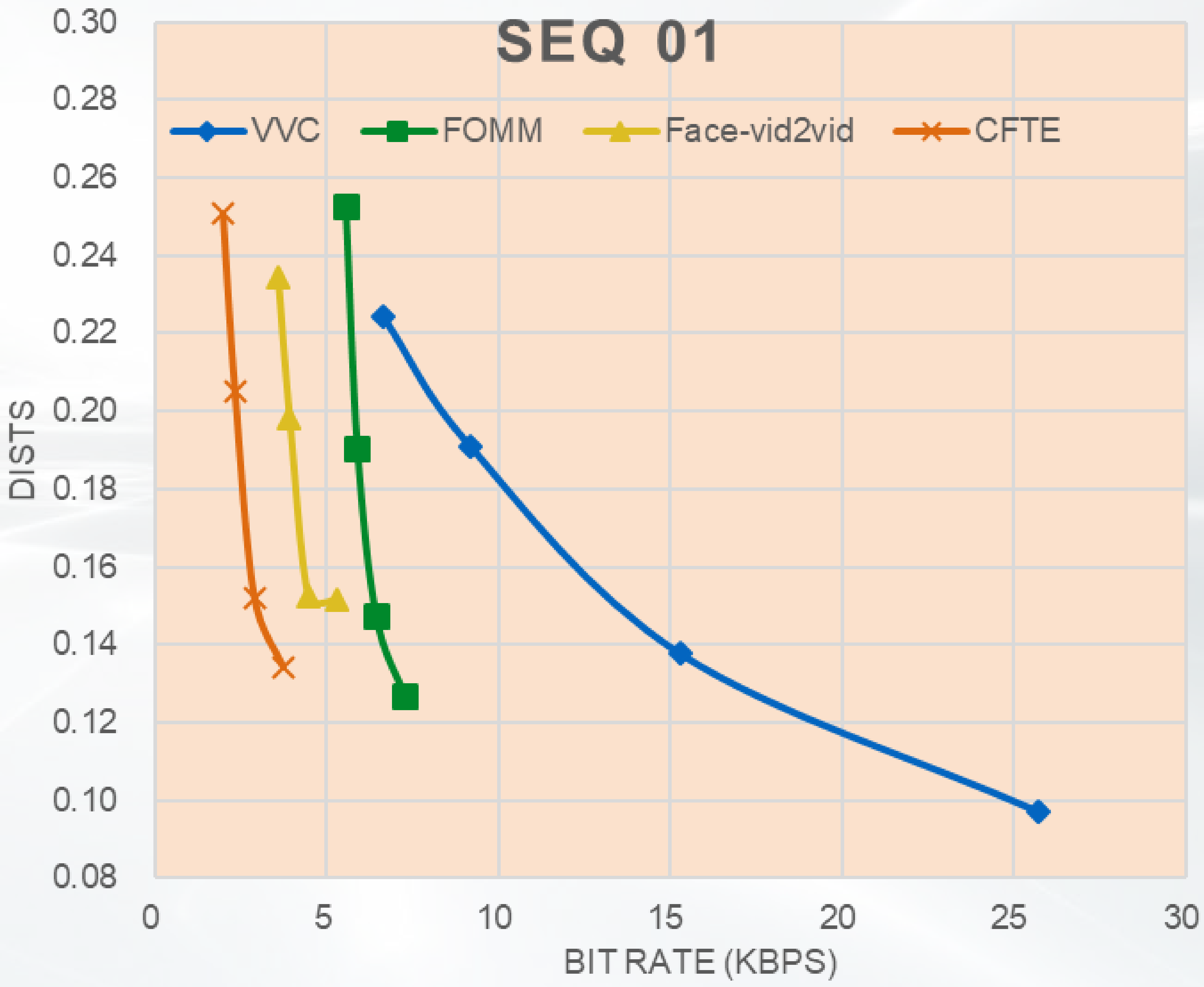
Bit rate	6.56k
LPIPS	0.2358
DISTS	0.1474
PSNR	20.18
SSIM	0.6815

