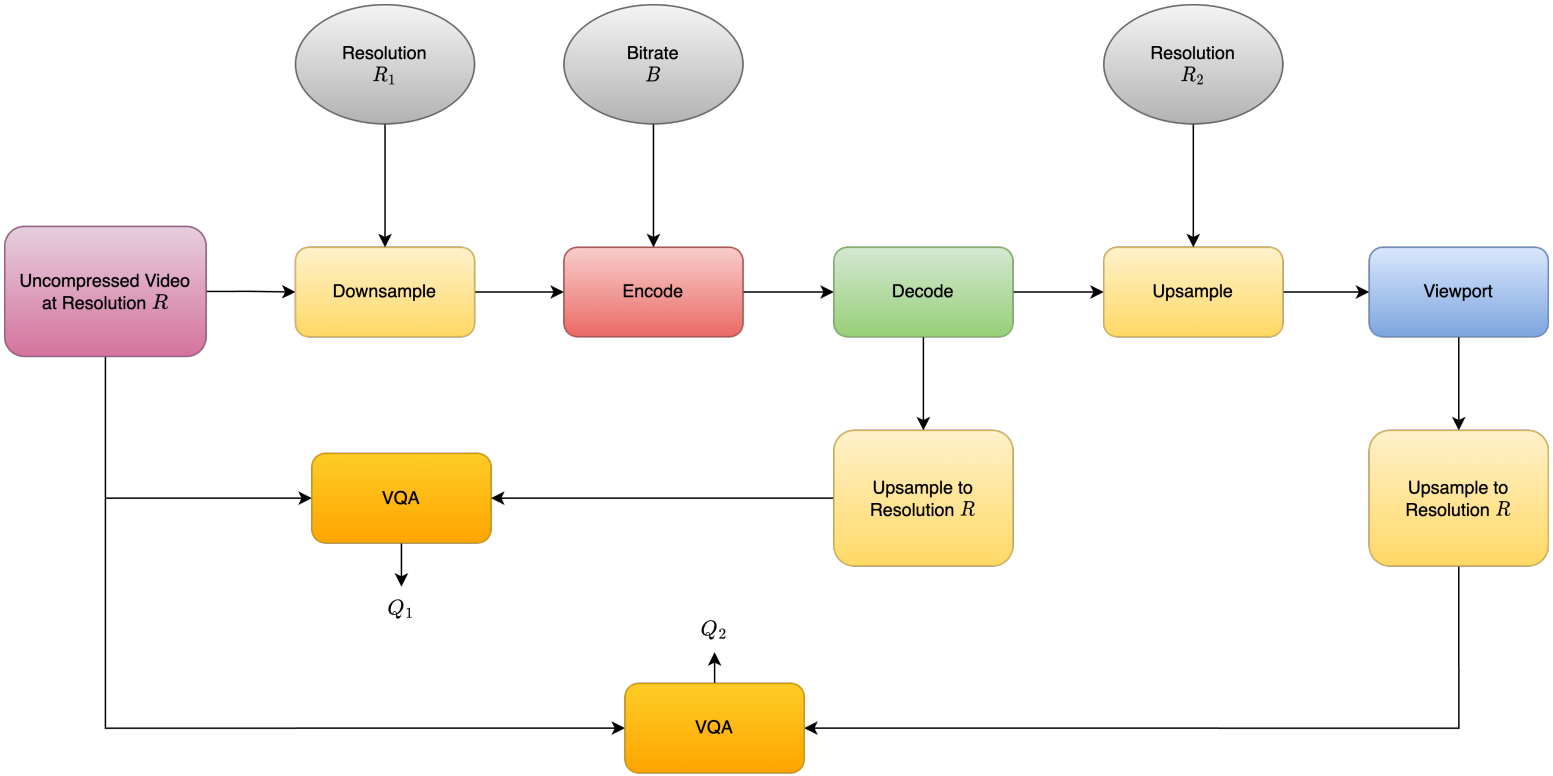
TEXAS
The University of Texas at Austin

Bitrate Ladder Construction using Visual Information Fidelity

Krishna Srikar Durbha
The University of Texas at Austin

Table of Contents:

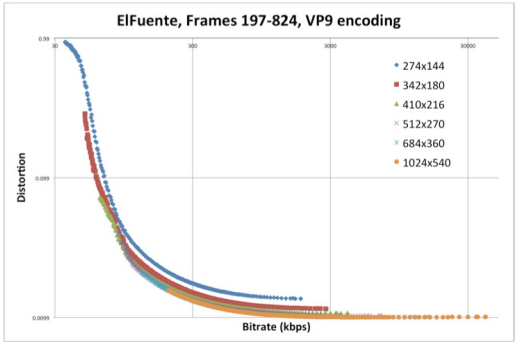
Video-Delivery Architecture

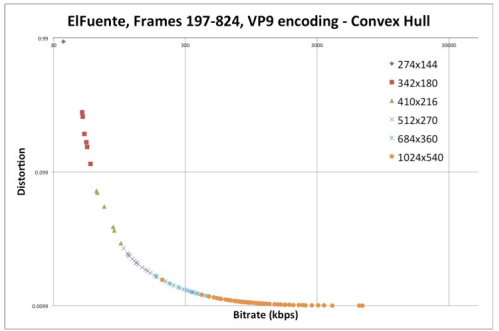


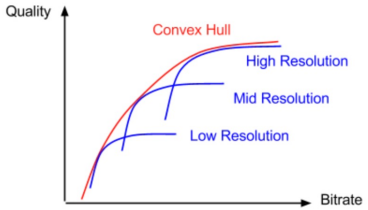
Fixed Bitrate Ladder

Bitrate (kbps)	Resolution
235	320x240
375	384x288
560	512x384
750	512x384
1050	640x480
1750	720x480
2350	1280x720
3000	1280x720
4300	1920x1080
5800	1920x1080

Per-Title Video Encoding







- Per-title encoding schemes design a convex hull tailored for each video, containing optimal bitrate-resolution pairs that display the highest visual quality at the target bitrate.
- The convex hull is where the encoding point achieves Pareto efficiency.
- The main disadvantage of this approach is the requirement for significant computation resources and time.

[1] https://netflixtechblog.com/per-title-encode-optimization-7e99442b62a2
[2] https://netflixtechblog.com/dynamic-optimizer-a-perceptual-video-encoding-optimization-framework-e19f1e3a277f

Previous Works

Objective	Quality-Metric	Features	ML-Algorithm	Time-Complexity	Cite
Predicting Rate-Quality Curves	PSNR	GLCM, TC	SVRM	-	[1]
	PSNR	GLCM, NCC, ALPD, NLP, TC, OF	SVRM	-	[2]
Predicting Bitrate Ladder	PSNR	GLCM, TC	Gaussian Processes Regression	0.18[1]	[3]
	PSNR	GLCM, TC, RsMSE	Gaussian Processes Regression	0.18[1]	[4]
	VMAF	GLCM, TC, NCC, RsMSE	Gaussian Processes Regression	-	[5]
	VMAF	E, h	Linear Regression	370fps[2]	[6]
	VMAF	E_Y, E_U ,E_V, h, L_Y, L_U, L_V, bitrate, bitstream-features, GLCM	Random-Forest + SVR	300fps[2]	[7]
	VMAF	Sobel(Y), Y, YDiff, U, V	Random-Forest	0.08-0.14[3]	[8]
	VMAF	Video	Neural Network	-	[9]
	VMAF	Video-Chunk	Conv-GRU	4.76-23.8fps[2]	[10]
	PSNR	GLCM, TC	Gaussian Processes Regression	0.08-0.14[3]	[11]
	PSNR VMAF	GLCM, TC, SI, TI, C, Noise, NCC	CNNs, and LSTMs	-	[12]

[1] Angeliki V. Katsenou, Mariana Afonso, Dimitris Agrafiotis, and David R. Bull, "Predicting video rate-distortion curves using textural features," in 2016 Picture Coding Symposium, PCS 2016, Nuremberg, Germany, December 4-7, 2016. 2016, pp. 1–5, IEEE.

[2] Angeliki V. Katsenou, Mariana Afonso, and David R. Bull, "Study of compression statistics and prediction of rate-distortion curves for video texture," Signal Process. Image Commun., vol. 101, pp. 116551, 2022.

[3] Angeliki V. Katsenou, Joel Sole, and David R. Bull, "Content-gnostic bitrate ladder prediction for adaptive video streaming," in Picture Coding Symposium, PCS 2019, Ningbo, China, November 12-15, 2019. 2019, pp. 1–5, IEEE.

[4] Angeliki V. Katsenou, Joel Sole, and David R. Bull, "Efficient bitrate ladder construction for content-optimized adaptive video streaming," IEEE Open Journal of Signal Processing, vol. 2, pp. 496–511, 2021.

[5] Angeliki V. Katsenou, Fan Zhang, Kyle Swanson, Mariana Afonso, Joel Sole, and David R. Bull, "Vmaf-based bitrate ladder estimation for adaptive streaming," in Picture Coding Symposium, PCS 2021, Bristol, United Kingdom, June 29 - July 2, 2021. 2021, pp. 1–5, IEEE.