CFTE

Bit rate	6.29k
LPIPS	0.1907
DISTS	0.0985
PSNR	19.37
SSIM	0.7013

Quality comparison @ similar bit rates: seq 01





VVC

Bit rate	5.20k
LPIPS	0.4074
DISTS	0.2618
PSNR	26.94
SSIM	0.7779

Face_Vid2Vid

Bit rate	5.27k
LPIPS	0.2001
DISTS	0.1198
PSNR	25.76
SSIM	0.7705

original

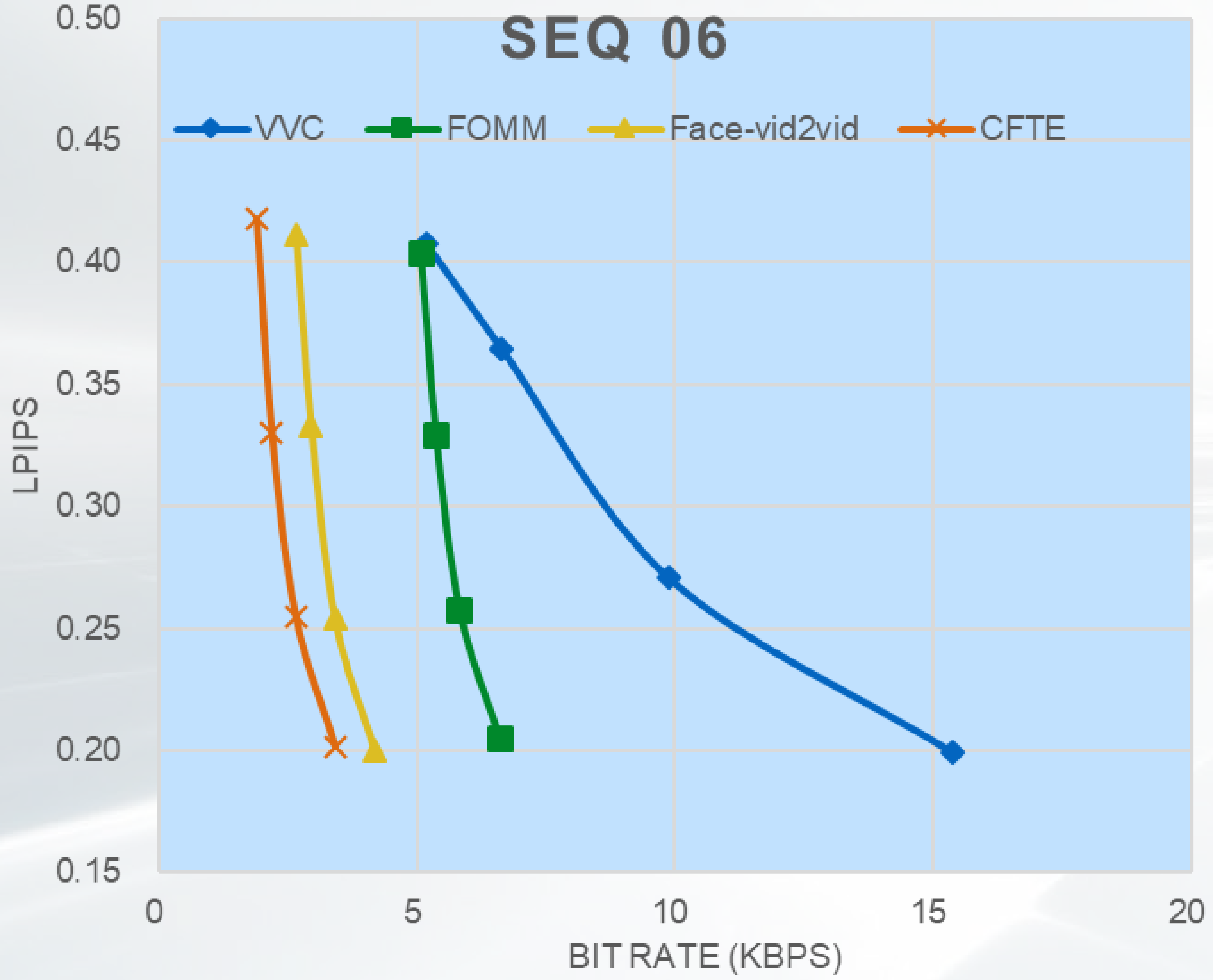
FOMM

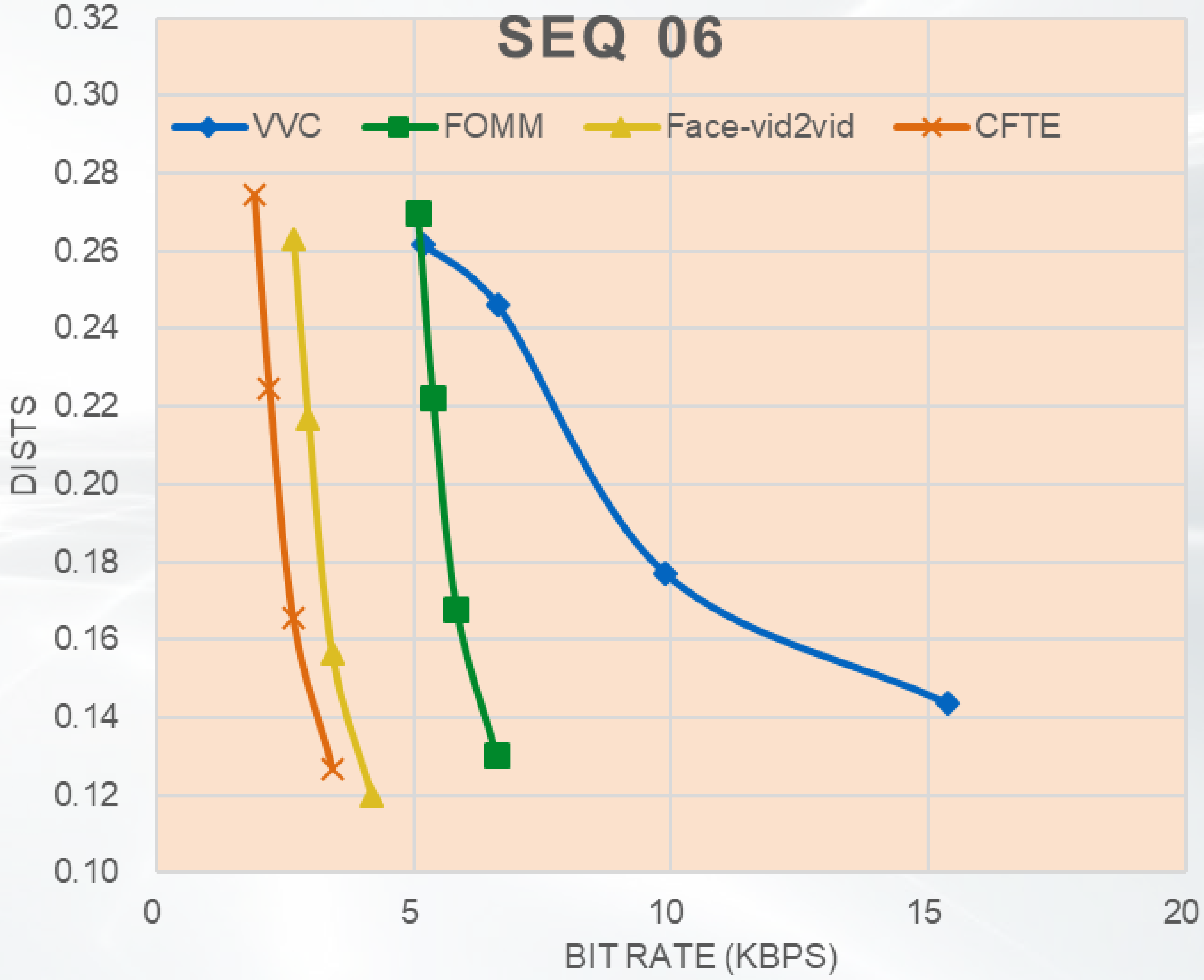
Bit rate	5.25k
LPIPS	0.3496
DISTS	0.2425
PSNR	25.51
SSIM	0.7396

CFTE

Bit rate	5.22k
LPIPS	0.1703
DISTS	0.0959
PSNR	25.75
SSIM	0.7717

Quality comparison @ similar bit rates: seq 06





original

VVC

Bit rate	18.71k
LPIPS	0.1218
DISTS	0.1011
PSNR	30.25
SSIM	0.8846

Face_Vid2Vid

Bit rate	4.67k
LPIPS	0.1235
DISTS	0.1016
PSNR	21.81
SSIM	0.7560

FOMM

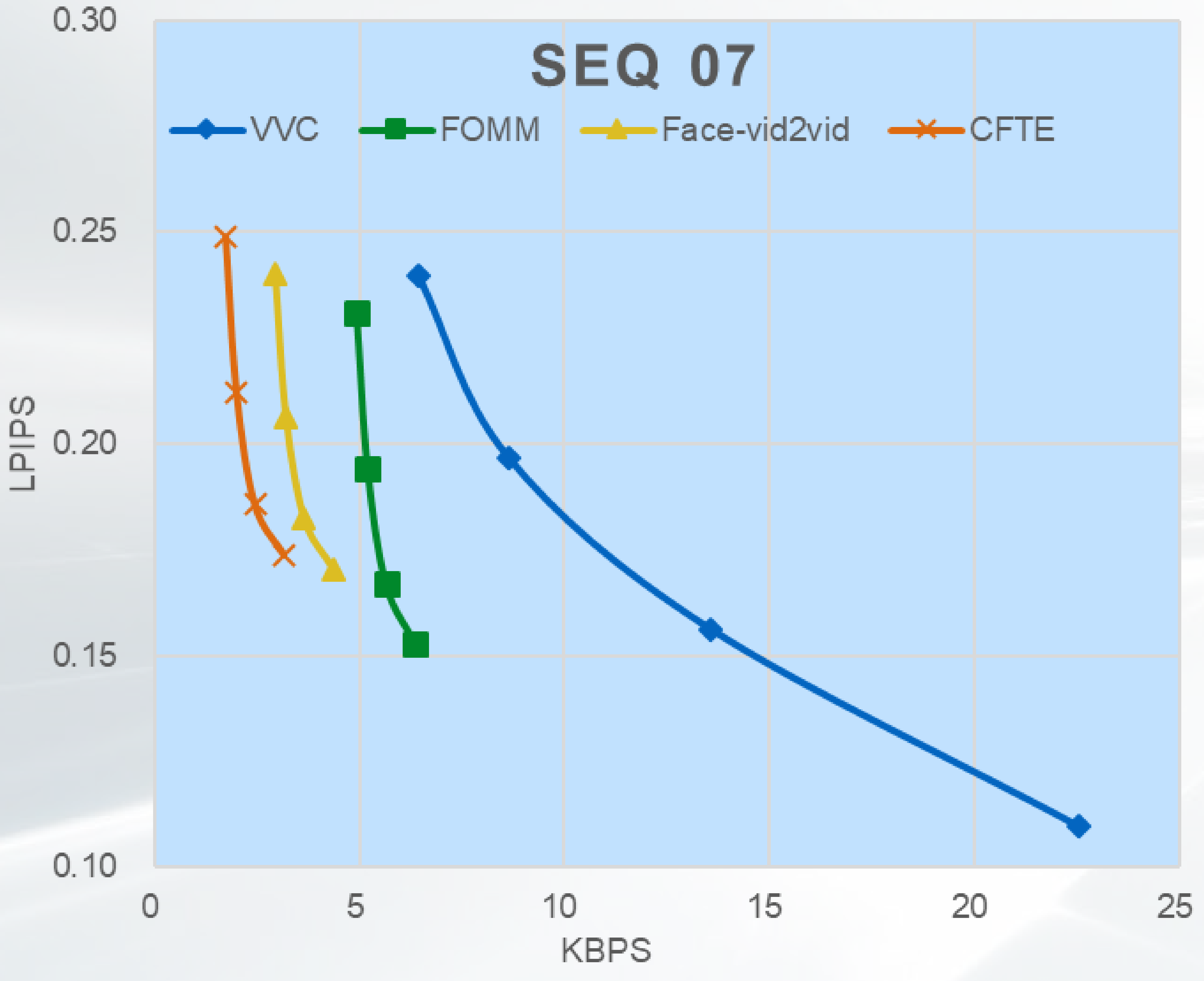
Bit rate	7.48k
LPIPS	0.1295
DISTS	0.1034
PSNR	22.27
SSIM	0.7692