[6] Vignesh V. Menon, Hadi Amirpour, Mohammad Ghanbari, and Christian Timmerer, "Perceptually-aware per-title encoding for adaptive video streaming," in IEEE International Conference on Multimedia and Expo, ICME 2022, Taipei, Taiwan, July 18-22, 2022. 2022, pp. 1–6, IEEE.

[7] Vignesh V. Menon, Jingwen Zhu, Prajit T. Rajendran, Hadi Amirpour, Patrick Le Callet, and Christian Timmerer, "Just noticeable difference-aware per-scene bitrate-laddering for adaptive video streaming," CoRR, vol. abs/2305.00225, 2023.

[8] Zhenwei Yang and Liquan Shen, "A multi-category task for bitrate interval prediction with the target perceptual quality," KSII Trans. Internet Inf. Syst., vol. 15, no. 12, pp. 4476–4491, 2021

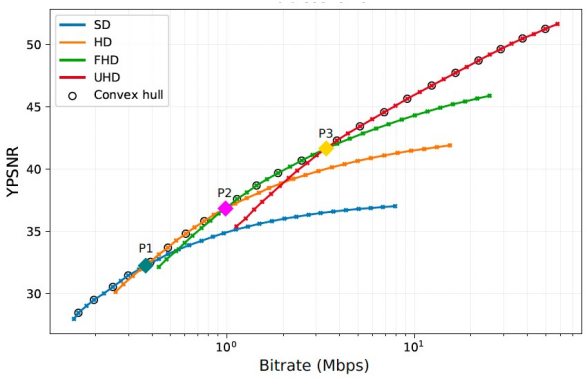
[9] Andreas Kah, Maurice Klein, Christoph Burgmair, Markus Rasokat, Wolfgang Ruppel, and Matthias Narroschke, "An algorithm for a quality-optimized bit rate ladder generation for video streaming services using a neural network," in Applications of Digital Image Processing XLV. SPIE, 2022, vol. 12226, pp. 338–346.

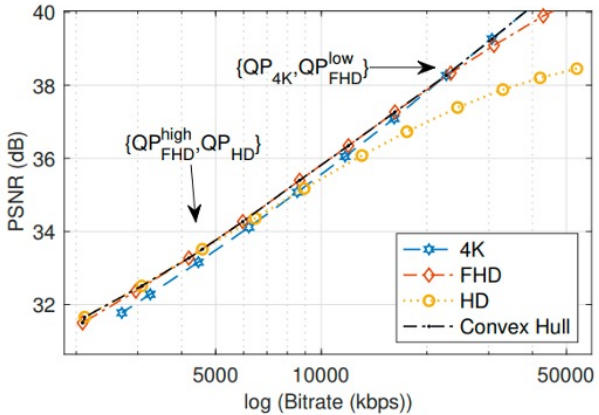
[10] Somdyuti Paul, Andrey Norkin, and Alan C. Bovik, "Efficient per-shot convex hull prediction by recurrent learning," CoRR, vol. abs/2206.04877, 2022.

[11] Fatemeh Nasiri, Wassim Hamidouche, Luce Morin, Nicolas Dhollande, and Jean-Yves Aubi´e, "Ensemble learning for efficient VVC bitrate ladder prediction," in 10th European Workshop on Visual Information Processing, EUVIP 2022, Lisbon, Portugal, September 11-14, 2022. 2022, pp. 1–6, IEEE.

[12] Ahmed Telili, Wassim Hamidouche, Sid Ahmed Fezza, and Luce Morin, "Benchmarking learning-based bitrate ladder prediction methods for adaptive video streaming," in Picture Coding Symposium, PCS 2022, San Jose, CA, USA, December 7-9, 2022. 2022, pp. 325–329, IEEE.