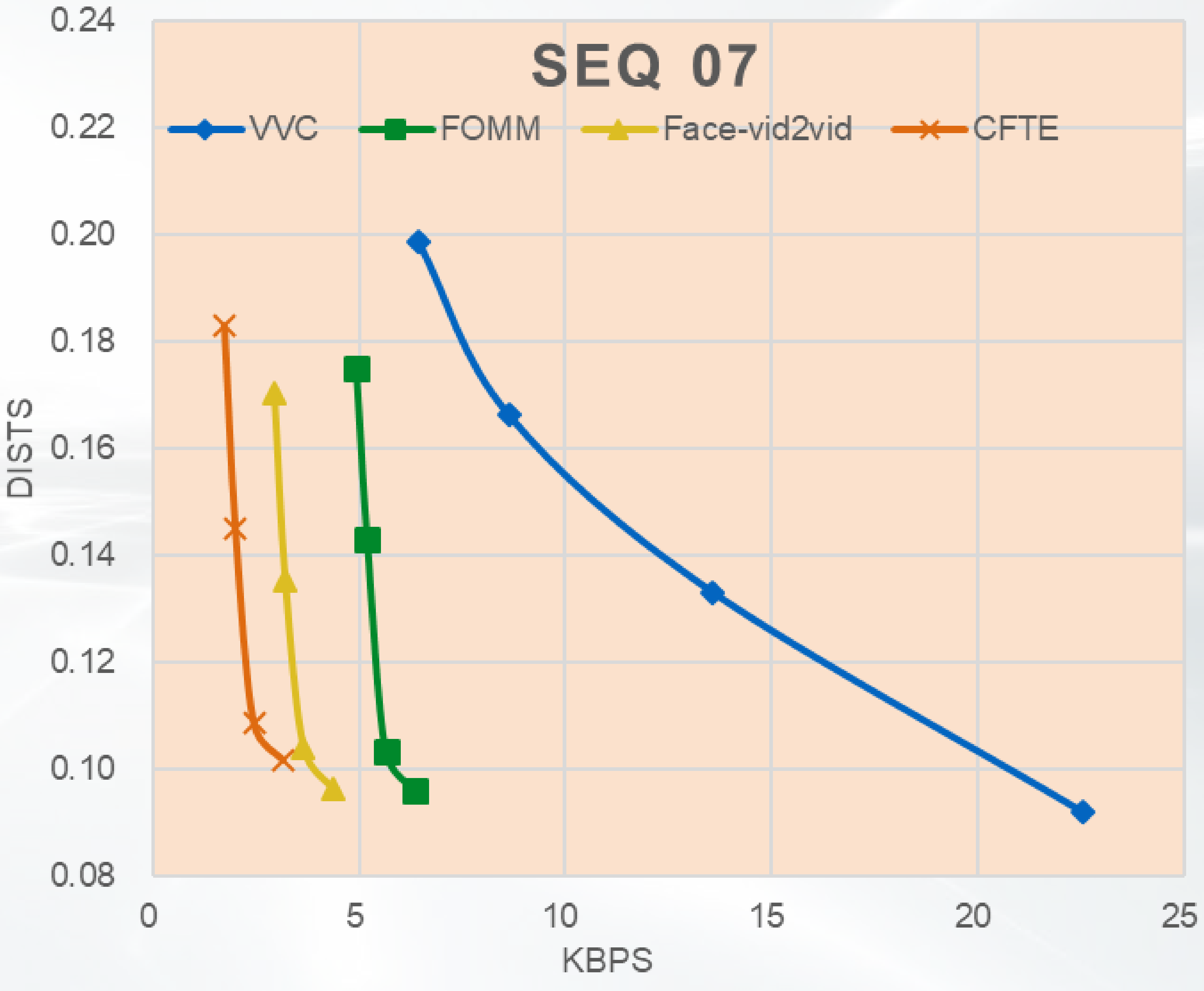
18% BR
of VVC

CFTE

Bit rate	3.31k
LPIPS	0.1187
DISTS	0.1002
PSNR	20.75
SSIM	0.7590

Bit rate comparison @ similar quality: seq 07





VVC

Bit rate	11.16k
LPIPS	0.1717
DISTS	0.1070
PSNR	30.68
SSIM	0.9113

original

Face_Vid2Vid

Bit rate	3.37k
LPIPS	0.1679
DISTS	0.1034
PSNR	22.71
SSIM	0.8203

FOMM

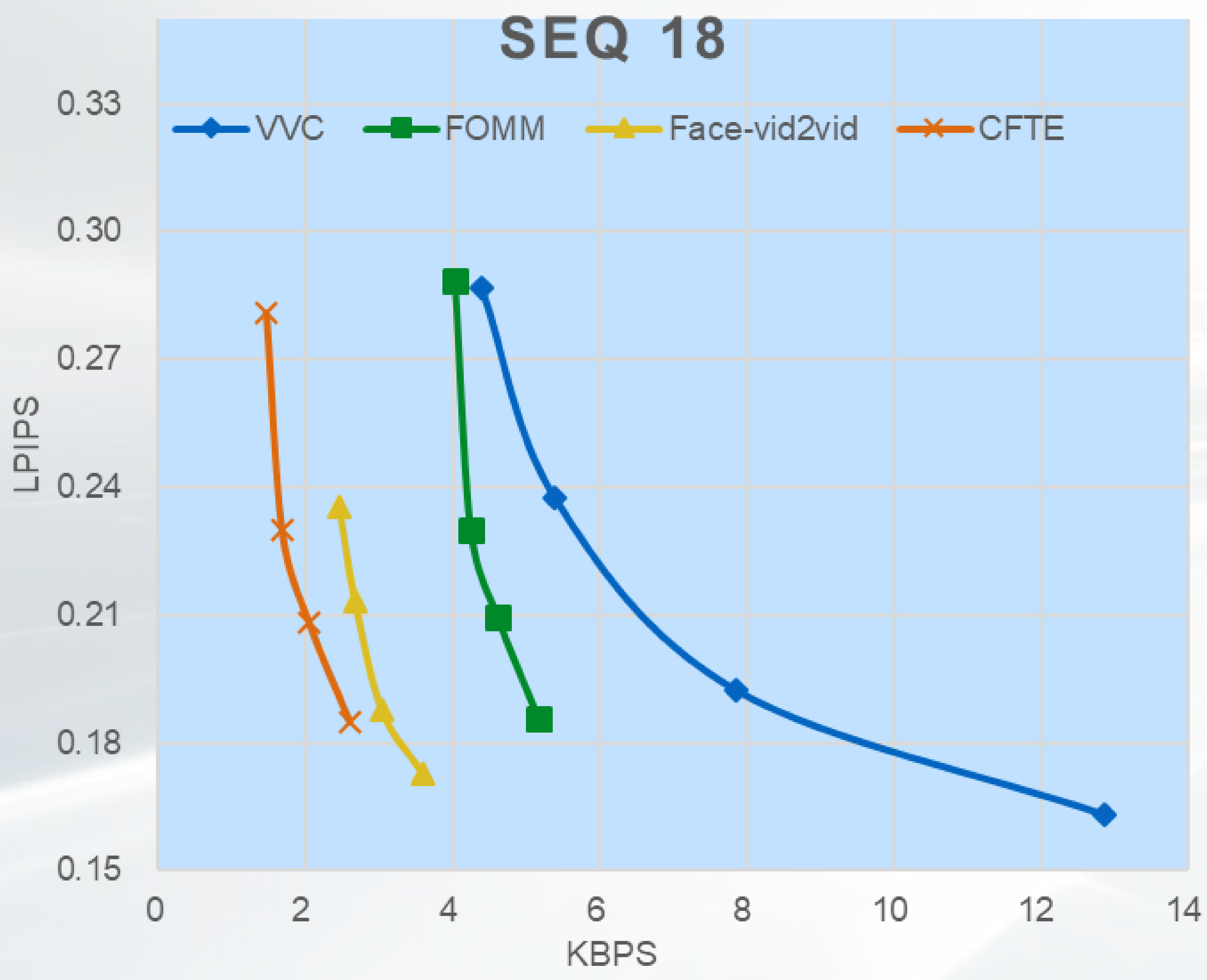
Bit rate	5.76k
LPIPS	0.1620
DISTS	0.1034
PSNR	23.89
SSIM	0.8417

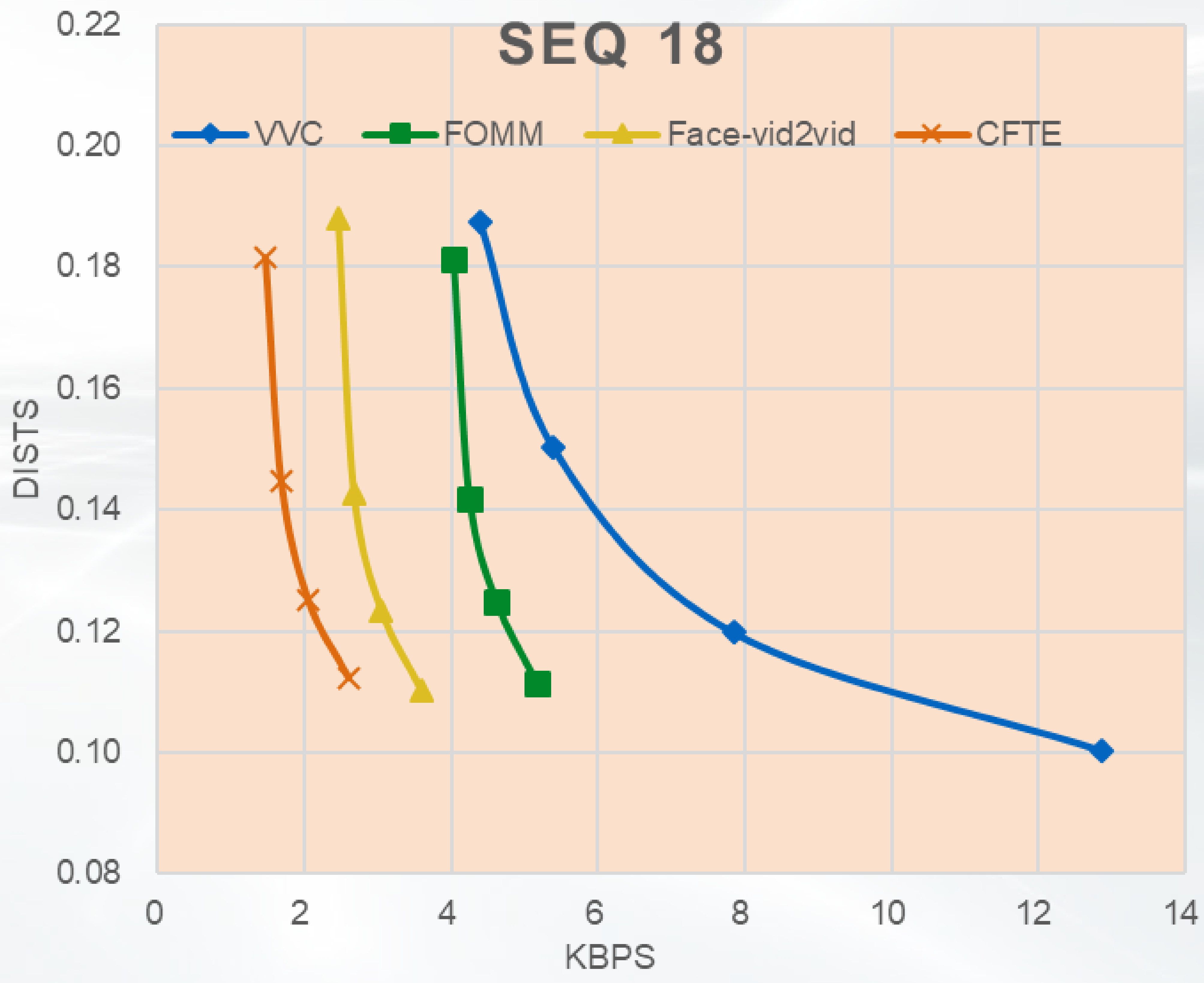
23% BR of VVC

CFTE

Bit rate	2.58k
LPIPS	0.1641
DISTS	0.1055
PSNR	23.13
SSIM	0.8314

Bit rate comparison @ similar quality: seq 18





Bit % of compact features/keypoints

1st-frame QP	FOMM	Face-vid2vid	CFTE
What is sent	10 x (2D-KP + Jacobian)	15 x (3D-KP) + exp + translation	4x4 CFTE map
QP = 52	92%	83%	72%
QP = 42	81%	65%	49%
QP = 32	61%	41%	27%

Alibaba Cloud
Worldwide Cloud Services Partner

CityU
香港城市大學
City University of Hong Kong
專業 創新 胸懷全球
Professional · Creative
For The World

PART 2:
THE CHALLENGES

The challenge of larger motion

When there is larger motion

Original

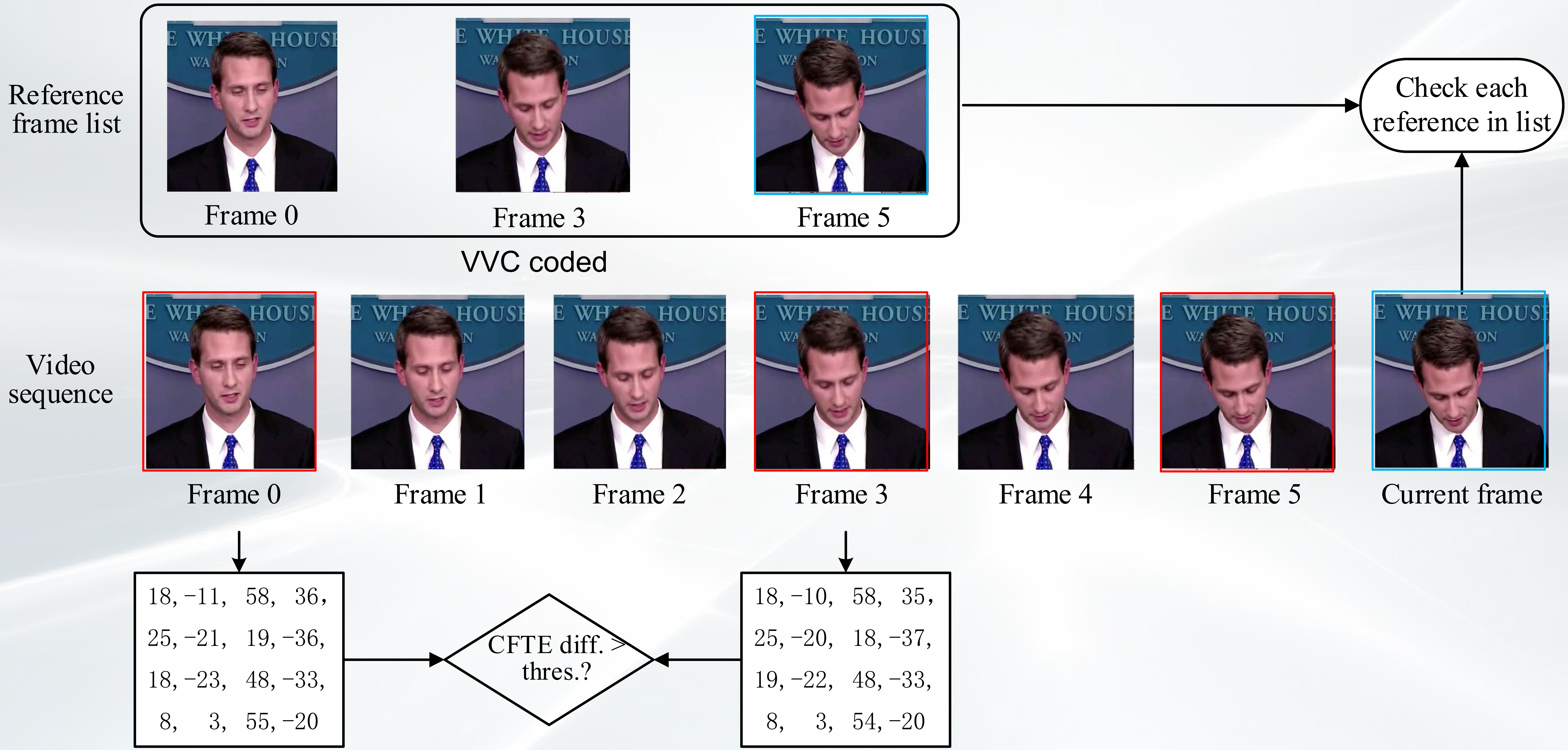
FOMM

All generative methods suffer from objectionable motion distortions

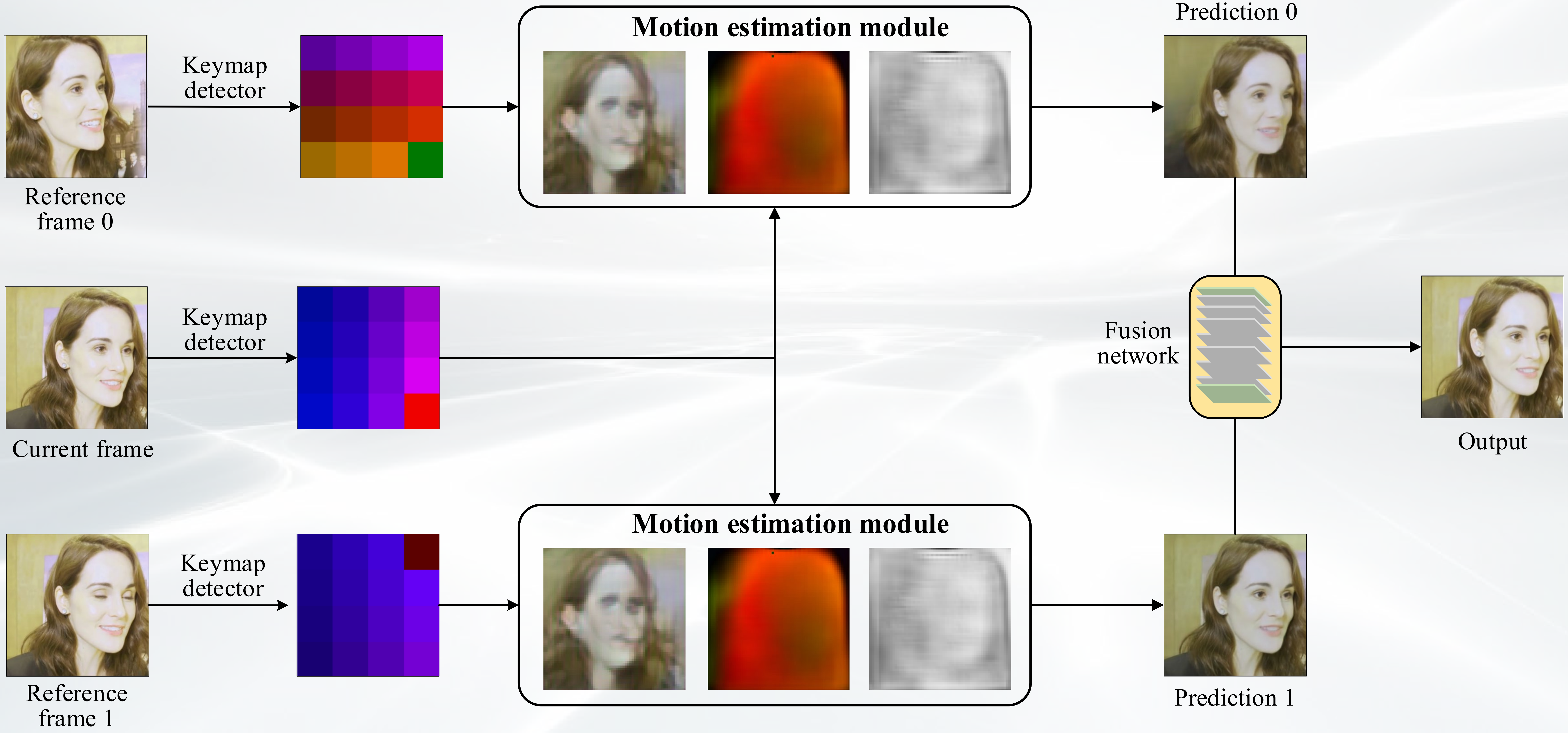
Face_vi2vid

CFTE

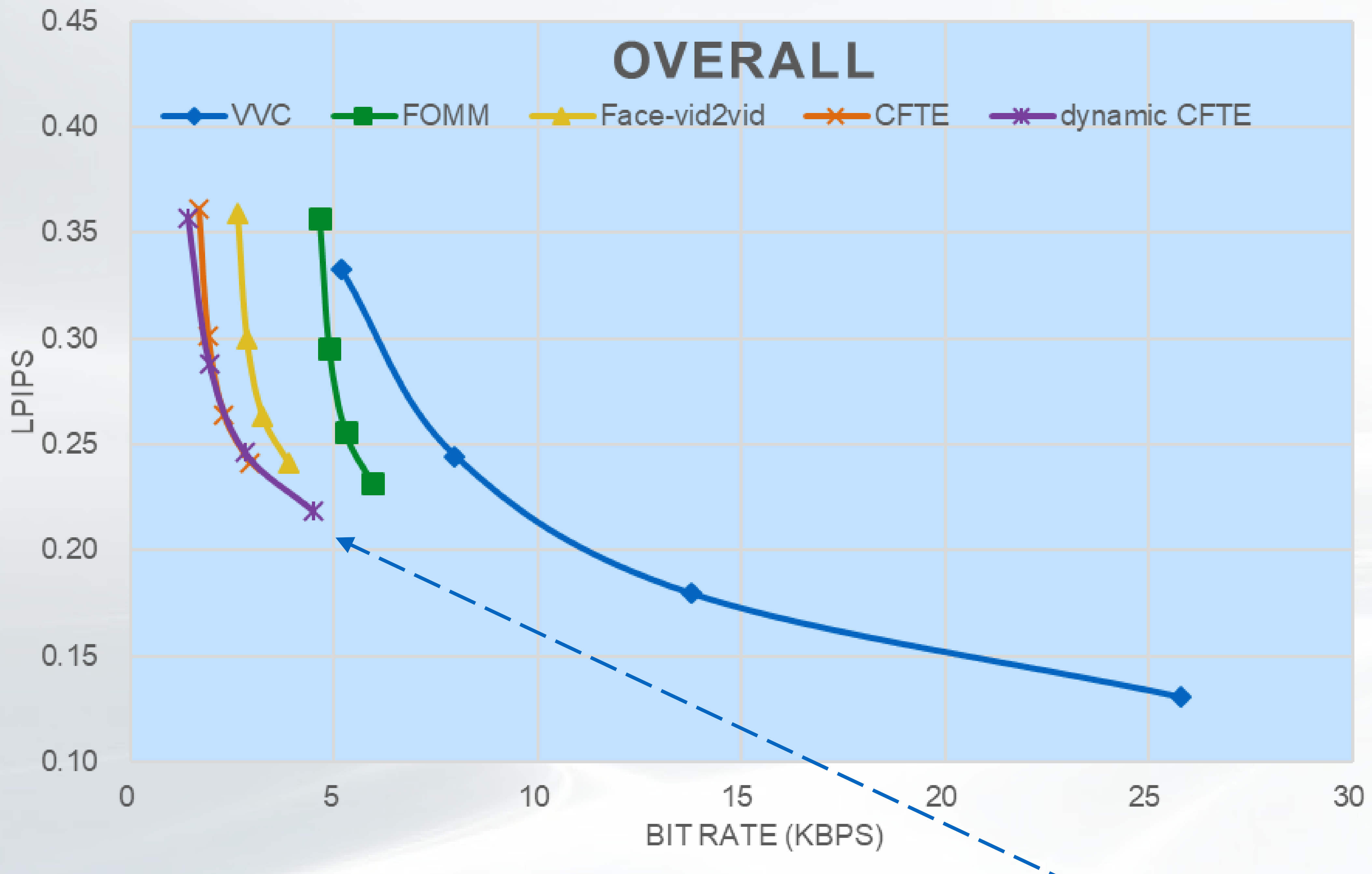
Dynamic reference refresh

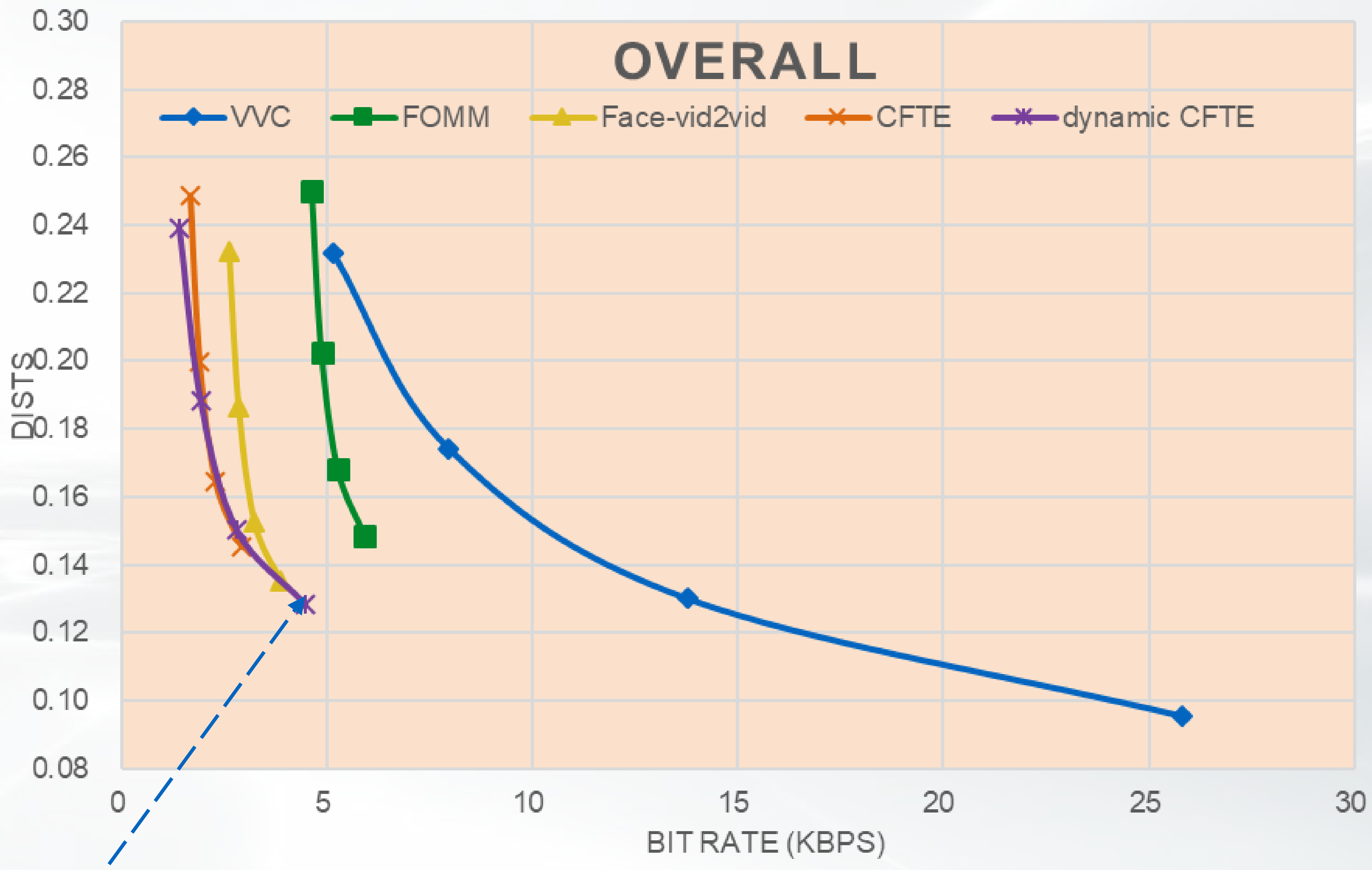


Multi-reference prediction



Rate distortion curves





Dynamic reference can extend CFTE's operation range towards higher quality

Alibaba Cloud