Predicting Cross-Over Points

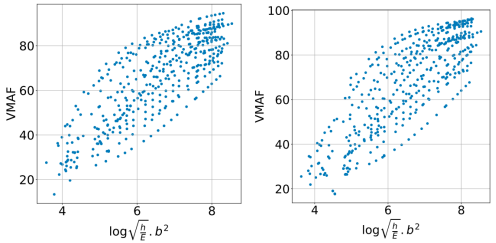




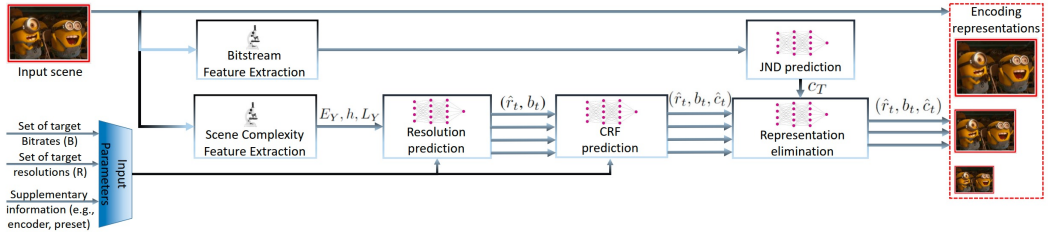
- Cross-Over points are defined as intersection points between RQ curves of different resolutions where switching happens low-resolution to higher resolution.
- They are defined as either a pair of QPs i.e one for each resolution or as a cross-over bitrate.
- These cross-over points are predicted using low-level features like Low-level features like GLCM, TC, NCC, etc extracted from uncompressed videos and are used to construct bitrate ladders.

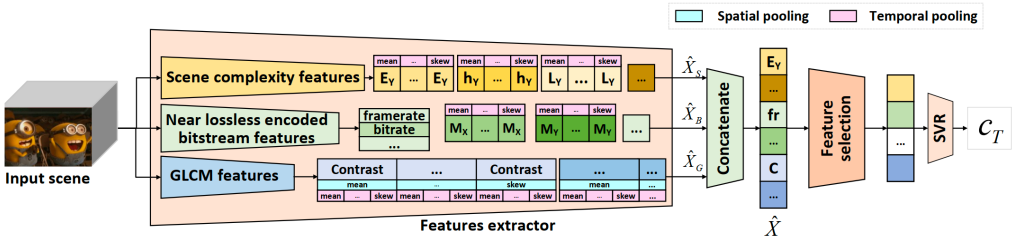
[1]: Angeliki V. Katsenou, Joel Sole, and David R. Bull, “Efficient bitrate ladder construction for content-optimized adaptive video streaming,” IEEE Open Journal of Signal Processing, vol. 2, pp. 496–511, 2021.
[2]: Ahmed Telili, Wassim Hamidouche, Sid Ahmed Fezza, and Luce Morin, “Benchmarking learning-based bitrate ladder prediction methods for adaptive video streaming,” in Picture Coding Symposium, PCS 2022, San Jose, CA, USA, December 7-9, 2022. 2022, pp. 325–329, IEEE.

Predicting Quality or Optimal Resolution



(a) 720p (b) 1080p



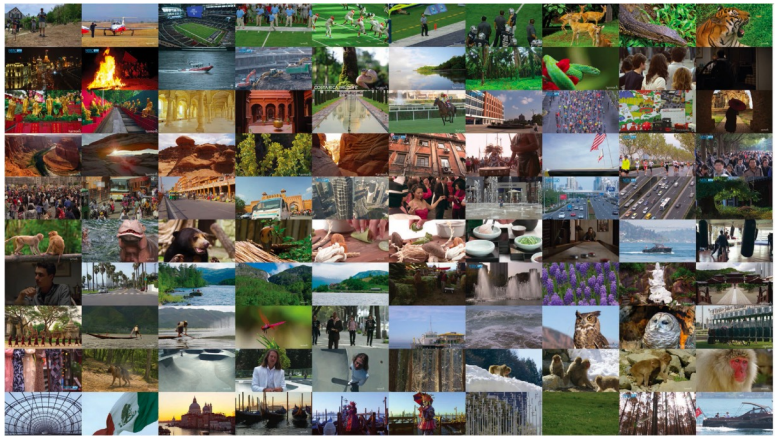


- DCT-based energy features are introduced by the authors in [1],[2] as a replacement for conventional features like SI and TI.
- These features are used in the construction of bitrate ladder by either modelling quality of compressed videos as a linear regression or directly predicting optimal resolution based on quality predictions.

[1] Vignesh V. Menon, Hadi Amirpour, Mohammad Ghanbari, and Christian Timmerer, "Perceptually-aware per-title encoding for adaptive video streaming," in IEEE International Conference on Multimedia and Expo, ICME 2022,
Taipei, Taiwan, July 18-22, 2022. 2022, pp. 1–6, IEEE.
[2] Vignesh V. Menon, Jingwen Zhu, Prajit T. Rajendran, Hadi Amirpour, Patrick Le Callet, and Christian Timmerer, "Just noticeable difference-aware per-scene bitrate-laddering for adaptive video streaming," CoRR, vol. abs/2305.00225, 2023.