Worldwide Cloud Services Partner

CityU
香港城市大學
City University of Hong Kong

Seq 13 original

VVC

BR	LPIPS	DISTS	PSNR	SSIM
4.44k	0.3868	0.2198	27.72	0.8216

FOMM

BR	LPIPS	DISTS	PSNR	SSIM
4.82k	0.2691	0.1625	23.84	0.7933

Face-vid2vid

BR	LPIPS	DISTS	PSNR	SSIM
4.27k	0.2634	0.1362	20.15	0.7276

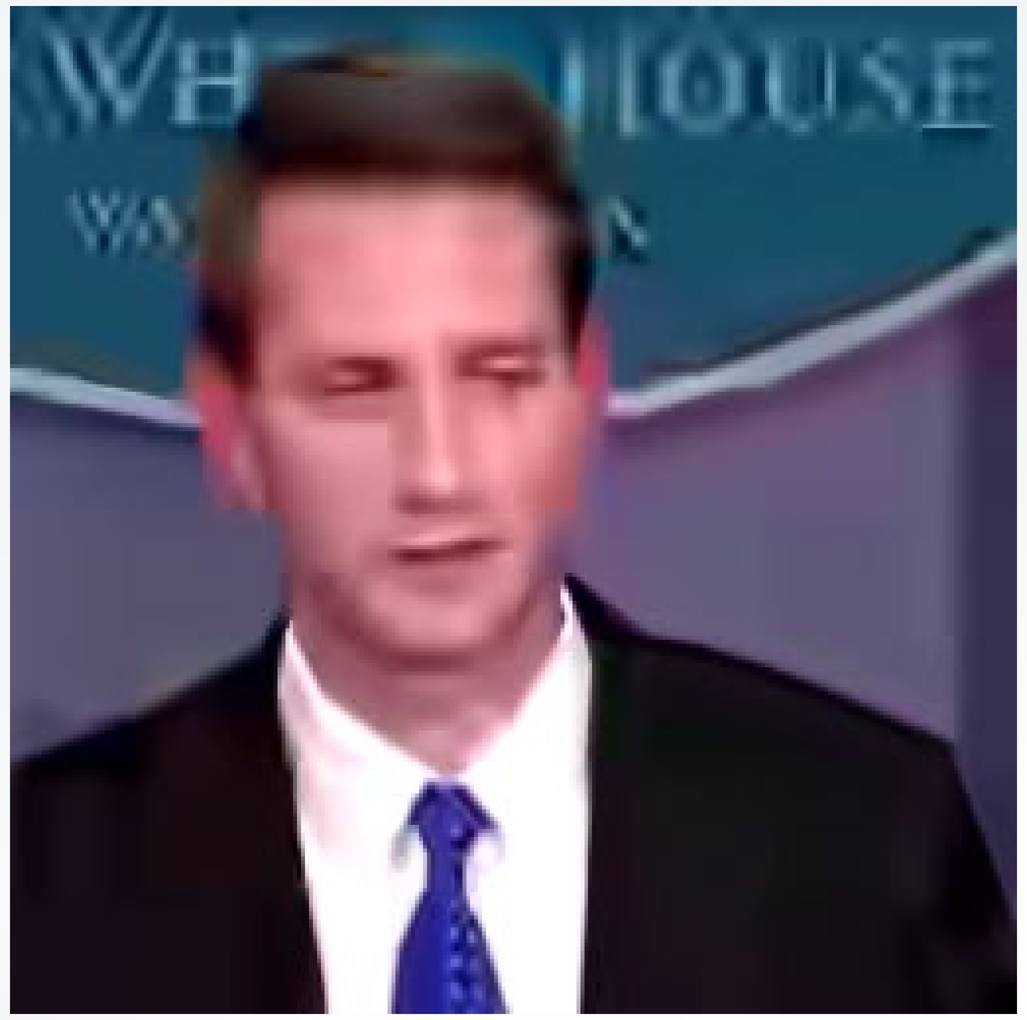
CFTE

BR	LPIPS	DISTS	PSNR	SSIM
4.03k	0.2022	0.1154	22.23	0.7865

Dynamic & multi ref

BR	LPIPS	DISTS	PSNR	SSIM
4.41k	0.1783	0.0932	24.91	0.8052

Seq 19 original



VVC

BR	LPIPS	DISTS	PSNR	SSIM
5.47k	0.3104	0.1988	25.85	0.7915

FOMM

BR	LPIPS	DISTS	PSNR	SSIM
5.90k	0.2628	0.1702	20.00	0.6753

Face-vid2vid

BR	LPIPS	DISTS	PSNR	SSIM
5.41k	0.3004	0.1602	16.74	0.6359

CFTE

BR	LPIPS	DISTS	PSNR	SSIM
5.17k	0.2487	0.1493	18.85	0.6776

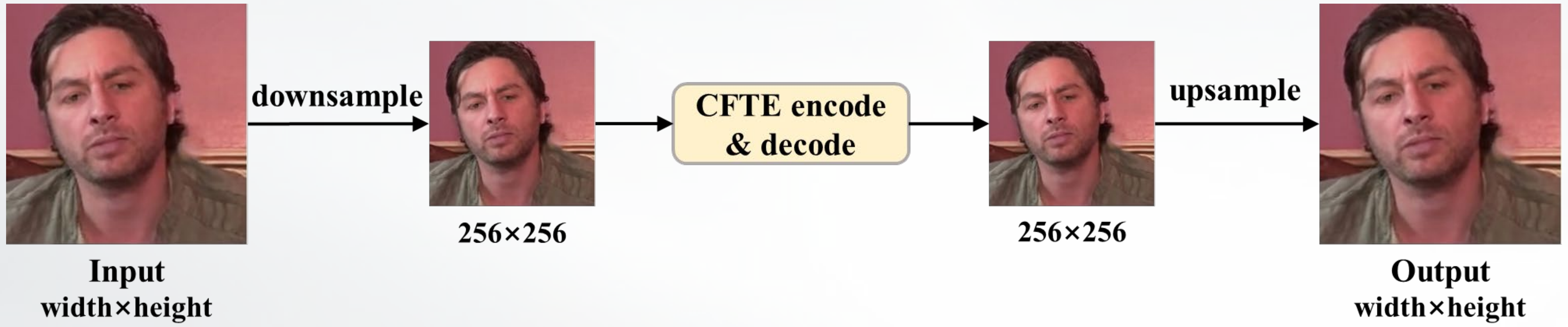
Dynamic & multi ref

BR	LPIPS	DISTS	PSNR	SSIM
5.56k	0.1637	0.0967	22.89	0.7241

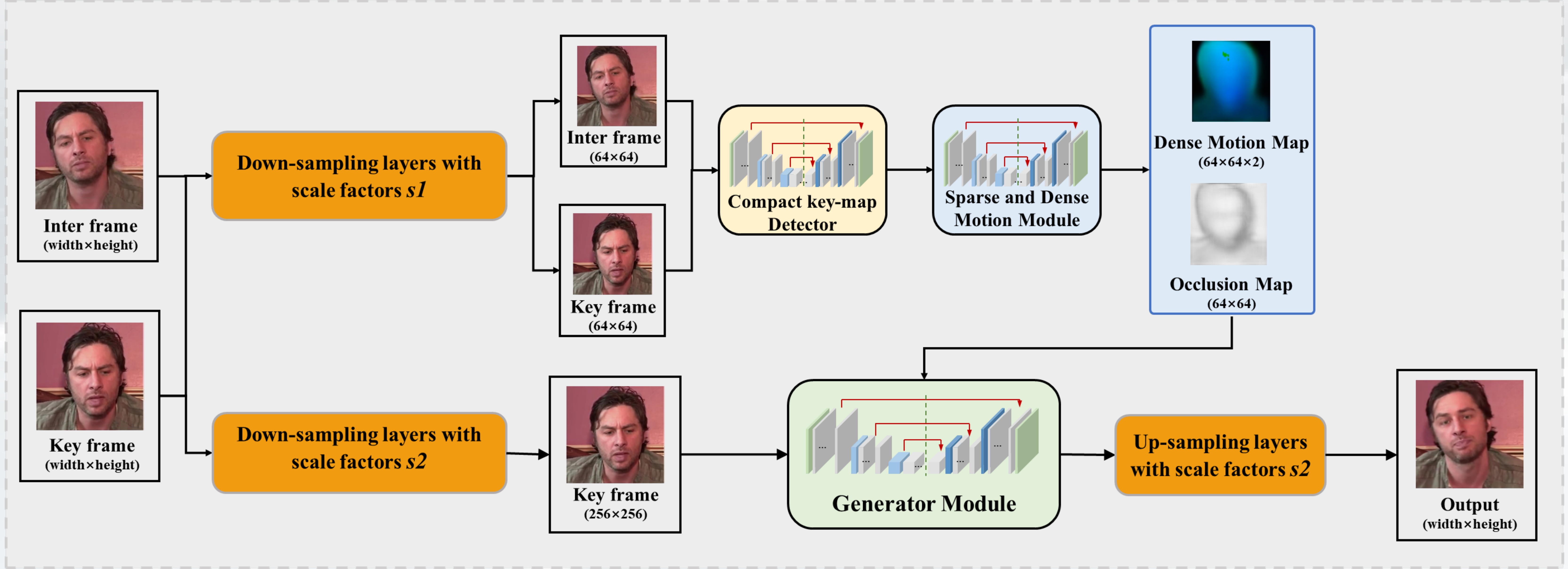
Adapting to larger resolutions

Resolution adaption

➤ Resize to 256x256 for coding (bicubic filters as pre- and post-processing)