Dataset:

- Name: BVT-100 4K
- Authors: M. Afonso, A. Katsenou, F. Zhang, and D. R. Bull
- 100 UHD video sequences downloaded from various public sources like Netflix Chimera, Ultra Video Group, Harmonic Inc, SJTU and AWS Elemental.
- All the sequences were spatially cropped to UHD (if originally 4K resolution, 4096 × 2160 pixels), converted to 4:2:0 chroma subsampling (if originally 4:2:2 or 4:4:4) and temporally cropped to 64 frames.
- Each sequence contains a single scene (without scene cuts) and the majority of the test sequences have a frame rate of 60 fps and bit depth of 10 bits per sample.



Experimental-Setup:

- The videos are compressed using **libx265** codec with the encoder preset set to **medium**.
- Resolutions of compressed videos are set to the following values {3860 × 2160, 1920 × 1080, 1280 × 720, 960 × 540, 768 × 432, 640 × 360} which almost have a similar aspect ratio of 16 : 9.
- The encoding is performed using constant-quality/CRF settings with the following CRF values {16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,37,39,41}.
- We estimated VMAF for each of the compressed video assuming that videos are viewed on a 4K screen.
- In our experiments, we consider rate-quality points that have a minimum VMAF value of 10 and maximum VMAF value of 95.

$$\mathcal{C} = \mathcal{S}.\mathcal{U} = \{S_i.\vec{U}_i : i \in I\}$$

$$\mathcal{E} = \mathcal{C} + \mathcal{N}$$

$$I(\vec{C}^N; \vec{E}^N | s^N) = \sum_{j=1}^{N} \sum_{i=1}^{N} I(\vec{C}_i; \vec{E}_j | \vec{C}^{i-1}, \vec{E}^{j-1}, s^N)$$

$$I(\vec{C}^N; \vec{E}^N | S^N = s^N) = \sum_{i=1}^{N} I(\vec{C}_i; \vec{E}_i | s_i)$$

$$I(\vec{C}^N; \vec{E}^N | s^N) = \frac{1}{2} \sum_{i=1}^{N} \log_2 \left(\frac{|s_i^2 \mathbf{C_U} + \sigma_n^2 \mathbf{I}|}{|\sigma_n^2 \mathbf{I}|} \right),$$

$$I(\vec{C}^N; \vec{E}^N | s^N) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{M} \log_2 \left(1 + \frac{s_i^2 \lambda_j}{\sigma_n^2}\right)$$

$$I_{k,b}^j = \frac{1}{N} \sum_{i=1}^{N} \log_2 \left(1 + \frac{s_i^2 \lambda_j}{\sigma_n^2}\right)$$

$$I_{k,b} = \frac{1}{N} \sum_{j=1}^{M} \sum_{i=1}^{N} \log_2 \left(1 + \frac{s_i^2 \lambda_j}{\sigma_n^2}\right)$$

$$I_k = \frac{1}{2} \sum_{b=1}^{2} I_{k,b}.$$

Our Approach

Feature Set Number	Features Set	No.of Features
1	$I_k[F_i]$	4
2	$I_{k,b}[F_i]$	8
3	$I_{k,b}^j[F_i]$	72
4	$I_k[F_i]$, $\lvert D_i\rvert$	5
5	$I_{k,b}[F_i]$, $\lvert D_i\rvert$	9
6	$I_{k,b}^j[F_i]$, $\lvert D_i\rvert$	73
7	$I_k[F_i]$, $\lvert D_i\rvert$, $I_k[D_i]$	9
8	$I_{k,b}[F_i]$, $\lvert D_i\rvert$, $I_{k,b}[D_i]$	17
9	$I_{k,b}^j[F_i]$, $\lvert D_i\rvert$, $I_{k,b}^j[D_i]$	145

Performance Comparisons:

We compare the performance against Apple's fixed bitrate ladder and reference bitrate ladder constructed using exhaustive encoding.

Our Approach:

- Bitrate Ladder constructed by predicting quality using VIF feature sets
- Bitrate Ladder constructed by predicting using Low-Level features and VIF feature sets

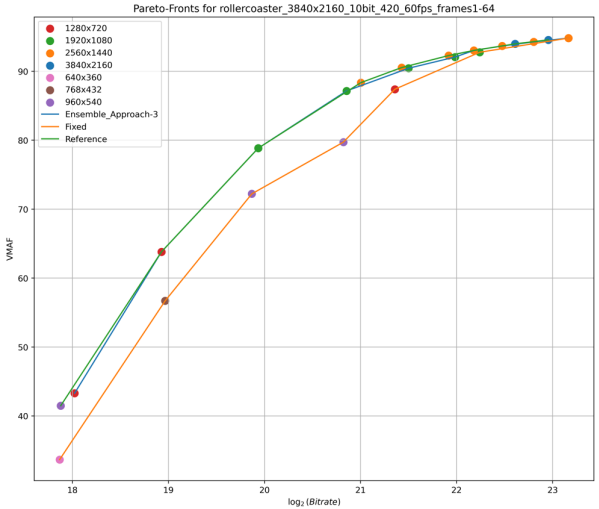
Comparisons:

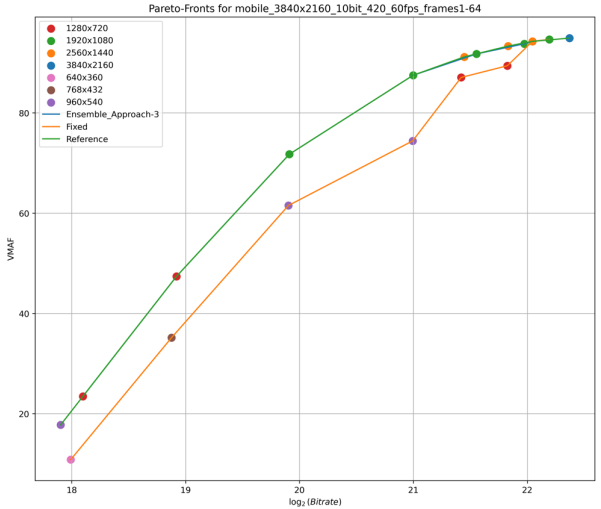
- Bitrate Ladder constructed by predicting Cross-Over Bitrates using Low-level features
- Bitrate Ladder constructed using by predicting quality using Low-level features.

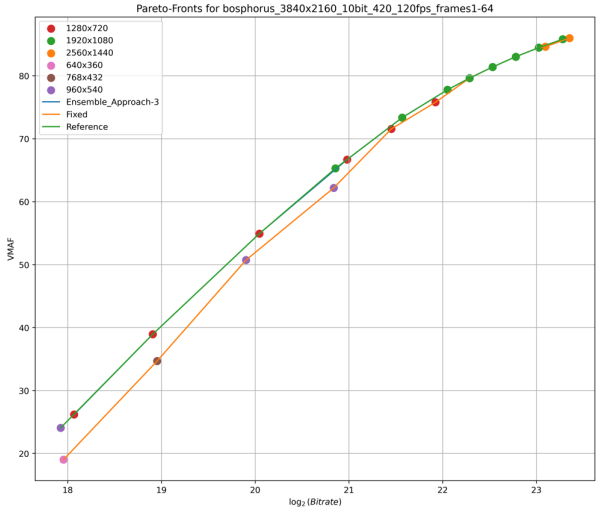
Feature	Formula	No.of Spatial Features	No.of Spatio-Temporal Features
GLCM	$F_2\{F_1\{correlation(GLCM)\}\}$, $F_2\{F_1\{homogeneity(GLCM)\}\}$, $F_2\{F_1\{contrast(GLCM)\}\}$, $F_2\{F_1\{energy(GLCM)\}\}$ where GLCM is calculated on blocks of size (64,64), $F_1 = \{mean, std\}$ and $F_2 = \{mean, std, skew, kurtosis\}$	8	32
TC	$F_2\{F_1\{Coherence\}\}$ where $F_1 = \{mean, std, skew, kurtosis\}$ and $F_2 = \{mean, std\}$	4	8
SI	$F_2\{F_1\{Sobel(Y)\}\}$ where $F_1 = \{mean, std\}$ and $F_2 = \{mean, std, skew, kurtosis\}$	2	8
TI	$F_2\{F_1\{(Y_2 - Y_1)\}\}$ where $F_1 = \{mean, std\}$ and $F_2 = \{mean, std, skew, kurtosis\}$	2	8
CTI	$F_2\{F_1\{Y\}\}$ where $F_1 = \{mean, std\}$ and $F_2 = \{mean, std, skew, kurtosis\}$	2	8
CF	$F_2\{(YUV)\}$ where $F_2 = \{mean, std, skew, kurtosis\}$	1	4
CI	$F_2\{F_1\{U\}\}$, $F_2\{W_R \times F_1\{V\}\}$ where $W_R = 5$, $F_1 = \{mean, std\}$ and $F_2 = \{mean, std, skew, kurtosis\}$	4	16
Texture-DCT	$F_2\{E_Y\}$, $F_2\{h_Y\}$, $F_2\{L_Y\}$, $F_2\{E_U\}$, $F_2\{h_U\}$, $F_2\{L_U\}$, $F_2\{E_V\}$, $F_2\{h_V\}$, $F_2\{L_V\}$ where $F_2 = \{mean\}$	9	9
Total		32	93