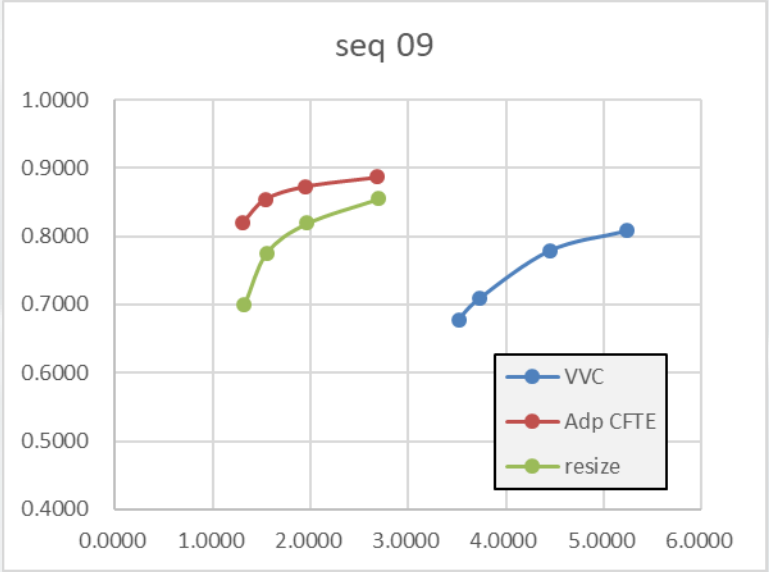


➤ Adaptive CFTE: embedding down-/up-sampling layers within the CFTE workflow



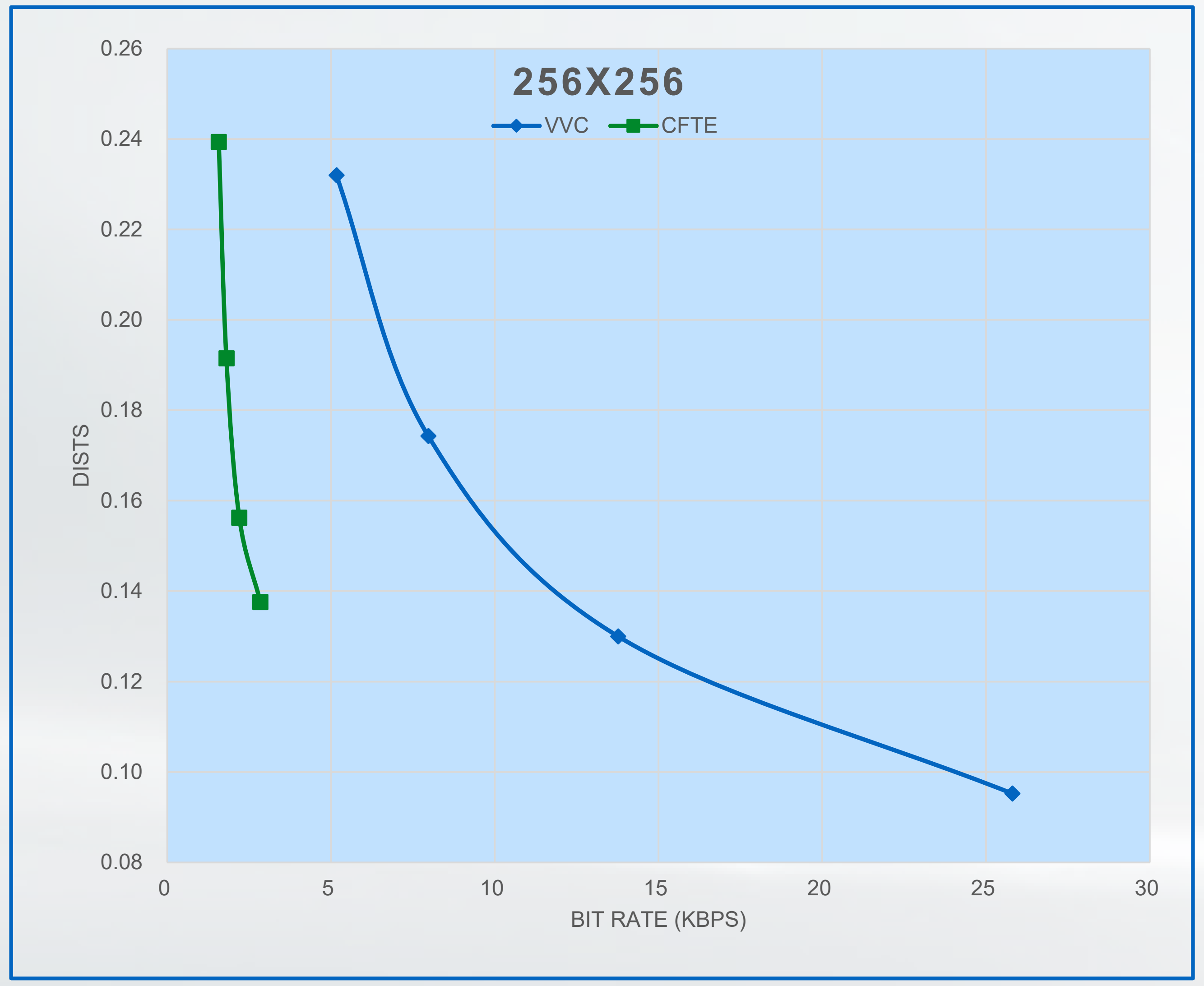
Objective performance: DISTS

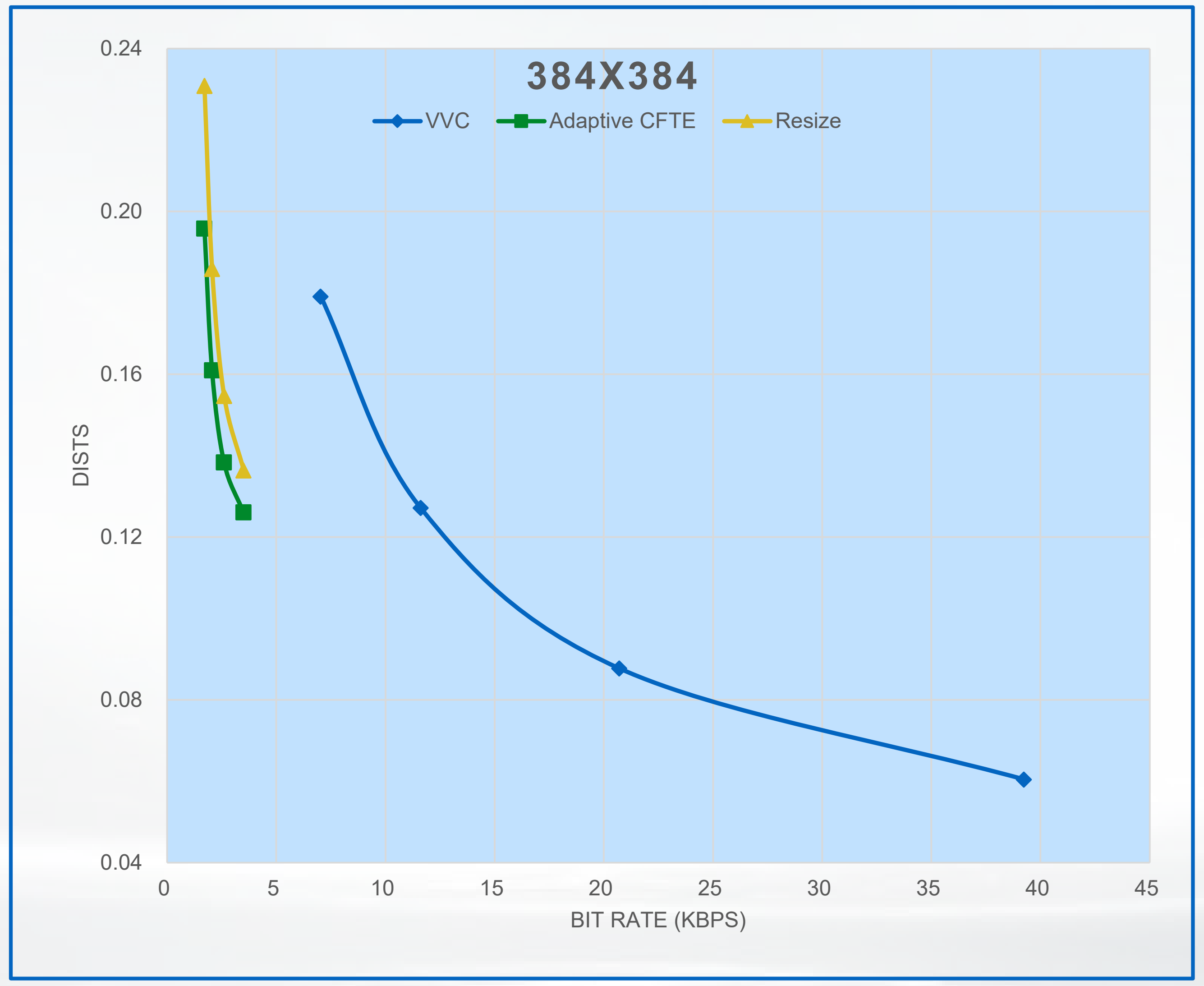
	384x384		512x512		640x640	
	Resize	Adaptive	Resize	Adaptive	Resize	Adaptive
Seq 01	-72.9%	-77.1%	-71.3%	-79.6%	-71.2%	-82.0%
Seq 02	-68.9%	-73.1%	-67.1%	-75.5%	-68.9%	-78.4%
Seq 03	-61.8%	-65.1%	-58.3%	-64.8%	-60.6%	-69.7%
Seq 04	-51.2%	-65.5%	-55.9%	-70.7%	-57.7%	-68.8%
Seq 05	-71.0%	-75.8%	-70.2%	-76.1%	-70.0%	-76.9%
Seq 06	-68.8%	-73.9%	-66.6%	-77.4%	-61.8%	-76.8%
Seq 07	-80.7%	-84.2%	-79.7%	-85.5%	-77.9%	-86.1%
Seq 08	-69.0%	-74.8%	-65.5%	-72.7%	-61.6%	-73.8%
Seq 09*	-68.8%	-72.9%	-67.3%	-74.6%	-64.8%	0.0%
Seq 10*	-69.4%	-74.0%	-67.9%	-73.8%	-65.0%	0.0%
Seq 11*	-68.3%	-74.1%	-68.2%	0.0%	-68.7%	0.0%
Seq 12*	-65.5%	-70.7%	-61.2%	-70.3%	-59.6%	0.0%
Seq 13	-67.3%	-70.3%	-64.3%	-72.2%	-61.6%	-72.6%
Seq 14*	0.0%	0.0%	0.0%	-53.6%	-56.8%	-69.0%
Seq 15	-66.8%	-73.4%	-64.3%	-74.8%	-59.3%	-75.5%
Seq 16	-56.6%	-64.4%	-56.2%	-66.2%	-56.6%	-69.4%
Seq 17	-68.7%	-74.1%	-66.7%	-76.6%	-65.6%	-75.9%
Seq 18	-66.7%	-74.2%	-62.5%	-72.1%	-63.7%	-73.9%
Seq 19	-60.3%	-65.3%	-54.7%	-61.2%	-56.8%	-63.4%
Seq 20	-60.4%	-66.7%	-59.3%	-66.4%	-60.4%	-71.1%
Average	**-63.2%**	**-68.5%**	**-61.4%**	**-68.2%**	**-63.4%**	**-59.2%**
Average*	**-66.1%**	**-71.9%**	**-64.2%**	**-72.8%**	**-63.6%**	**-74.3%**