Results

Features Set	BL vs Fixed Bitrate Ladder		BL vs Reference Bitrate Ladder	
	BD-Rate (in %)	BD-VMAF (in dB)	BD-Rate (in %)	BD-VMAF (in dB)
Feature-Set-1, b, $\frac{w}{3840}$, $\frac{h}{3840}$	$-11.045/17.688$	$2.978/4.112$	$0.505/0.465$	$-1.107/1.34$
Feature-Set-2, b, $\frac{w}{3840}$, $\frac{h}{3840}$	$-10.809/17.578$	$2.967/4.159$	$0.525/0.534$	$-1.135/1.509$
Feature-Set-3, b, $\frac{w}{3840}$, $\frac{h}{3840}$	$-12.995/18.043$	$3.748/3.848$	$0.355/0.461$	$-0.495/1.019$
Feature-Set-4, $\frac{w}{3840}$, $\frac{h}{3840}$	$-12.972/16.069$	$3.399/3.872$	$0.367/0.35$	$-0.741/1.07$
Feature-Set-5, b, $\frac{w}{3840}$, $\frac{h}{3840}$	$-13.12/15.783$	$3.499/3.776$	$0.379/0.34$	$-0.692/1.09$
Feature-Set-6, b, $\frac{w}{3840}$, $\frac{h}{3840}$	$-13.259/17.967$	$3.756/3.776$	$0.378/0.459$	$-0.506/1.036$
Feature-Set-7, b, $\frac{w}{3840}$, $\frac{h}{3840}$	$-12.757/15.208$	$3.427/3.7$	$0.454/0.464$	$-0.709/1.184$
Feature-Set-8, b, $\frac{w}{3840}$, $\frac{h}{3840}$	$-10.943/17.208$	$2.997/4.06$	$0.516/0.521$	$-1.069/1.367$
Feature-Set-9, b, $\frac{w}{3840}$, $\frac{h}{3840}$	$-10.417/17.466$	$3.065/3.607$	$0.605/0.693$	$-0.931/1.174$
Low-Level-Features, Feature-Set-1, b, $\frac{w}{3840}$, $\frac{h}{3840}$	$-14.597/12.24$	$3.606/3.08$	$0.26/0.338$	$-0.688/1.054$
Low-Level-Features, Feature-Set-2, b, $\frac{w}{3840}$, $\frac{h}{3840}$	$-15.795/12.685$	$4.088/3.239$	$0.188/0.262$	$-0.28/0.82$
Low-Level-Features, Feature-Set-3, b, $\frac{w}{3840}$, $\frac{h}{3840}$	$-16.245/12.825$	$4.231/3.275$	$0.126/0.251$	$-0.073/0.75$
Low-Level-Features, Feature-Set-4, $\frac{w}{3840}$, $\frac{h}{3840}$	$-14.713/10.95$	$3.588/2.695$	$0.285/0.436$	$-0.724/1.153$
Low-Level-Features, Feature-Set-5, b, $\frac{w}{3840}$, $\frac{h}{3840}$	$-15.214/12.625$	$3.873/3.144$	$0.212/0.321$	$-0.434/1.094$
Low-Level-Features, Feature-Set-6, b, $\frac{w}{3840}$, $\frac{h}{3840}$	$-14.649/16.074$	$4.118/3.397$	$0.214/0.416$	$-0.233/0.832$
Low-Level-Features, Feature-Set-7, b, $\frac{w}{3840}$, $\frac{h}{3840}$	$-14.047/10.42$	$3.458/2.543$	$0.34/0.505$	$-0.827/1.272$
Low-Level-Features, Feature-Set-8, b, $\frac{w}{3840}$, $\frac{h}{3840}$	$-14.685/16.199$	$4.141/3.398$	$0.231/0.41$	$-0.192/0.795$
Low-Level-Features, Feature-Set-9, b, $\frac{w}{3840}$, $\frac{h}{3840}$	$-14.883/16.626$	$4.174/3.548$	$0.196/0.419$	$-0.134/0.866$
Low-Level-Features, b, $\frac{w}{3840}$, $\frac{h}{3840}$	$-13.042/15.417$	$3.468/3.629$	$0.416/0.38$	$-0.601/0.991$
Low-Level-Features (Cross-Over bi-rates)	$-16.838/14.388$	$4.275/3.865$	$0.037/0.156$	$-0.102/0.699$