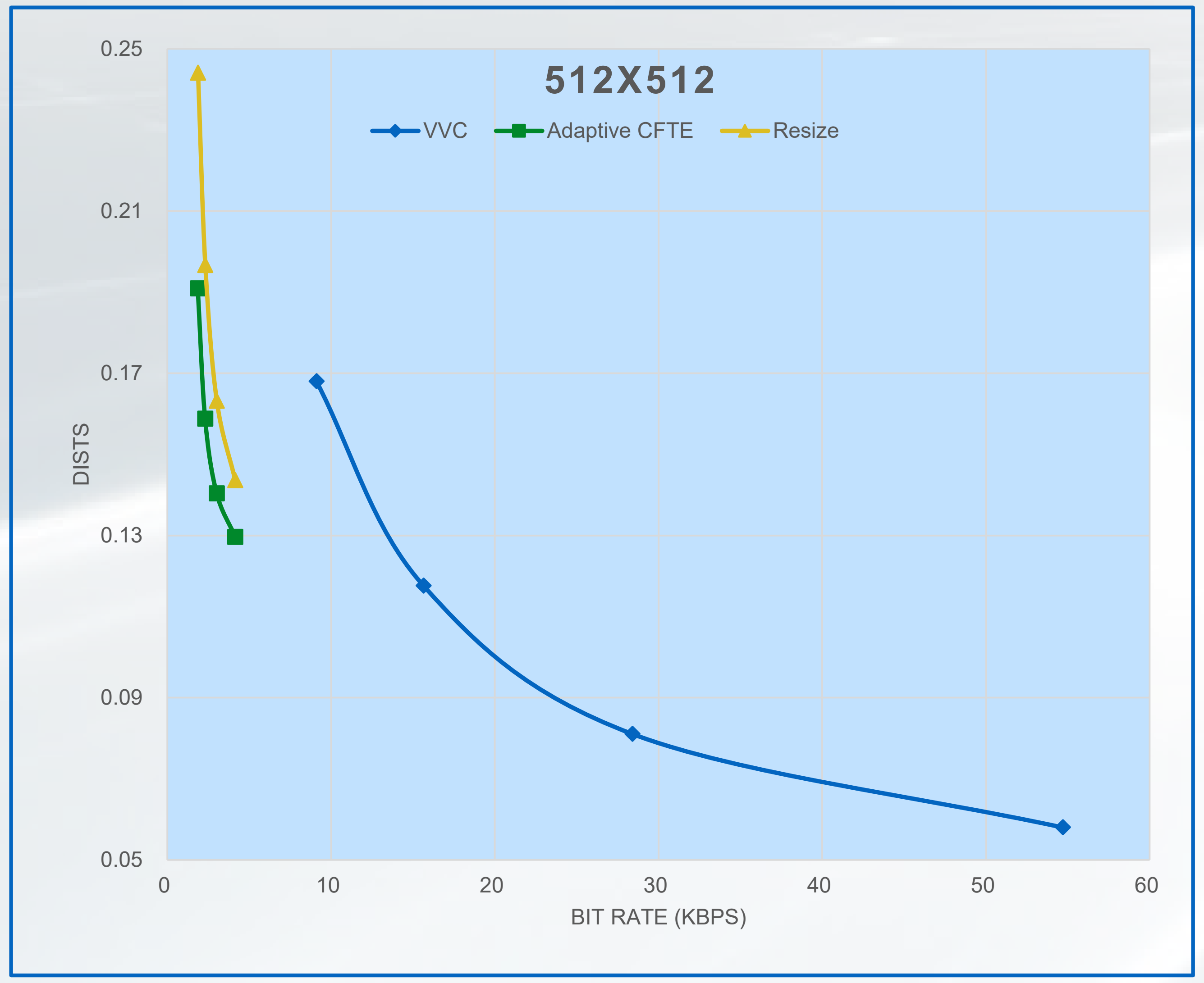


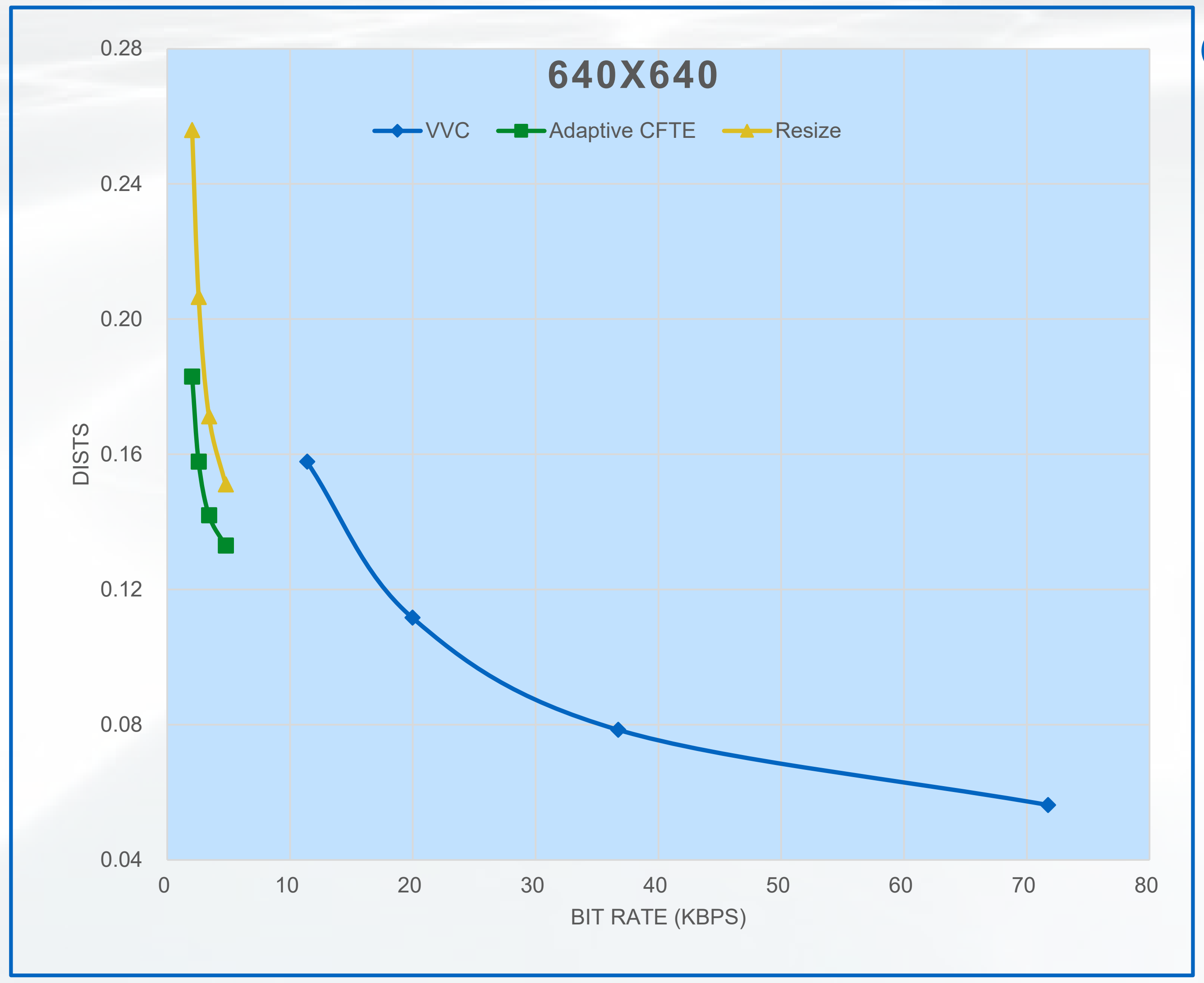
* Unreliable BD-rate calculation due to non-overlapped RD curves, removed from average* calculation

Rate distortion performance: DISTS









By absorbing scaling within the CFTE process, adaptive CFTE shows robust performance for all resolutions

Visual quality @ similar rate: 384x384

VVC

Bit rate	4.37k
DISTS	0.3049
LPIPS	0.5029

Resize

Bit rate	4.08k
DISTS	0.1191
LPIPS	0.2467

Adaptive CFTE

Bit rate	4.08k
DISTS	0.0967
LPIPS	0.2196

Visual quality @ similar rate: 512x512

VVC

Bit rate	6.85k
DISTS	0.2457
LPIPS	0.3649

Resize

Bit rate	6.98k
DISTS	0.1014
LPIPS	0.2462

Adaptive CFTE

Bit rate	6.97k
DISTS	0.0899
LPIPS	0.2307

Visual quality @ similar rate: 640x640

VVC

Bit rate	8.43k
DISTS	0.2134
LPIPS	0.3920

Resize

Bit rate	8.23k
DISTS	0.1106
LPIPS	0.2615

Adaptive CFTE

Bit rate	8.22k
DISTS	0.0956
LPIPS	0.2471

Complexity challenge

Computational and model complexity

		FOMM	Nvidia	CFTE
Encoder	Parameter Number	38.9M	68.1M	43.6M
	Macs per pixel	14.5G	26.2G	18.7G
	Inference speed	28fps	11fps	15fps
Decoder	Parameter Number	82.4M	96.7M	85.8M
	Macs per pixel	33.7G	39.2G	36.8G
	Inference speed	21fps	8fps	13fps

Inference 256x256 video on Tesla-V100 and 22 core CPU (Intel(R) Xeon(R) Platinum 8163 CPU @ 2.50GHz)

CONCLUDING REMARKS

Concluding remarks

Generative face compression has shown much promise

- Preserve clearer facial features @ ultra-low bit rate ranges
- Significant BD rate reduction over VVC
- Face composition in 3D space

But it also faces many challenges

- Avoidance of objectionable distortions
- Higher quality reconstruction, esp. expression, local motion, etc
- Complexity reduction esp. @ decoder side

So does AI-based video compression

- Expanding beyond head-and-shoulder scenario
- General-purpose high performance video compression using AI-based methodology
- Quality metrics beyond PSNR and SSIM, e.g. AI-based

Acknowledgment

I'd like to thank my wonderful collaborators:

Dr. Shiqi Wang, Assistant Prof., *City University of Hong Kong*
Dr. Zhao Wang, *Alibaba Group*
Bolin Chen, Ph.D. student, *City University of Hong Kong*
Binzhe Li, Ph.D. student, *City University of Hong Kong*

I have learned a lot working with you all!

References

1. B. Bross, J. Chen, J.-R. Ohm, G. J. Sullivan, and Y.-K. Wang, “Developments in international video coding standardization after AVC, with an overview of Versatile Video Coding (VVC),” Proceedings of the IEEE, 2021.
2. B. Bross, et al. "Overview of the versatile video coding (VVC) standard and its applications." IEEE Transactions on Circuits and Systems for Video Technology 31.10 (2021): 3736-3764.
3. J. Ballé, V. Laparra, and E. P. Simoncelli. "End-to-end optimized image compression." In International Conference on Learning Representations (ICLR), 2017.
4. D. Minnen, J. Ballé, and G. Toderici. "Joint autoregressive and hierarchical priors for learned image compression." In Advances in Neural Information
5. Processing Systems, pages 10771–10780, 2018.
6. 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, pp. 11006–11015, 2019.
7. G. Lu, X. Zhang, W. Ouyang, L. Chen, Z. Gao, and D. Xu, “An end-to-end learning framework for video compression,” IEEE transactions on pattern analysis and machine intelligence, 2020.
8. P. Eisert, T. Wiegand, and B. Girod, “Model-aided coding: a new approach to incorporate facial animation into motion-compensated video coding,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 10, no. 3, pp. 344–358, 2000.
9. A Siarohin, S Lathuilière, S Tulyakov, "First order motion model for image animation." Advances in Neural Information Pro-cessing Systems 32 (2019): 7137-7147.
10. B. Chen, et al. "Beyond Keypoint Coding: Temporal Evolution Inference with Compact Feature Representation for Talking Face Video Compression." Pro-ceedings of the IEEE Data Compression Conference, 2022.
11. T.-C. Wang, A. Mallya, and M.-Y. Liu. "One-shot free-view neural talking-head synthesis for video conferencing." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.
12. R. Zhang, et al. "The unreasonable effectiveness of deep features as a perceptual metric." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
13. K. Ding, et. al., “Image quality assessment: Unifying structure and texture similarity," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020
14. K. Ding, et al. "Comparison of full-reference image quality models for optimization of image processing systems." International Journal of Computer Vision 129.4 (2021): 1258-1281
15. Video database: https://ibug.doc.ic.ac.uk/resources/300-VW/

Q & A