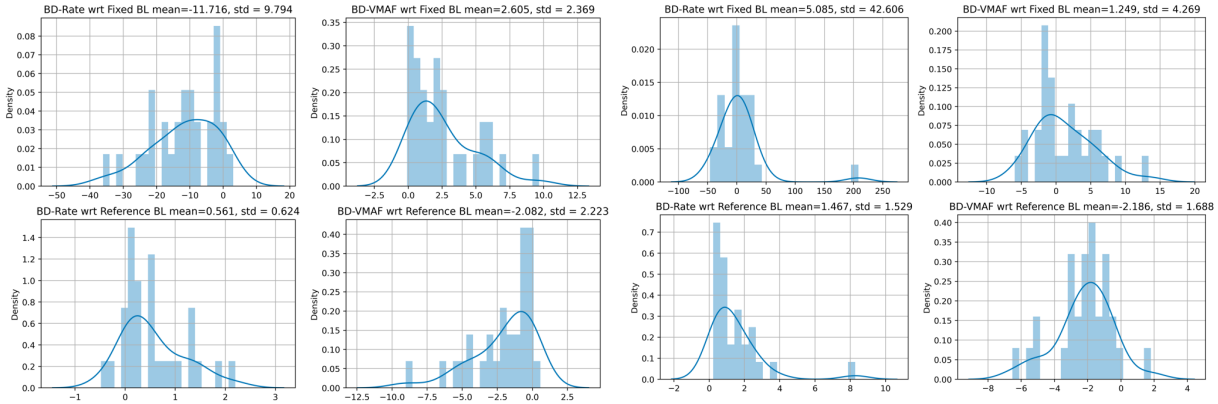

Pareto-Fronts for rollercoaster_3840x2160_10bit_420_60fps_frames1-64
1280x720
1920x1080
2560x1440
3840x2160
640x360
768x432
960x540
Ensemble_Approach-3
Fixed
Reference
VMAF
log2 (Bitrate)


Pareto-Fronts for mobile_3840x2160_10bit_420_60fps_frames1-64
1280x720
1920x1080
2560x1440
3840x2160
640x360
768x432
960x540
Ensemble_Approach-3
Fixed
Reference
VMAF
log2 (Bitrate)


Pareto-Fronts for bosphorus_3840x2160_10bit_420_120fps_frames1-64
1280x720
1920x1080
2560x1440
640x360
768x432
960x540
Ensemble_Approach-3
Fixed
Reference
VMAF
log2 (Bitrate)

Future Works:

- Improving the performance of models.
- Exploring more bitrate ladder correction algorithms.
- Developing a robust evaluation metrics for different methods

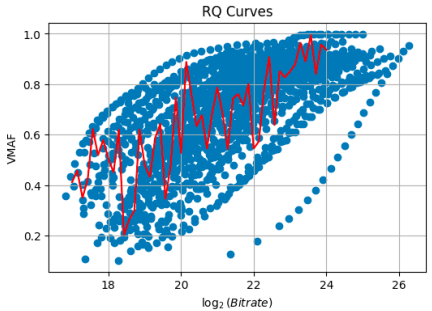


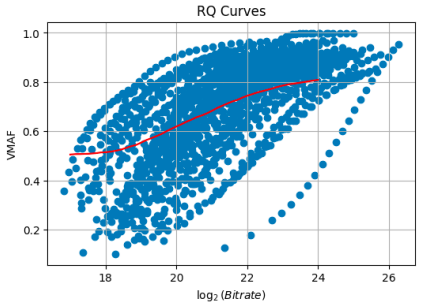
Bitrate (in kbps)	**Resolution**
10500	(3840, 2160)
9000	(3840, 2160)
8000	(3840, 2160)
7000	(3840, 2160)
6000	(1920, 1080)
5000	(3840, 2160)
4000	(2560, 1440)
3000	(2560, 1440)
2000	(1920, 1080)
1000	(1920, 1080)
500	(768, 432)
250	(960, 540)

(a) Predicted Bitrate Ladder before correctic

Bitrate (in kbps)	**Resolution**
10500	(3840, 2160)
9000	(3840, 2160)
8000	(3840, 2160)
7000	(3840, 2160)
6000	(1920, 1080)
5000	(1920, 1080)
4000	(1920, 1080)
3000	(1920, 1080)
2000	(1920, 1080)
1000	(1920, 1080)
500	(768, 432)
250	(768, 432)

(b) Predicted Bitrate Ladder after correction





Thanks to Meta Platforms for funding our research.

Thanks to:
- Hassene Tmar
- Cosmin Stejerean
- Haixiong Wang
- Ioannis Katsavounidis
- Alan C. Bovik